



HAL
open science

Processus empiriques, estimation non paramétrique et données censurées.

Vivian Viallon

► **To cite this version:**

Vivian Viallon. Processus empiriques, estimation non paramétrique et données censurées.. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2006. Français. NNT: . tel-00119260

HAL Id: tel-00119260

<https://theses.hal.science/tel-00119260>

Submitted on 8 Dec 2006

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.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : Mathématiques

Option : Statistique

présentée par

Vivian VIALLO

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS 6

Sujet de la thèse :

**Processus empiriques, estimation
non paramétrique et données
censurées.**

Soutenue le 01/12/2006

Directeur de thèse	PAUL DEHEUELS	<i>Université Paris VI</i>
Rapporteurs	PHILIPPE VIEU LAURENT BORDES	<i>Université de Toulouse Université de Pau</i>
Examineurs	INGRID VAN KEILEGOM ODILE PONS MICHEL BRONIATOWSKI	<i>Université de Louvain INRA Jouy-en-Josas Université Paris VI</i>

Remerciements

Je tiens tout d'abord à remercier le Professeur Paul Deheuvels pour m'avoir encadré durant ces trois années de thèse. Je tiens à lui témoigner ma gratitude pour m'avoir toujours laissé libre de choisir parmi les multiples sujets de recherche qu'il pouvait me proposer. Merci aussi de m'avoir insufflé la rigueur mathématique nécessaire à l'exercice délicat de la rédaction de ce manuscrit.

J'adresse un remerciement tout particulier à Philippe Vieu et Laurent Bordes pour avoir accepté d'être mes rapporteurs, ainsi qu'à Ingrid Van Keilegom, Odile Pons et Michel Broniatowski pour avoir accepté de faire partie de mon jury.

Je tiens aussi à remercier Joël Coste pour m'avoir proposé le poste d'AHU que j'occupe actuellement, et pour m'avoir laissé le temps nécessaire à la préparation de ma soutenance. Je voudrais aussi témoigner ma gratitude à Mounir Mesbah pour m'avoir « mis sur les rails » du poste d'AHU.

Je suis également reconnaissant à Armelle Guillou pour le temps qu'elle a su m'accorder à divers moments de ma thèse, ainsi que pour la pédagogie dont elle a fait preuve pendant ses cours.

Un grand merci à Stéphane Ragusa pour m'avoir poussé vers la voie de la recherche (un choix dont je ne peux que me féliciter aujourd'hui), et pour m'avoir proposé un sujet d'étude épidémiologique aussi stimulant que celui des scores de risque de maladie. Merci aussi évidemment à Françoise Clavel-Chapelon, directrice de l'équipe E3N, pour m'avoir accueilli dans cette jeune et dynamique équipe. Qu'il me soit permis de remercier Catherine Com-Nougué pour son écoute et ses conseils, pas seulement en biostatistique. Mes pensées vont aussi à l'ensemble de mes collègues E3N, et assimilés, par ordre d'"ancienneté" : Maryvonne, Marie et Lyan, Agnès, Marina ($\times 2$), Bertrand, Nadia, Julien, Greg, Violaine, Séverine, Mathieu (U605), Alban, Marie-Noël, Gaël, etc. Merci pour tous ces bons moments, les parties de foot à midi, les picnics au parc des Hautes-Bruyères, les parties de tarot du mois d'août... bref l'ambiance décontractée, et par là, propice au travail.

Je tiens également à remercier les doctorants du LSTA (Davit, David, Salim, Alexandre, Pierre, Anne, etc) pour leur aide lorsque le besoin s'en faisait sentir, mais aussi et surtout leur bonne humeur. Je tiens évidemment à exprimer un merci tout spécial, doublé d'un grand respect, à Mohammed pour sa motivation communicative, sa patience et sa sérénité, faisant de ces premières expériences de co-rédaction un bonheur quotidien. J'aimerais remercier aussi Mmes Lamart et Durande ainsi que M. Epron pour leur sympathie et leur disponibilité de chaque instant.

Un grand merci évidemment à mes amis (et plus), de tout bord. Par ordre de proximité géographique à l'heure où j'écris : Virginie, Philippe et Claire, Bertrand et Ségo, Olive et Julie, Maxime et Yasmine, Juan Carlos, Cyril et Lucie, Pascale et Pierre, Rémy, Fif, Mickael, Mimi, Matthieu, Fred, Ludo, Pierre-Louis et Charlotte, Sylvain et Lila (et Matéo aussi!), Benoit et Virginie (et Louison aussi!), LS, Matthieu et Blanka, P'tit Jules et Nadège, Raoul... Merci pour tous ces bons moments.

Encore quelques lignes, sans doute trop courtes, pour exprimer toute la gratitude que j'ai pour mes parents, mes grands-parents et mes oncles et tante Jean-Paul, Mizou et Patou pour tout ce qu'ils ont fait pour moi. Mes pensées vont enfin à ma soeur et mon frère.

Abréviations

- a.p.c.r.* : à partir d'un certain rang.
a.s. : almost surely.
c.d.f. : conditional distribution function.
i.i.d. : indépendante et identiquement distribuée.
I.P.C.W. : inverse probability of censoring weighted.
L.L.F.U. : loi du logarithme fonctionnelle uniforme.
L.L.U. : loi du logarithme uniforme.
p.s. : presque sûrement.
U.F.L.L. : uniform functional limit law.

Table des matières

Introduction Générale	1
0.1 Le processus empirique uniforme	2
0.2 Le processus empirique généralisé	5
0.3 Les estimateurs non paramétriques à noyau	6
0.3.1 Les modèles additifs	10
0.4 Le cas des données censurées	13
0.4.1 Définition	13
0.4.2 Quantités mathématiques invoquées	15
0.4.3 L'estimateur de Kaplan-Meier	16
1 A U.F.L.L. for the increments of the normed sample quantile process.	19
1.1 Introduction	19
1.1.1 Motivations	19
1.1.2 Notation and Main Results	20
1.2 The Uniform Case	22
1.3 Some Applications	23
1.3.1 The k -Spacings	23
1.3.2 A Law of the Logarithm for Nearest-Neighbor Density Estimators	25
1.4 Proofs	25
1.4.1 Proof of Theorem 1.2.1	25
1.4.2 Proof of Theorem 1.1.1	30
1.4.3 Proof of Theorem 1.3.2	33
1.4.4 Proof of Theorem 1.3.1	35
1.4.5 Proof of Theorem 1.3.3	35
1.5 Appendix	37
1.5.1 Useful facts	37
2 L.L.F.U pour les incréments du processus empirique avec censure.	39
2.1 Présentation générale- Principaux résultats	39
2.1.1 Introduction- Notations	39
2.1.2 Résultats de la littérature	42
2.1.3 Nouveaux résultats	43
2.2 Preuves des théorèmes 2.1.5, 2.1.6 et 2.1.7	45
2.2.1 Quelques notations supplémentaires	45
2.2.2 Un résultat d'approximation utile	46
2.2.3 Preuve du théorème 2.1.5	49
2.2.4 Preuve du théorème 2.1.6	53
2.2.5 Preuve du théorème 2.1.7	54

3	L.L.U. pour les estimateurs non paramétriques à noyau.	55
3.1	Introduction-Résultats	55
3.1.1	Notations et hypothèses	55
3.1.2	Résultats	59
3.1.3	Un résultat utile.	60
3.1.4	Corollaires	61
3.2	Démonstration du théorème 3.1.3	62
3.2.1	Le cas $\mathcal{F} = \{\psi\}$	62
3.2.2	Démonstration du théorème 3.1.3 dans le cas général	76
3.3	Démonstration du corollaire 3.1.3	79
3.4	Appendice	80
3.4.1	Classes de fonctions séparables point par point	80
3.4.2	Faits importants	80
4	L.L.U. pour un estimateur de la régression en données censurées.	83
4.1	Introduction	83
4.2	Définition de l'estimateur	85
4.3	Résultats principaux	87
4.3.1	Notations et hypothèses supplémentaires	87
4.3.2	Résultats	90
4.4	Applications : construction de bandes de confiance	91
4.4.1	Résultat nécessaire	91
4.4.2	Hypothèses nécessaires supplémentaires	92
4.4.3	Construction des bandes de confiance	92
4.4.4	Motivation de l'introduction de \mathcal{V}_I et $\log_{\theta, K}$ - Choix de θ	94
4.4.5	Illustration	96
4.5	Démonstrations	98
4.5.1	Démonstration du théorème 4.3.1	98
4.5.2	Démonstration du lemme 4.4.1	100
4.6	Appendice	101
4.6.1	Justification de l'écriture additive de l'estimateur de Kaplan-Meier	101
4.6.2	Etude du biais des estimateurs à noyau de la régression	102
5	Régression additive en données censurées.	105
5.1	Introduction	105
5.2	Notations	106
5.3	Présentation des hypothèses et résultats	109
5.4	Démonstration du théorème 5.3.1	110
5.4.1	Introduction	110
5.4.2	Cas où la densité f des covariables et la fonction \bar{G} sont connues	110
5.4.3	Cas où la densité des covariables est inconnue et la fonction \bar{G} connue	118
5.4.4	Cas où la densité des covariables et la fonction \bar{G} sont inconnues	119
A	Les classes à nombre de recouvrement polynomial	121
A.1	Nombres de recouvrement d'un ensemble	121
A.2	Classes d'ensembles de Vapnik-Červonenkis	122
A.2.1	Définitions et notations	122
A.2.2	Propriétés fondamentales des classes VC	123
A.2.3	Construction de classes VC	125
A.3	Les classes de fonctions VC	126

A.3.1	Définitions	127
A.3.2	Propriétés-Exemples	127
A.4	Un résultat utile	127
Perspectives de recherche		129
A.5	Fonctionnelles de la loi conditionnelle en données censurées	129
A.5.1	Estimateurs des lois conditionnelles avec censure	129
A.5.2	Le cas « âge en échelle de temps »	129
A.6	Modèles additifs et régression censurées	130
A.7	Constructions de scores de risques de pathologies	130
Bibliographie		130

Introduction Générale

La théorie des processus empiriques joue un rôle central en statistique, puisqu'elle concerne l'ensemble des résultats limites généraux se rapportant aux échantillons aléatoires. De ce fait, elle comporte d'innombrables applications à des problèmes particuliers. Parmi les plus importantes propriétés de ces modèles mathématiques, ayant fait l'objet de recherches approfondies depuis l'origine de la statistique moderne, mentionnons, parmi d'autres, les résultats liés aux théorèmes de Glivenko-Cantelli (Glivenko (1933), Cantelli (1933)), aux théorèmes de Donsker (Donsker (1951)), aux lois du logarithme itéré (Chung (1949), voir Deheuvels (1991) et la bibliographie attenante), ou encore aux lois limites fonctionnelles, globales ou locales, (Finkelstein (1971), Deheuvels (1992), Deheuvels et Mason (1992), Deheuvels (2000)). Pour une description approfondie des résultats classiques ayant trait aux processus empiriques, et à leurs applications, nous renvoyons, parmi de nombreux autres textes de référence possibles, aux ouvrages de Csörgő et Révész (1981), de Pollard (1984), de Shorack et Wellner (1986), de van der Vaart et Wellner (1996), de Dudley (1999), et aux bibliographies citées dans ces livres. Dans le mémoire qui suit, nous avons entrepris d'approfondir certains de ces résultats. Nous avons voulu également montrer dans quelle mesure nos contributions permettaient d'établir de nouvelles propriétés de certains estimateurs non paramétriques, notamment dans le cas où ces estimateurs sont basés sur des données *censurées* (voir plus loin pour des définitions précises de ce concept).

Au début des années 1980, Stute (voir Stute (1982b), Stute (1982a), Stute (1986a), Stute (1986b)) fut l'un des premiers statisticiens (voir également Csörgő et Révész (1981) et Deheuvels (1974)), à faire un usage systématique de méthodes issues de la théorie des processus empiriques dans l'étude des propriétés asymptotiques d'estimateurs fonctionnels non paramétriques. Ses travaux ont concerné principalement les *estimateurs à noyau* (voir plus loin). Depuis cette époque, de nombreux auteurs, parmi lesquels nous citerons Deheuvels, J.H.J. Einmahl, U. Einmahl et Mason (voir, entre autres, Einmahl et Mason (2000), Deheuvels et Mason (2004), Einmahl et Mason (2005)), ont introduit des techniques nouvelles pour aborder ces problèmes, comme celles requérant l'usage de lois limites fonctionnelles, ou encore, mettant en œuvre des variantes locales de la théorie des processus

empiriques indexés par des ensembles ou par des fonctions. Dans cette perspective, les idées combinatoires de Vapnik et Červonenkis (Vapnik et Červonenkis (1971), Vapnik et Červonenkis (1981), Vapnik (1982)) ont conduit à des innovations majeures. Ces travaux ont permis d'établir de nouvelles propriétés des estimateurs fonctionnels classiques (concernant, par exemple, l'estimation de la densité ou de la régression). En particulier, des lois fonctionnelles uniformes du logarithme ont permis d'aborder de manière systématique la convergence en norme sup de tels estimateurs. Dans cette thèse, nous nous sommes particulièrement intéressés à l'extension de ces résultats à des cas n'ayant pas encore été traités dans la littérature scientifique. Dans cet esprit, le **chapitre 1** qui suit, généralisant les travaux de Deheuvels et Mason (1992), est consacré à l'étude du *processus des quantiles normé*, qui nous permet d'établir des propriétés nouvelles de l'estimateur de la densité basé sur les k -plus proches voisins. Dans le **chapitre 2** suivant, nous abordons le cas du *processus de Kaplan-Meier*, qui prolonge la notion de processus empirique standard aux *données censurées*. Nous aboutissons à des lois uniformes du logarithme pour certaines fonctionnelles de la loi de survie censurée. Ces résultats généralisent des travaux antérieurs de Deheuvels et Einmahl (1996), et de Deheuvels et Einmahl (2000). Le **chapitre 3** est consacré à l'étude du processus empirique local généralisé, indexé par des classes de fonctions. L'application naturelle de ces notions permet d'étudier les estimateurs à noyau (de type Parzen-Rosenblatt (Parzen (1962), Rosenblatt (1956)) pour la densité, et les estimateurs de Nadaraya-Watson (Nadaraya (1964), Watson (1964)) pour la régression). Nous faisons usage ici de la méthodologie introduite dans les articles de Einmahl et Mason (2000), Einmahl et Mason (2005), Deheuvels et Mason (2004), et Mason (2004). Dans la suite de ces idées, nous présentons, dans le **chapitre 4**, des résultats originaux portant sur un estimateur non paramétrique de la régression en données censurées introduit par Carbonez *et al.* (1995), et généralisé par Kohler *et al.* (2002). Le **chapitre 5** est, quant à lui, consacré à l'étude d'un estimateur de la fonction de régression censurée, lorsque celle-ci admet une représentation additive. Pour construire cet estimateur, nous faisons usage d'une méthode d'intégration marginale (Newey (1994), Tjøstheim et Auestad (1994)), en association avec l'estimateur défini au **chapitre 4**.

Dans la suite du présent chapitre introductif, nous définissons les principaux objets mathématiques utilisés tout au long de ce mémoire. Nous présentons également un bref résumé de quelques-uns parmi les résultats, qui seront exposés, et démontrés, dans les chapitres suivants.

0.1 Le processus empirique uniforme

Soit U_1, U_2, \dots des variables *indépendantes et idem-distribuées* [i.i.d.] de loi uniforme sur $(0, 1)$. Pour tout entier $n \geq 1$, notons $U_n(t) := n^{-1} \sharp \{U_i \leq t, 1 \leq i \leq n\}$, pour $t \in \mathbb{R}$, la version continue à droite de la fonction de répartition empirique uniforme associée à $\{U_1, \dots, U_n\}$. Ici, $\sharp A$ désigne la cardinalité de A . Pour $n \geq 1$ et $t \in \mathbb{R}$, nous définissons le processus empirique uniforme basé sur l'échantillon U_1, \dots, U_n par

$$\alpha_n(t) := n^{1/2} (U_n(t) - t). \quad (1)$$

Par suite, nous notons $\xi_n(h, t; s)$, pour $0 \leq t \leq 1$, $-1 \leq s \leq 1$ et $0 < h \leq t \wedge (1 - t)$, les incréments du processus empirique uniforme, définis par

$$\xi_n(h, t; s) := \alpha_n(t + hs) - \alpha_n(t). \quad (2)$$

D'autre part, nous désignons par

$$V_n(t) = \begin{cases} \inf \{u \geq 0 : U_n(u) \geq t\} & \text{pour } 0 < t \leq 1, \\ 0 & \text{pour } t \leq 0, \\ V_n(1) & \text{pour } t > 1, \end{cases}$$

la version continue à gauche de la fonction de quantile empirique uniforme. Le processus empirique des quantiles uniforme est alors défini comme suit. Nous posons, pour $n \geq 1$ et $t \in \mathbb{R}$,

$$\beta_n(t) := n^{1/2} (V_n(t) - t). \quad (3)$$

De manière analogue, pour $0 \leq t \leq 1$, $-1 \leq s \leq 1$ et $0 < h < t \wedge (1 - t)$, les incréments du processus empirique de quantiles uniforme sont définis par

$$\zeta_n(h, t; s) := \beta_n(t + hs) - \beta_n(t). \quad (4)$$

Le processus empirique uniforme, le processus empirique des quantiles uniforme, ainsi que leurs incréments respectifs, ont fait l'objet de recherches approfondies (voir par exemple les livres de Csörgő et Révész (1981), Csörgő (1983) et Shorack et Wellner (1986)). Ceci est justifié, tout particulièrement, par l'intérêt que ces processus présentent d'un point de vue statistique. En effet, de nombreuses propriétés asymptotiques d'estimateurs non paramétriques classiques (par exemple : les estimateurs à noyau, les estimateurs dits des k -plus proches voisins, ou encore les estimateurs par projection tronquée (voir notamment Bosq et Lecoutre (1987) et Scott (1992)), mais aussi de L -Statistiques (Hajek *et al.* (1999))), peuvent être déduites de propriétés limites établies pour l'un ou l'autre de ces quatre processus aléatoires. D'autres applications statistiques relatives aux processus empiriques sont détaillées dans l'ouvrage de van der Vaart et Wellner (1996).

Dans l'étude des incréments du processus empirique, la taille de la *fenêtre* h joue, comme on peut s'y attendre, un rôle central, et influe notamment très largement sur les vitesses de convergence dans les lois limites (voir par exemple les travaux successifs de Deheuvels et Mason (1990), Deheuvels et Mason (1991), Deheuvels et Mason (1992), ainsi que ceux de Deheuvels et Mason (1994) et Deheuvels et Mason (1995)). Dans ce qui suit, nous travaillerons principalement sur des incréments $h = h_n$, vérifiant les conditions asymptotiques (H.1-2-3), données ci-dessous (ces conditions sont souvent dites de Csörgő-Révész-Stute (Csörgő et Révész (1981), Stute (1982b)) ou [CRS]). Soit $\{h_n\}_{n \geq 1}$ une suite de réels positifs. Nous supposons que

$$(H.1) \quad h_n \downarrow 0 \text{ and } nh_n \uparrow \infty \text{ lorsque } n \uparrow \infty;$$

$$(H.2) \quad \log(1/h_n)/\log \log n \rightarrow \infty \text{ lorsque } n \rightarrow \infty;$$

$$(H.3) \quad nh_n/\log n \rightarrow \infty \text{ lorsque } n \rightarrow \infty.$$

Pour l'exposé des résultats du **chapitre 1**, il sera commode d'introduire les notions et notations additionnelles suivantes. Pour un intervalle $(a, b) \subset \mathbb{R}$ spécifié, nous désignons par $AC(a, b)$ l'ensemble des fonctions absolument continues sur $[a, b]$, et par $\mathcal{B}(a, b)$ l'ensemble des fonctions bornées sur $[a, b]$. La *norme sup* correspondante sera notée $\|f\| := \sup_{a \leq s \leq b} |f(s)|$, pour tout $f \in \mathcal{B}(a, b)$. Pour chaque ensemble non vide de fonctions $\mathcal{G} \subset \mathcal{B}(a, b)$, et pour tout $\varepsilon > 0$, nous définissons l' ε -voisinage uniforme \mathcal{G}^ε de \mathcal{G} par

$$\mathcal{G}^\varepsilon := \left\{ g \in \mathcal{B}(a, b) : \inf_{f \in \mathcal{G}} \|g - f\| < \varepsilon \right\}. \quad (5)$$

Soit \mathbb{S}_0 l'ensemble de Strassen (cf. Strassen (1964)), i.e., la boule unité de l'espace de Hilbert auto-reproduisant associé au processus de Wiener standard sur $[-1, 1]$, soit $\{W(u) : u \in [-1, 1]\}$. Ce dernier processus est commodément défini par

$$W(u) = \begin{cases} W_1(u) & \text{pour } u \geq 0, \\ W_2(-u) & \text{pour } u < 0, \end{cases}$$

où $\{W_1(t) : t \geq 0\}$ et $\{W_2(t) : t \geq 0\}$ désignent deux processus de Wiener standards indépendants. En d'autres termes,

$$\mathbb{S}_0 := \left\{ f \in AC(-1, 1), f(0) = 0 \text{ et } \int_{-1}^1 \dot{f}(s)^2 ds \leq 1 \right\}, \quad (6)$$

où $\dot{f} = \frac{d}{ds}f$ désigne la dérivée de Lebesgue de f .

Rappelons maintenant la notion de *relative compacité* pour les suites de parties d'un espace métrique \mathcal{E} , muni d'une distance d . Nous adoptons la notation suivante. Pour $A \subseteq \mathcal{E}$ et $x \in \mathcal{E}$, nous posons

$$d(x, A) = \inf_{y \in A} d(x, y).$$

Définition 0.1.1. Soit (\mathcal{E}, d) un espace métrique, et \mathcal{E}_1 une partie compacte de (\mathcal{E}, d) . Soit $\{\mathcal{U}_n\}_{n \geq 1}$ une suite de parties bornées non vides de \mathcal{E} . On dit que la suite $\{\mathcal{U}_n\}_{n \geq 1}$ est *relativement compacte d'ensemble d'adhérence \mathcal{E}_1* , lorsque les deux conditions suivantes sont vérifiées.

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{u \in \mathcal{U}_n} d(u, \mathcal{E}_1) &= 0, \\ \forall x \in \mathcal{E}_1, \quad \lim_{n \rightarrow \infty} \inf d(x, \mathcal{U}_n) &= 0. \end{aligned}$$

Dans le **chapitre 1** qui suit, nous faisons usage d'un résultat de Varron (2005) portant sur le processus ξ_n (ce dernier résultat se trouve être une extension naturelle des théorèmes de Deheuvels et Mason (1992)), ainsi que d'une version uniforme en h de la représentation de Bahadur locale (voir Bahadur (1966) et Deheuvels et Mason (1992)). Ceci nous permet d'établir une variante de ces résultats pour le processus ζ_n . De manière explicite, nous définissons, pour tout $n \geq 1$, l'ensemble de fonctions

$$\tilde{\zeta}_n := \left\{ \frac{\zeta_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}}, h \leq t \leq 1 - h, h'_n \leq h \leq h''_n \right\},$$

où $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ sont deux suites de constantes telles que $0 < h'_n \leq h''_n < \infty$, et vérifient, chacune, les conditions (H.1, 2, 3). Nous obtenons alors le théorème suivant.

Théorème 0.1.1. Soient $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$, deux suites de constantes réelles, vérifiant $0 < h'_n \leq h''_n < \infty$, ainsi que les conditions (H.1-2-3). Supposons de plus que l'une au moins des deux conditions (H.4)(i)-(ii) ci-dessous est vérifiée. On pose, pour $x \in \mathbb{R}^+$, $\log_2 x := \log_+ \log_+ x$, et $\log_+ x := \log(x \vee e)$.

$$(H.4)(i) \quad \frac{\sqrt{n} h'_n \log(1/h'_n)}{\log n \sqrt{\log_2 n}} \rightarrow \infty \text{ lorsque } n \rightarrow \infty;$$

$$(H.4)(ii) \quad \frac{\sqrt{h''_n \log(1/h''_n)}}{h'_n \log(1/h'_n)} = o\left(\frac{\sqrt{n}}{\log n}\right) \text{ lorsque } n \rightarrow \infty.$$

Alors, la suite $\{\tilde{\zeta}_n, n \geq 1\}$ est presque sûrement relativement compacte dans $\mathcal{B}[-1, 1]$, avec comme ensemble d'adhérence l'ensemble de Strassen \mathbb{S}_0 .

Nous établissons également dans le **chapitre 1** un résultat similaire à celui du théorème 0.1.1 dans le cas non uniforme, après avoir introduit des définitions adaptées pour le *processus empirique des quantiles normé*. En corollaire, nous en déduisons une loi du logarithme uniforme pour les *espacements* d'un échantillon. Ce dernier résultat nous permet de déduire directement les propriétés limites des estimateurs du type des *k-plus proches voisins* des lois correspondantes obtenues pour les estimateurs à noyau « classiques » (le Théorème 1.3.3 du **chapitre 1** fournit un exemple de cette méthodologie).

Remarquons que le résultat, mentionné plus haut, de Varron (2005), portant sur le processus ξ_n , nous permettra de traiter, selon un procédé analogue, le cas des incréments du *processus empirique de Kaplan-Meier* (voir le §0.4.3 qui suit pour les définitions appropriées), de façon à obtenir des propriétés asymptotiques des estimateurs non-paramétriques de certaines fonctionnelles de la loi de survie censurée.

Un autre « type » de processus empirique nous sera très utile pour l'étude des estimateurs à noyau. Il s'agit du processus empirique généralisé, introduit ci-dessous.

0.2 Le processus empirique généralisé

Par rapport au cas univarié, où les observations prennent leurs valeurs dans \mathbb{R} , l'étude des processus empiriques devient plus délicate dans \mathbb{R}^d , $d \geq 1$, par exemple, en raison de problèmes liés à la transformation de quantiles. Le cadre théorique général permettant l'étude de variables aléatoires i.i.d. à valeurs dans un espace mesurable plus général que \mathbb{R} (comme, par exemple, \mathbb{R}^d) repose sur la notion de *processus empirique généralisé*. Nous présentons la définition correspondante dans le cas multidimensionnel, en nous limitant au cadre strict des applications que nous allons développer.

Soit $(\mathbf{X}_i, \mathbf{Z}_i)_{i \geq 1}$, une suite de couples i.i.d. de variables aléatoires, à valeurs dans \mathbb{R}^d et \mathbb{R}^q , respectivement, avec $q, d \geq 1$. Soit par ailleurs \mathcal{G} , un ensemble de fonctions boréliennes réelles, définies sur \mathbb{R}^{d+q} . Pour tout $n \geq 1$, définissons le processus empirique d'ordre n , associé à $(\mathbf{X}_i, \mathbf{Z}_i)_{1 \leq i \leq n}$, et indexé par \mathcal{G} , par

$$\Upsilon_n(g) := n^{-1/2} \sum_{i=1}^n \left\{ g(\mathbf{X}_i, \mathbf{Z}_i) - \mathbb{E}(g(\mathbf{X}_i, \mathbf{Z}_i)) \right\}, \quad \text{pour } g \in \mathcal{G}. \quad (7)$$

Il est maintenant bien connu (voir par exemple van der Vaart et Wellner (1996) ou Dudley (1999)) que pour des ensembles de fonctions \mathcal{G} convenables, les théorèmes de type Glivenko-Cantelli [resp. Donsker] admettent des généralisations, par rapport au « cas trivial » des classes d'indicatrices de produits d'intervalles. Les ensembles \mathcal{G} correspondants sont dits former des classes de *Glivenko-Cantelli* [*GC*] (resp., de *Donsker*). Dans le cas du théorème de Glivenko-Cantelli, on exprime ainsi, par définition, que

$$n^{-1/2} \sup_{g \in \mathcal{G}} |\Upsilon_n(g)| \rightarrow 0,$$

presque sûrement lorsque $n \rightarrow \infty$, pourvu que \mathcal{G} soit une classe *GC* (dans le cas des classes de Donsker, on obtient une convergence vers un processus gaussien indexé par \mathcal{G}). Parmi les classes *GC*, on retrouve les classes de fonctions présentant un nombre de recouvrement polynomial, et, plus particulièrement, les classes dites de *Vapnik-Červonenkis* [*VC*]. Les notions de *nombre de recouvrement*, et de classes *VC* sont rappelées en

annexe. On consultera également Vapnik et Červonenkis (1971), Vapnik et Červonenkis (1981), Vapnik (1982), Pollard (1984), van der Vaart et Wellner (1996) et Dudley (1999).

Une question naturelle qui se pose au vu des développements précédents est la suivante. « Etant donnée une classe \mathcal{G} à nombre de recouvrement polynomial, quelle est la vitesse de convergence presque sûre de $n^{-1/2} \sup_{g \in \mathcal{G}} |\Upsilon_n(g)|$ vers 0 ? » Des réponses appropriées à cette question, dans des perspectives d'application à la statistique non paramétrique, ont été apportées dans les travaux successifs de Einmahl et Mason (2000), Deheuvels et Mason (2004), Einmahl et Mason (2005), ainsi que Varron (2005), pour des familles $\mathcal{G} = \mathcal{G}_n$ de fonctions appropriées, dépendant de $n \geq 1$, et présentant toutes la propriété de posséder un nombre de recouvrement polynomial. Dans les situations les plus favorables, il est également possible d'obtenir la constante associée à cette vitesse de convergence, ce qui permet la construction de bandes de confiance asymptotiques, simultanées et presque sûres, pour $\mathbb{E}(g(\mathbf{X}, \mathbf{Z}))$, lorsque g varie dans \mathcal{G} .

Comme conséquence de l'étude du processus $\{\Upsilon_n(g) : g \in \mathcal{G}\}$, pour des choix appropriés de $\mathcal{G} = \mathcal{G}_n$, Einmahl et Mason (2000), Einmahl et Mason (2005), Deheuvels et Mason (2004), Mason (2004) (voir également Varron (2005)), ont obtenu des résultats portant sur la consistance uniforme d'estimateurs non paramétriques de la densité $f_{\mathbf{X}}$ de \mathbf{X} , ainsi que d'estimateurs de la fonction de régression généralisée, $\mathbb{E}(\psi(\mathbf{Z}) \mid \mathbf{X} = x)$, de $\psi(\mathbf{Z})$ sachant $\mathbf{X} = x$. Ici, ψ désigne une fonction spécifiée, mesurable à valeurs réelles, vérifiant des hypothèses additionnelles convenables.

Dans le paragraphe suivant, nous discutons le cas de ces estimateurs non paramétriques. Nous montrons notamment que, pour un choix convenable des classes \mathcal{G} de fonctions, la vitesse de convergence asymptotique de (7) est, essentiellement, caractérisée par la variance des estimateurs.

0.3 Les estimateurs non paramétriques à noyau

Soit $(\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2), \dots$, une suite i.i.d. de répliques aléatoires indépendantes d'un *couple aléatoire générique* $(\mathbf{X}, Z) \in \mathbb{R}^d \times \mathbb{R}$, $d \geq 1$. Le couple (\mathbf{X}, Z) de variables aléatoires est supposé admettre une densité jointe $f_{\mathbf{X}, Z}$ sur $\mathbb{R}^d \times \mathbb{R}$ ainsi qu'une densité marginale (densité du vecteur \mathbf{X}) notée $f_{\mathbf{X}}$. Soit ψ une fonction mesurable, telle que la variable aléatoire $\psi(Z)$ vérifie $\mathbb{E}[|\psi(Z)|] < \infty$, ainsi que d'autres conditions qui seront spécifiées plus loin. Nous pouvons définir (sous réserve de conditions de régularité justifiant la validité de (8) ci-dessous) la fonction de régression de $\psi(Z)$ sachant \mathbf{X} par

$$m_\psi(\mathbf{x}) := \mathbb{E}(\psi(Z) \mid \mathbf{X} = \mathbf{x}) = \frac{\int_{\mathbb{R}} \psi(z) f_{\mathbf{X}, Z}(\mathbf{x}, z) dz}{\int_{\mathbb{R}} f_{\mathbf{X}, Z}(\mathbf{x}, z) dz}. \quad (8)$$

Remarquons que le choix $\psi = \mathbb{I}_{[-\infty, t]}$, $t \in \mathbb{R}$, dans (8), conduit à remplacer $m_\psi(\mathbf{x})$ par la fonction de répartition conditionnelle,

$$\mathbb{P}(\psi(Z) \leq t \mid \mathbf{X} = \mathbf{x}). \quad (9)$$

Ainsi, par des choix convenables de ψ , il est possible d'inclure dans le modèle (8) la fonction de répartition conditionnelle (9), sous la condition d'existence de la densité jointe $f_{\mathbf{X}, Z}$ (alors que le couple $(\mathbf{X}, \mathbb{I}_{\{Z \leq t\}})$, $t \in \mathbb{R}$, n'a pas nécessairement de densité à proprement parler).

Pour un choix de fenêtre $h > 0$ approprié et un *noyau* K défini sur \mathbb{R}^d , c'est à dire une fonction mesurable telle que

$$\int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1, \quad (10)$$

nous considérons les estimateurs à noyau de Akaike-Parzen-Rosenblatt (Akaike (1954), Rosenblatt (1956), Parzen (1962)) et de Nadaraya-Watson (Nadaraya (1964), Watson (1964)). Ces estimateurs de $f_{\mathbf{X}}(\mathbf{x})$, et de $m_{\psi}(\mathbf{x})$, sont définis, respectivement, par

$$f_{\mathbf{X};n}(\mathbf{x}; h) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right), \quad (11)$$

$$m_{\psi;n}(\mathbf{x}; h) = \begin{cases} \frac{\sum_{i=1}^n \psi(Z_i) K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)} =: \frac{r_{\psi;n}(\mathbf{x}; h)}{f_{\mathbf{X};n}(\mathbf{x}; h)}, & \text{pour } f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n \psi(Z_i), & \text{pour } f_{\mathbf{X};n}(\mathbf{x}; h) = 0. \end{cases} \quad (12)$$

La convergence de ces estimateurs vers les « vraies » fonctions, $f_{\mathbf{X}}$ et m_{ψ} , a été étudiée de manière intensive, au cours des vingt dernières années. Nous renvoyons, tout particulièrement, à Collomb (1981), Prakasa Rao (1983), Devroye et Györfi (1985), Silverman (1986), Bosq et Lecoutre (1987), Nadaraya (1989), Roussas (1990), Härdle (1990), Scott (1992), Wand et Jones (1995), Eggermont et La Riccia (2001) et Györfi *et al.* (2002), pour des détails et références additionnels. Nous rappelons ci-dessous un résultat classique ayant trait à la convergence de ces estimateurs, qui nous permettra d'évoquer le problème bien connu du *fléau de la dimension* (en anglais *curse of dimensionality*). Notons au passage que les conditions introduites dans la proposition 0.3.1 qui suit sont suffisantes, mais non nécessaires, pour la convergence des estimateurs. En préalable, rappelons une définition. Un noyau K défini sur \mathbb{R}^d est dit d'ordre ℓ , avec $\ell \geq 1$ entier, si

$$\begin{aligned} \int_{\mathbb{R}^d} K(\mathbf{u})d\mathbf{u} &= 1, \\ \int_{\mathbb{R}^d} u_1^{j_1} \dots u_d^{j_d} K(\mathbf{u})d\mathbf{u} &= 0, \quad j_1, \dots, j_d \geq 0, \quad j_1 + \dots + j_d = 0, \dots, \ell - 1, \\ \int_{\mathbb{R}^d} |u_1^{j_1} \dots u_d^{j_d}| K(\mathbf{u})d\mathbf{u} &< \infty, \quad j_1, \dots, j_d \geq 0, \quad j_1 + \dots + j_d = \ell. \end{aligned}$$

Proposition 0.3.1. *Soient $\{h_n\}_{n \geq 1}$ une suite de constantes positives telle que $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, et K un noyau d'ordre deux, borné et à support compact, défini sur \mathbb{R}^d . Supposons que les fonctions $f_{\mathbf{X},Z}$ et $f_{\mathbf{X}}$ sont trois fois continûment différentiables, et positives. Alors, pour toute fonction ψ mesurable bornée,*

$$|\mathbb{E}m_{\psi;n}(\mathbf{x}; h_n) - m_{\psi}(\mathbf{x})| = \mathcal{O}(h_n^2), \quad (13)$$

$$\mathbb{E}|m_{\psi;n}(\mathbf{x}; h_n) - \mathbb{E}m_{\psi;n}(\mathbf{x}; h_n)|^2 = \mathcal{O}\left(\frac{1}{nh_n^d}\right). \quad (14)$$

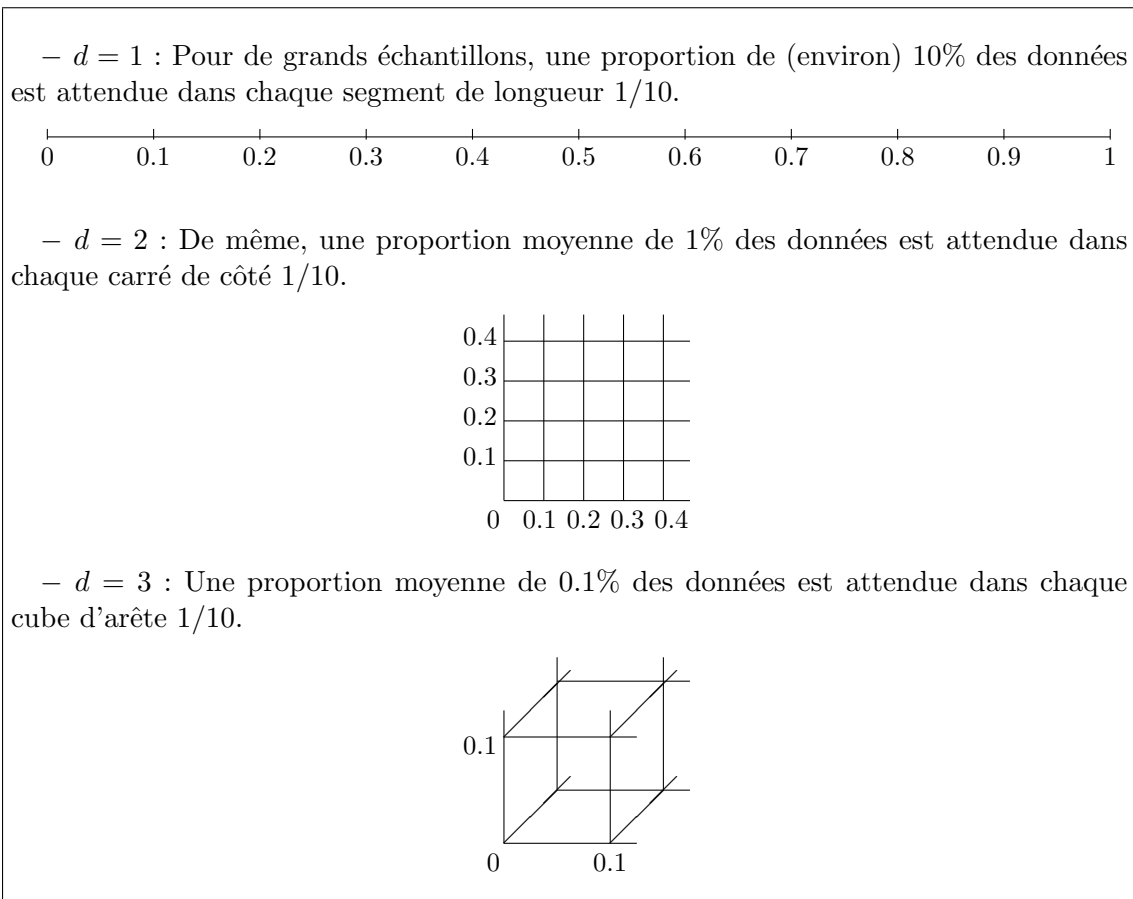
Un bon choix de $\{h_n\}_{n \geq 1}$ est primordial pour optimiser la vitesse de convergence asymptotique de $m_{\psi;n}$ vers m_{ψ} . La vitesse optimale requiert un équilibre entre les ordres de grandeur du terme de biais (13) et du terme de variance (14). Dans le cas de la convergence au sens de \mathbb{L}_2 , et, en vertu du résultat de la Proposition 0.3.1, un choix de $\{h_n\}_{n \geq 1}$ assurant un ordre de grandeur optimal de la convergence vers 0 de la norme \mathbb{L}_2 de $|m_{\psi;n}(\mathbf{x}; h_n) - m_{\psi}(\mathbf{x})|$

est tel que $h_n \sim n^{-1/(4+d)}$. On note ici $u_n \sim v_n$ pour exprimer que u_n/v_n et v_n/u_n sont tous deux $\mathcal{O}(1)$ lorsque $n \rightarrow \infty$. La vitesse de convergence de $m_{\psi;n}$ vers m_ψ , pour la norme \mathbb{L}_2 , est, lorsque $h_n \sim n^{-1/(4+d)}$, en

$$\mathcal{O}(n^{-4/(4+d)}). \quad (15)$$

Pour $d = 1$, cet ordre de convergence est en $\mathcal{O}(n^{-4/5}) \gg n^{-1}$. Le symbole $u_n \gg v_n$ étant utilisé ici pour exprimer que $u_n/v_n \rightarrow \infty$. On notera au passage que $\mathcal{O}(n^{-1})$ correspond à la vitesse « idéale » (ne pouvant être améliorée) de convergence dans \mathbb{L}_2 de l'estimateur paramétrique d'une fonction de régression linéaire (sous l'hypothèse que la régression appartient bien à ce modèle).

Un fait crucial est que la vitesse en (15) est une fonction décroissante de d , la dimension des covariables. Cette propriété est connue sous le nom de *fléau de la dimension* et reflète l'éparpillement relatif des données dans des espaces de dimension élevée. Pour mettre ce phénomène encore mieux en évidence, prenons l'exemple simple où la covariable \mathbf{X} est uniformément distribuée sur $[0, 1]^d$, avec $d = 1, 2, 3$.



Dans le cas général, où $d \geq 1$ est quelconque, considérons le problème de l'estimation de $m_\psi(\mathbf{x}_0)$, pour un point $\mathbf{x}_0 \in \mathbb{R}^d$ fixé. Le phénomène d'éparpillement évoqué ci-dessus se traduit par le fait que, pour un noyau K à support compact et à h constant, le nombre d'observations rentrant en compte dans l'estimation $m_{\psi;n}(\mathbf{x}_0, h)$ de $m_\psi(\mathbf{x}_0)$ se trouve être une fonction décroissante de la dimension d (ceci est une conséquence facile de la relation (12)). Comme l'estimateur $m_{\psi;n}$ se rattache à la classe générale des estimateurs à moyenne locale, cette propriété a comme conséquence naturelle la croissance du terme de variance lorsque d augmente (voir (14)). Une solution pour contourner le fléau de la dimension

est de travailler, si possible, sous des hypothèses de *modèles additifs* (voir le paragraphe correspondant dans la suite de notre exposé).

Dans le **chapitre 3**, nous présenterons des *lois uniformes du logarithme* pour les estimateurs $f_{\mathbf{X};n}$ et $m_{\psi;n}$ définis en (11) et (12). Ces résultats feront, tout naturellement, intervenir la variable de positionnement et la taille de la fenêtre. Nous emprunterons dans une large mesure les idées développées dans les articles de Einmahl et Mason (2000), Einmahl et Mason (2005), Deheuvels et Mason (2004), elles-mêmes suivies par Varron (2005). Comme ces auteurs, nous attacherons une importance primordiale à l'étude de la vitesse de convergence uniforme du « terme aléatoire » obtenu par les estimateurs centrés (par des facteurs n'étant pas nécessairement égaux à l'espérance de l'estimateur). En effet, il peut être montré que, pour les tailles de la fenêtre h les plus appropriées aux applications statistiques (vérifiant les conditions [CRS] avec un ordre de convergence convenable vers 0), et sous des conditions générales de régularité, le « terme de biais » (i.e., le facteur de centrage moins la fonction à estimer) est, bien souvent, négligeable (en norme uniforme) par rapport au « terme aléatoire ». On consultera, à ce sujet, l'étude détaillée de Deheuvels et Mason (2004).

Concrètement, en comparant (7) et (11) [resp. (12)], on constate que le choix de la classe de fonctions \mathcal{G} , comprenant les fonctions g définies par

$$g(\mathbf{X}, Z) = (nh^d)^{-1}K((\mathbf{x} - \mathbf{X}_i)/h),$$

[resp. $g(\mathbf{X}, Z) = (nh^d)^{-1}\psi(Z)K((\mathbf{x} - \mathbf{X}_i)/h)$] permet de traiter le terme aléatoire

$$f_{\mathbf{X};n}(\mathbf{x}; h) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h),$$

[resp. $r_{\psi;n}(\mathbf{x}; h) - \mathbb{E}r_{\psi;n}(\mathbf{x}; h)$], à partir de l'étude de Υ_n . Dans cet esprit, nous établirons dans le **chapitre 3**, le théorème 0.3.1 suivant. Rappelons que, par définition, un noyau K est toujours d'intégrale égale à 1, vérifiant (10). Ce noyau est dit de carré intégrable, si

$$\int_{\mathbb{R}^d} K(\mathbf{u})^2 d\mathbf{u} < \infty.$$

Un *pavé* de \mathbb{R}^d sera, par convention, un produit I d'intervalles non vides, de la forme

$$I = \prod_{j=1}^d [a_j, b_j].$$

Désignant par $\|\mathbf{u}\|_{\mathbb{R}^d}$ la norme euclidienne de $\mathbf{u} \in \mathbb{R}^d$, on pose, pour tout $\alpha > 0$,

$$I^\alpha = \{\mathbf{x} \in \mathbb{R}^d : \inf_{\mathbf{y} \in I} \|\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^d} < \alpha\}.$$

Nous rappelons également qu'un sous-ensemble \mathcal{E}_0 d'un ensemble topologique \mathcal{E} est dit relativement compact si sa fermeture est compacte. Nous renvoyons enfin à l'Appendice du **chapitre 3** et l'annexe de ce mémoire pour les définitions de classes de fonctions séparables point par point, et de type Vapnik-Červonenkis [VC].

Théorème 0.3.1. *Soit K un noyau borné à support compact et de carré intégrable. Supposons K de la forme $K(\mathbf{u}) = \phi(P(\mathbf{u}))$, où P est un polynôme des d variables composantes de \mathbf{u} , et ϕ une fonction réelle, à variable réelle et à variation bornée sur \mathbb{R} . Soient deux*

suites $\{h'_n\}_{n \geq 1}$, $\{h''_n\}_{n \geq 1}$ vérifiant (H.1-2-3), avec $h'_n \leq h''_n$, et soit $\alpha > 0$ une constante positive. Supposons qu'il existe un pavé I de \mathbb{R}^d tel que

- (i) $f_{\mathbf{X}}$ est strictement positive sur I^α ;
- (ii) $\forall \mathbf{x} \in I^\alpha, \lim_{\substack{\mathbf{x}' \rightarrow \mathbf{x} \\ \mathbf{x}' \in I^\alpha}} f_{\mathbf{X},Z}(\mathbf{x}', z) = f_{\mathbf{X},Z}(\mathbf{x}, z)$ pour presque tout $z \in \mathbb{R}$.

Soit enfin \mathcal{F} une classe de fonctions mesurables réelles et bornées. Si \mathcal{F} est séparable point par point et de type VC, et si la classe de fonctions $\{m_\psi/f_{\mathbf{X}} : \psi \in \mathcal{F}\}$ est relativement compacte par rapport à la topologie engendrée par la norme uniforme sur I^α , alors on a presque sûrement,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \{m_{\psi;n}(\mathbf{x}; h) - \widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h)\}}{\sqrt{2 \log(1/h^d)}} \\ &= \left\{ \int_{\mathbb{R}^d} K^2(t) dt \sup_{\mathbf{x} \in I} \frac{\sup_{\psi \in \mathcal{F}} \sigma_\psi^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right\}^{1/2}, \end{aligned} \quad (16)$$

$$\text{où } \sigma_\psi^2(\mathbf{x}) = \text{Var}(\psi(Z) \mid \mathbf{X} = \mathbf{x}) \quad \text{et} \quad \widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h) = \mathbb{E}r_{\psi;n}(\mathbf{x}; h)/\mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h). \quad (17)$$

Au cours de ce même **chapitre 3**, des résultats analogues à ceux du théorème 0.3.1 seront présentés dans le cas de l'estimation de la densité.

Remarque 0.3.1. Comme on pouvait s'y attendre, au vu des remarques précédentes, le résultat (16) illustre le fait que l'ordre de grandeur du « terme aléatoire » croît (ce qui implique la décroissance de la vitesse de convergence correspondante des estimateurs) avec la dimension d .

0.3.1 Les modèles additifs

Dans la littérature récente, plusieurs idées ont été introduites pour pallier le problème posé par le fléau de la dimension. Parmi celles-ci, on peut, tout d'abord, mentionner les méthodes générales ayant pour objet la réduction de la dimensionnalité des données (typiquement la *projection pursuit* de Friedman et Stuetzle (1981) ou encore Hall (1989)). En deuxième lieu il faut évoquer un ensemble de méthodes, introduisant des hypothèses sur la structure de la « vraie » fonction de régression m_ψ (par exemple, les *modèles additifs généralisés*, introduits par Winsberg et Ramsay (1980) et repris par Stone (1985), Hastie et Tibshirani (1986) et Hastie et Tibshirani (1990), rentrent dans cette catégorie).

C'est ce dernier type de modèle que nous étudierons dans le **chapitre 5**. Nous y considérerons des données, pour lesquelles la variable d'intérêt est censurée. Ces *modèles additifs généralisés* reposent sur l'hypothèse suivante, où ψ désigne une fonction mesurable réelle fixée. Nous supposons que

$$m_\psi(\mathbf{x}) = m_{add}(\mathbf{x}) := \mu + \sum_{\ell=1}^d m_\ell(x_\ell). \quad (18)$$

Dans la représentation (18), les fonctions m_ℓ , $\ell = 1, \dots, d$, sont, habituellement, appelées *composantes additives de la régression*. Un des avantages de ce modèle est que son interprétation est aisée, du fait que chaque variable concomitante agit « séparément » sur la variable d'intérêt Z .

Dans le modèle (18), les fonctions m_ℓ sont définies à une constante additive près. Cette propriété impose naturellement que les procédures d'analyse statistiques soient développées sous la condition d'identifiabilité

$$\mathbb{E}m_\ell(X_\ell) = 0, \ell = 1, \dots, d. \quad (19)$$

Cette dernière condition implique que $\mu = \mathbb{E}(\psi(Z))$.

Remarque 0.3.2. *Notons, au passage, que d'autres conditions d'identifiabilité que celle donnée dans (19) sont envisageables. Par exemple, la condition*

$$m_\ell(\mathbb{E}X_\ell) = 0, \ell = 1, \dots, d, \quad (20)$$

implique que le terme constant, μ , de (18), vérifie l'identité $\mu = \mathbb{E}(\psi(Z) | \mathbf{X} = \mathbb{E}\mathbf{X})$.

Un exemple intéressant d'application est obtenu lorsque Z représente le délai précédant l'apparition d'une maladie, et pour le choix de ψ donné par $\psi = \mathbb{I}_{[0,t]}$ pour un $t > 0$ fixé. Le problème de l'estimation de $m_\psi(\mathbf{x})$ est alors connu sous le nom de la construction du score de risque de maladie (voir, par exemple, Viallon et al. (2006b), et les références ci-incluses). Dans ce cadre, les interprétations de μ selon que la condition d'identifiabilité est (19) ou (20) sont les suivantes. Sous (19), le terme constant, μ , de (18), soit $\mu = \mathbb{E}(\mathbb{I}_{\{Z \leq t\}})$, représente le risque moyen de développer la maladie avant t . Sous (20), ce même terme constant est tel que $\mu = \mathbb{E}(\mathbb{I}_{\{Z \leq t\}} | \mathbf{X} = \mathbb{E}\mathbf{X})$, et représente le risque de développer la maladie avant t pour un « individu moyen ». L'intérêt de chacune de ces deux quantités dépend du cadre de l'étude, et la condition d'identifiabilité initiale est donc à choisir en fonction de celle-ci.

Certaines statistiques ont été proposées, dans la littérature scientifique récente, pour tester le caractère additif de m_ψ , et donc l'adéquation du modèle (18) aux données (voir Sperlich et al. (2002), Eubank et al. (1995) et Derbort et al. (2002)).

Méthodes d'estimation pour les composantes additives

Dans le cadre des modèles additifs, diverses méthodes ont été développées pour estimer les composantes additives m_ℓ , $1 \leq \ell \leq d$, de la représentation (18). Ces méthodes permettent de construire des estimateurs de la fonction de régression m_ψ . Nous présentons dans ce paragraphe deux de ces techniques. En premier, nous considérons la méthode d'*intégration marginale* (Newey (1994), Tjøstheim et Auestad (1994), Linton et Nielsen (1995)). En second, nous considérons la méthode reposant sur l'algorithme dit du *backfitting* (voir Hastie et Tibshirani (1990)).

La méthode d'intégration marginale Dans ce paragraphe, nous rappelons le principe de cette méthode d'estimation, que nous adapterons au cas des données censurées dans le **chapitre 5**. Nous supposons, tout d'abord, que $d \geq 2$ et que la fonction de régression m_ψ puisse s'écrire sous la forme additive donnée en (18). Nous pouvons alors reconstruire la fonction de régression additive m_ψ de la façon suivante. Soient q_1, \dots, q_d , d fonctions de densité réelles. En posant $q = \prod_{\ell=1}^d q_\ell$ et $q_{-\ell} = \prod_{j \neq \ell} q_j$, on peut écrire $m_\psi(\mathbf{x})$ sous la forme

$$m_\psi(\mathbf{x}) = \sum_{\ell=1}^d \eta_\ell(x_\ell) + \int_{\mathbb{R}^d} m_\psi(\mathbf{z})q(\mathbf{z})d\mathbf{z}, \quad (21)$$

ceci, en faisant usage des notations

$$\eta_\ell(x_\ell) := \int_{\mathbf{R}^{d-1}} m_\psi(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbf{R}^d} m_\psi(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad \ell = 1, \dots, d, \quad (22)$$

avec $\mathbf{x}_{-\ell} = (x_1, \dots, x_{\ell-1}, x_{\ell+1}, \dots, x_d) \in \mathbf{R}^{d-1}$. En effet, en intégrant la fonction de régression du modèle (18) par rapport à $q_{-\ell}$, il vient

$$\int_{\mathbf{R}^{d-1}} m_\psi(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} = \mu + m_\ell(x_\ell) + \sum_{j \neq \ell} \int_{\mathbf{R}} m_j(z_j) q_j(z_j) dz_j.$$

D'autre part, en intégrant m_ψ par rapport à q , on obtient

$$\int_{\mathbf{R}^d} m_\psi(\mathbf{x}) q(\mathbf{x}) d\mathbf{x} = \mu + \sum_{\ell=1}^d \int_{\mathbf{R}} m_\ell(z_\ell) q_\ell(z_\ell) dz_\ell.$$

Par une soustraction, membre à membre, de ces deux dernières équations, on obtient ensuite que

$$\int_{\mathbf{R}^{d-1}} m_\psi(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbf{R}^d} m_\psi(\mathbf{x}) q(\mathbf{x}) d\mathbf{x} = m_\ell(x_\ell) - \int_{\mathbf{R}} m_\ell(z_\ell) q_\ell(z_\ell) dz_\ell. \quad (23)$$

En combinant (22) et (23), on observe alors que

$$\eta_\ell(x_\ell) = m_\ell(x_\ell) - \int_{\mathbf{R}} m_\ell(z_\ell) q_\ell(z_\ell) dz_\ell, \quad (24)$$

ce qui permet d'aboutir à (21).

Remarque 0.3.3. *Il ressort de (24), que, pour chaque valeur de $\ell = 1, \dots, d$, les fonctions η_ℓ et m_ℓ sont égales, à une constante additive près. Les fonctions η_ℓ sont donc des composantes additives de m_ψ , au même titre que les m_ℓ . Ces composantes additives correspondent, en fait, à une condition d'identifiabilité différente de celle fournie par (19). Notons que le choix $q_\ell = f_\ell$, $\ell = 1, \dots, d$, avec f_ℓ qui désigne la fonction densité de la variable aléatoire réelle X_ℓ , entraîne $\eta_\ell = m_\ell$. En pratique, cependant, f_ℓ n'est généralement pas connue, et $\eta_\ell \neq m_\ell$.*

Soit $h > 0$ fixé. Les estimateurs des composantes de m_ψ , définies par (22), sont obtenus en intégrant l'estimateur de la fonction de régression multivariée défini par (12). On obtient ainsi

$$\hat{\eta}_{\ell,n}(x_\ell, h) := \int_{\mathbf{R}^{d-1}} \hat{m}_{\psi;n}(\mathbf{x}, h) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbf{R}^d} \hat{m}_{\psi;n}(\mathbf{x}, h) q(\mathbf{x}) d\mathbf{x}. \quad (25)$$

Un estimateur de la fonction de régression m_ψ , sous l'hypothèse additive, peut alors être construit, en posant

$$\hat{m}_{\psi,add}(\mathbf{x}, h) := \sum_{\ell=1}^d \hat{\eta}_{\ell,n}(x_\ell, h) + \int_{\mathbf{R}^d} \hat{m}_{\psi;n}(\mathbf{x}, h) q(\mathbf{x}) d\mathbf{x}. \quad (26)$$

Cette méthode a, notamment, été étudiée par Linton et Nielsen (1995). En faisant usage d'un estimateur *localement linéaire* (voir par exemple Fan et Gijbels (1996)), ces auteurs ont établi un théorème central limite pour l'*estimateur intégré* ainsi obtenu. Pour des fonctions de régression bivariées ($d = 2$), deux fois continûment différentiables, ils obtiennent

une vitesse de convergence en $\mathcal{O}(n^{-4/5})$, sous l'hypothèse additionnelle que la densité du vecteur de covariables est continûment différentiable. On notera que cette vitesse est du même ordre que celle des estimateurs non paramétriques de la régression *univariée*, obtenus sous des conditions de régularité analogues. On échappe ainsi au « fléau de la dimension ». On consultera Newey (1994), Tjøstheim et Auestad (1994), Chen *et al.* (1996), Linton *et al.* (1997) et Severance-Lossin et Sperlich (1999).

L'algorithme du backfitting La méthode communément implémentée dans les logiciels statistiques repose sur un algorithme itératif, connu sous le nom de « *backfitting* » (Hastie et Tibshirani (1990)). Ce dernier algorithme a été programmé dans plusieurs logiciels classiques, par exemple, dans la procédure *gam* de *SAS*, ou dans les « packages » *gam* ou *mgecv* de *R*. A chaque étape, d'indice (ou d'ordre) $p = 0, 1, \dots$, de l'algorithme, une régression univariée est effectuée, l'étape $p + 1$ consistant en une nouvelle régression univariée sur les résidus obtenus au cours de l'étape d'indice p . Le principe de l'algorithme peut ainsi être résumé de la manière suivante.

1. Dans l'étape d'ordre $p = 0$, on estime $\mu = E(Y)$ par \bar{Y}_n et pour $\ell = 1, \dots, d$, on pose $\hat{m}_{\ell,0}(X_{i,\ell}) = 0$.

2. Pour les étapes d'indice $p = 1, 2, \dots$ et jusqu'à convergence des $\hat{m}_{\ell_0,p}$, $\ell_0 = 1, \dots, d$, on pose pour $i = 1, \dots, n$, $r_{i,\ell_0} = Y_i - \bar{Y}_n - \sum_{\ell=1}^{\ell_0-1} \hat{m}_{\ell,p}(X_{i,\ell}) - \sum_{\ell=\ell_0+1}^d \hat{m}_{\ell,p-1}(X_{i,\ell})$. La quantité $\hat{m}_{\ell_0,p}(X_{i,\ell_0})$ est obtenue en effectuant la régression des r_{i,ℓ_0} par rapport aux X_{i,ℓ_0} .

A chaque étape de l'algorithme, il faut déterminer des paramètres de lissage optimaux. Deutsch (1983) a montré la convergence de cet algorithme vers l'unique approximation de m_{ψ} dans \mathbb{L}^2 . Cette méthode a aussi été étudiée par Breiman et Friedman (1985), et par Hastie et Tibshirani (1990) et Härdle et Hall (1993). Bien que la méthode du backfitting soit très utilisée en pratique, le fait que les estimateurs backfitting soient définis comme solutions d'un algorithme itératif a rendu difficile l'étude théorique de leurs propriétés statistiques. Citons, parmi les résultats disponibles, ceux d'Opsomer et Ruppert (1997), qui donnent, dans le cas d'un modèle additif bivarié, les expressions du biais et de la variance asymptotiques pour des estimateurs backfitting basés sur des polynômes locaux. Dans les travaux d'Opsomer (2000), on peut trouver une extension de ce résultat au cas où le modèle additif est obtenu par la somme de d fonctions univariées.

Severance-Lossin et Sperlich (1999) ont comparé, au moyen de simulations, la méthode d'intégration marginale à celle du backfitting. Il ressort de leurs travaux qu'aucune des deux méthodes ne surclasse l'autre, leurs performances respectives dépendant du modèle sous-jacent. En revanche, il apparaît que des effets de bord se font davantage ressentir dans le cas de l'estimateur intégré. De plus, les problèmes liés à la présence de covariables corrélées sont plus sérieux avec cette dernière méthode d'estimation. Une voie logique pour améliorer l'estimateur obtenu par la méthode d'intégration est alors de combiner cette méthode avec une itération de backfitting (Linton *et al.* (1997), Kim *et al.* (1998) et Sperlich *et al.* (2002)).

0.4 Le cas des données censurées

0.4.1 Définition

Dans de nombreuses applications statistiques, on est amené à faire intervenir une *variable de durée*, c'est-à-dire un délai séparant un instant initial de l'instant où un événement final

est observé. C'est tout particulièrement le cas pour les applications médicales liées à l'étude de *durées de survie*. Pour cette raison, et bien que le champ d'application des méthodes que nous allons étudier ne se limite nullement à la médecine ou à la biologie, nous utiliserons tout au long de ce mémoire la terminologie associées aux *durées de survie*. Celle-ci est la plus communément usitée dans la littérature statistique, puisque c'est précisément dans le domaine médical que les avancées méthodologiques liées à ces variables ont été développées en premier. Nous parlerons de *données censurées* lorsque la durée de survie n'est connue que lorsqu'elle est limitée par une durée limite d'observation. Par exemple, dans le cas dit de *censure à droite*, seul le minimum entre la *durée de survie* et une *durée limite supérieure d'observation* est connu, ainsi que l'indicateur exprimant que la durée de survie a été censurée ou non. Ce minimum constitue, en quelque sorte, une *durée de participation*, qui est observée en ne donnant qu'une information partielle sur la durée de survie. Il existe plusieurs autres catégories de modèles de censure, obtenus par des variations du principe, décrit ci-dessus, de la censure à droite. Parmi ceux-ci, mentionnons les suivants, en y incluant, en premier, le modèle de censure à droite.

- CENSURE À DROITE. Il y a censure à droite lorsque la *durée de survie* est supérieure ou égale à la *durée de participation*. Un exemple typique est celui où l'événement considéré est le décès du patient malade (ou la panne d'un appareil dans des études de fiabilité), et la durée d'observation est une durée totale d'hospitalisation. L'expérimentateur peut également fixer une date de fin d'expérience. Ainsi, pour toute observation sans « événement » observé avant cette date, on sait simplement que la durée de survie est supérieure à la durée de participation observée.
- CENSURE À GAUCHE. Il y a censure à gauche lorsque la *durée de survie* est inférieure ou égale à la *durée de participation*. Ce modèle est adapté au cas où l'incorporation d'un patient dans une étude est conditionnée par un *événement initial* (par exemple, par la date associée à l'apparition de certains symptômes). Il est alors fréquent de ne connaître qu'un minorant de la date de ce premier événement. La durée de survie, prise comme la différence entre le temps où se produit l'événement d'intérêt et le temps où s'est produit l'événement initial, est ici supérieure ou égale à la durée effectivement observée par le statisticien.
- CENSURE PAR INTERVALLE. Les deux phénomènes ci-dessus sont conjugués.

Les trois catégories de censure décrits ci-dessus peuvent être déclinés en fonction du mode de censure. On obtient les types suivants :

- CENSURE DE TYPE I. L'expérimentateur fixe une date (non aléatoire) de fin d'expérience. La durée de participation maximale est alors fixée (non aléatoire) et vaut, pour chaque observation, la différence entre la date de fin d'expérience, et la date d'entrée du patient dans l'étude. Le nombre d'événements observés est, quant à lui, aléatoire. Ce modèle est souvent utilisé dans les études épidémiologiques.
- CENSURE DE TYPE II. L'expérimentateur fixe a priori le nombre d'événements à observer. La date de fin d'expérience devient alors aléatoire, le nombre d'événements étant, quant à lui, non aléatoire. Ce modèle est souvent utilisé dans les études de fiabilité.

- CENSURE ALÉATOIRE. C'est typiquement ce modèle qui est utilisé pour les essais thérapeutiques. Dans ce type d'expérience, la date d'inclusion du patient dans l'étude est fixée, mais la date de fin d'observation est inconnue (celle-ci correspond, par exemple, à la durée d'hospitalisation du patient). Ici, le nombre d'événements observés et la durée totale de l'expérience sont aléatoires.

Dans notre mémoire, nous nous intéresserons uniquement au cas des censures à droite de type aléatoire. Celui-ci correspond à un modèle fréquemment utilisé dans la pratique, ce qui justifie amplement qu'on y attache de l'intérêt.

0.4.2 Quantités mathématiques invoquées

Nous introduisons maintenant les notations qui seront utilisées dans les chapitres traitant des données censurées. D'une manière générale, nous supposerons disposer de copies indépendantes et identiquement distribuées (Y_i, C_i, \mathbf{X}_i) , $i = 1, 2, \dots$, du vecteur aléatoire (Y, C, \mathbf{X}) où Y désigne la variable aléatoire d'intérêt. Celle-ci étant interprétée comme une durée de survie, il est naturel de supposer qu'elle ne prend que des valeurs positives, toutefois, nous ne ferons pas cette hypothèse en général. Ici, on désigne par C le temps (aléatoire) de censure, et par \mathbf{X} , un vecteur de variables aléatoires concomitantes, dont la structure de dépendance avec Y et C dépendra du contexte. Ainsi, nous travaillerons, successivement, *sans conditionnement* et *avec conditionnement* par rapport à \mathbf{X} . Plus explicitement, nous supposerons, dans certains cas, que la variable \mathbf{X} est indépendante des variables Y et C , et dans d'autres cas, nous nous affranchirons de cette hypothèse. Nous travaillerons sous l'hypothèse générale d'indépendance entre (Y, \mathbf{X}) et C . Dans le cas où la covariable \mathbf{X} est absente, nous supposerons l'indépendance entre Y et C . L'échantillon observé est alors $(Z_i, \delta_i, \mathbf{X}_i)$, $i = 1, \dots, n$, pour une taille d'échantillon $n \geq 1$, où, pour $i = 1, \dots, n$,

$$\begin{aligned} Z_i &:= Y_i \wedge C_i, \\ \delta_i &:= \mathbb{I}_{\{Y_i \leq C_i\}}, \end{aligned} \tag{27}$$

\mathbb{I}_E désignant la fonction indicatrice de E . Au vu de (27), Z_i s'interprète comme le *temps de participation* (ou le *temps de suivi*) du patient dans l'étude. L'indicatrice δ_i vaut 1 en l'absence de censure, lorsque $Z_i = Y_i$, et 0 autrement.

Remarque 0.4.1. *Dans la suite de cette introduction, afin de ne pas alourdir inutilement les notations, nous travaillerons sous les hypothèses du **chapitre 2**, correspondant à l'absence de la covariable \mathbf{X} . Nous verrons dans le **chapitre 4** que les quantités que nous allons maintenant introduire, dans ce contexte simplifié, peuvent être généralisées sans trop de difficulté dans le cas plus complexe où Y et C sont effectivement liées à un vecteur de covariables \mathbf{X} .*

Pour $t \in \mathbb{R}$, soient $F(t) := \mathbb{P}(Y \leq t)$ et $G(t) := \mathbb{P}(C \leq t)$ les (versions continues à droite des) fonctions de répartition de Y et C . L'hypothèse d'indépendance entre Y et C implique alors que $H(t) := \mathbb{P}(Z \leq t) = 1 - (1 - F(t))(1 - G(t))$. De nombreuses applications requièrent l'estimation de certaines *fonctionnelles locales* de F . Un exemple important en est donné par le *taux de mortalité* (ou *taux de hasard instantané*), λ , défini comme suit. On suppose que F est absolument continue sur \mathbb{R} , de densité $f(t) = dF(t)/dt$. Le taux de mortalité est alors défini par

$$\lambda(t) := \frac{f(t)}{1 - F(t)} \quad \text{pour } t \in \mathbb{R}. \tag{28}$$

0.4.3 L'estimateur de Kaplan-Meier

Un estimateur jouant un rôle essentiel dans le cadre des données censurées est l'estimateur du maximum de vraisemblance généralisée non paramétrique de $F(t)$, le « *product limit estimator* » (estimateur « *produit-limite* »), introduit par Kaplan et Meier (1958), et défini par

$$F_n^*(t) = 1 - \prod_{i:Z_i \leq t} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)} \right)^{\delta_i}, \text{ pour } t \in \mathbb{R}, \quad (29)$$

où $N_n(x) = \sum_{i=1}^n \mathbb{I}_{\{Z_i \geq x\}}$, et avec les conventions $\prod_{\emptyset} = 1$ et $0^0 = 1$. On notera que ces conventions attribuent à $F_n^*(t)$ la valeur 1 lorsque $t > \max\{Z_1, \dots, Z_n\}$. Par analogie au cas non censuré, on peut définir le processus empirique de Kaplan-Meier a_n^* comme suit : pour tout $n \geq 1$ et tout $t \in \mathbb{R}$, on pose

$$a_n^*(t) := n^{1/2}(F_n^*(t) - F(t)). \quad (30)$$

On peut ensuite, comme pour le processus empirique usuel, définir les incréments du processus empirique de Kaplan-Meier par

$$\eta_n^*(h, t; s) := a_n^*(t + hs) - a_n^*(t), \quad (31)$$

pour $h > 0$, $0 \leq s \leq 1$ et $t \in \mathbb{R}$. Deheuvels et Einmahl (1996) et Deheuvels et Einmahl (2000) ont obtenu des lois limites fonctionnelles associées à η_n^* . Nous établirons une extension de leurs résultats dans le **chapitre 2**. Dans ce chapitre, nous généralisons le résultat, obtenu par Varron (2005) pour le processus empirique usuel, aux incréments du processus de Kaplan-Meier η_n^* . Nous en déduisons des lois uniformes du logarithme pour des estimateurs de la *densité* et du *taux de mortalité* (ou *taux de hasard instantané*), dans le contexte des données censurées.

Il est souvent commode de faire usage de certaines variantes de la définition (29), adoptée ci-dessus pour l'estimateur de Kaplan-Meier. Par exemple, l'estimateur F_n^* est la solution d'un algorithme itératif de type *EM* (pour « *Expectation Maximisation* »). En tant que tel, il est connu sous le nom de « *Self Consistent Estimator* » (SCE) (voir notamment Efron (1967), Turnbull (1974) et Tsai et Crowley (1985)). D'autre part, il est possible de réécrire l'estimateur de Kaplan-Meier sous forme additive. Plus précisément, en définissant l'estimateur de Kaplan-Meier G_n^* de G (en remplaçant formellement dans (29) les δ_i par $1 - \delta_i$, pour $i = 1, \dots, n$) par

$$G_n^*(z) = 1 - \prod_{i:Z_i \leq z} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)} \right)^{1-\delta_i}, \quad (32)$$

on peut montrer l'identité

$$F_n^*(t) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{Z_i \leq t\}}}{1 - G_n^*(Z_i)}. \quad (33)$$

La validité de ces formules repose sur des notions que nous ne pouvons résumer brièvement dans le cadre de la présente introduction, et utilise principalement le fait que l'estimateur de Kaplan-Meier est la solution d'un certain système d'équations différentielles. Nous renvoyons pour plus de détails à l'annexe du **chapitre 4**, qui suit, ainsi qu'à l'ouvrage de van der Laan et Robins (2002) et à l'article de Satten et Datta (2001).

Remarque 0.4.2. *Les estimateurs de ce type sont connus dans la littérature anglo-saxonne sous le nom de « Inverse Probability of Censoring Weighted [IPCW] estimates ». Ils font partie de la catégorie des estimateurs du type « synthetic data », qui comprend également les estimateurs SCE (voir, à ce sujet, van der Laan et Robins (2002)).*

A partir de l'expression (33), Carbonez *et al.* (1995) puis Kohler *et al.* (2002) ont déduit une version conditionnelle de l'estimateur de Kaplan-Meier. Cette version représente une alternative à celle proposée par Beran (1981), et étudiée, notamment, par Dabrowska (1989) et Deheuveld et Derzko (2006). De plus, les idées de Carbonez et Kohler conduisent à l'obtention d'un estimateur pour la fonction de régression (en remplaçant formellement $\mathbb{I}_{\{Z_i \leq y\}}$ par Z_i – voire $\psi(Z_i)$ – dans (33)). Dans le **chapitre 4** figurent des résultats analogues à (16) pour cet estimateur. La variance obtenue est alors « dilatée », dans le sens où le terme $\sigma_\psi^2(\mathbf{x})$ défini en (17) est remplacé par $\mathbb{E}\{\psi^2(Y)/(1 - G(Y)) \mid \mathbf{X} = \mathbf{x}\} - m_\psi^2(\mathbf{x})$. Suivant les idées développées par Deheuveld et Mason (2004), nous dérivons des bandes de confiance asymptotiques simultanées de ces lois uniformes du logarithme. Des simulations viennent illustrer nos résultats.

Enfin, dans le **chapitre 5**, nous obtenons la vitesse de convergence presque sûre d'un estimateur de la fonction de régression, sous une hypothèse additive, et dans un contexte de données censurées. La construction de notre estimateur combine les méthodes d'estimation introduites au **chapitre 4** avec la méthode d'intégration marginale évoquée plus haut.

Chapitre 1

A uniform in bandwidth functional limit law for the increments of the normed sample quantile process - application to the Nearest-Neighbor empirical density function.

Foreword. This chapter is an adaptation of a paper submitted to *Journal of Nonparametric Statistics* (Viallon (2006a)), tailored to fit in the present manuscript.

Abstract. In the present chapter, we establish a functional limit law of the logarithm for the increments of the normed sample quantile process based upon a random sample of size $n \rightarrow \infty$. We extend a limit law obtained by Deheuvels et Mason (1992) for kernel-type functional estimators. Our results hold uniformly over the bandwidth h , restricted to vary in $[h'_n, h''_n]$, where $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ are appropriate non-random sequences. We treat the case where the sample observations follow possibly non-uniform distributions. As a consequence of our theorems, we provide uniform limit laws for nearest-neighbor density estimators, in the spirit of those given by Deheuvels et Mason (2004) for kernel-type estimators.

1.1 Introduction

1.1.1 Motivations

The theory of empirical processes has been extensively investigated over the past decades. Several authors (see Deheuvels (1974), Csörgő et Révész (1981), Stute (1982a)) have underlined its relevance to the study of kernel nonparametric functional estimators, such as the Parzen-Rosenblatt density function estimator (refer to Parzen (1962) and Rosenblatt (1956)), or the Nadaraya-Watson regression function estimator (see, e.g., Nadaraya (1964) and Watson (1964)). Lately, these results have been significantly extended by, especially, Einmahl et Mason (2000), Einmahl et Mason (2005), Mason (2004), Deheuvels et Mason (2004) and Varron (2005). By combining empirical process arguments with combinatorial techniques initiated by Vapnik and Červonenkis (Vapnik et Červonenkis (1971), Vapnik

(1982)), these authors have established various versions of the uniform consistency of kernel type estimators, the latter being also shown to hold uniformly over the bandwidth parameter, with some restrictions. A direct statistical application of these results is given by the construction of limiting *certainty bands* for the density and the regression functions (see, e.g., Deheuvels et Mason (2004)). A parallel field of study is that of empirical quantile processes, for which a series of functional limit laws have been provided by Deheuvels et Mason (1992) and Deheuvels (1992). The motivation of the present study is that the recent refinements of the functional limit laws, initiated by Deheuvels, J.Einmahl, U.Einmahl and Mason, have been mostly written in the framework of the usual empirical (distribution) processes, and cover only in part the case of empirical quantile processes. We will therefore orient our work towards bridging the remaining gaps in this theory. As a main result, we shall provide, in the sequel, a uniform in bandwidth functional limit law of the logarithm for the increments of the normed empirical quantile processes. This result will be then applied to establish a uniform in bandwidth law of the logarithm for Nearest-Neighbor density function estimators. We mention that our results should provide themselves useful in the construction of nonparametric *goodness-of-fit* tests (see, e.g., Theorem 1.3.2 below). Likewise, some applications to the study of Lorenz process and score function estimators, in the spirit of Csörgő (1983), may be derived from our results. We refer to Csörgő (1983) for other examples of applications of quantile processes, which have some relevance to our work.

1.1.2 Notation and Main Results

Let X_1, X_2, \dots be independent and identically distributed [iid] random variables with distribution function $F(x) := \mathbb{P}(X_1 \leq x)$, for $x \in \mathbb{R}$, and quantile function $Q(t) := \inf\{x \in \mathbb{R} : F(x) \geq t\}$, for $0 \leq t \leq 1$. For $x \in \mathbb{R}$, we denote by $F_n(x) := n^{-1} \# \{X_i \leq x, 1 \leq i \leq n\}$ the empirical distribution function based upon X_1, \dots, X_n and by $Q_n(t) := \inf\{x \in \mathbb{R} : F_n(x) \geq t\}$, for $0 < t < 1$, the corresponding empirical quantile function. Here $\#E$ denotes the cardinality of E . We denote by $a_n(x) := n^{1/2}(F_n(x) - F(x))$, for $x \in \mathbb{R}$, the empirical process of order $n \geq 1$, and we let (see, e.g., Csörgő (1983))

$$b_n(t) := n^{1/2}(Q_n(t) - Q(t))/q(t) \quad \text{for } 0 < t < 1, \quad (1.1)$$

denote the normed sample quantile process. Here,

$$q(t) := \frac{d}{dt} Q(t) = \frac{1}{f(Q(t))}, \quad (1.2)$$

stands for the quantile density function (see, e.g., Parzen (1979)), and

$$f(x) = \frac{d}{dx} F(x),$$

denotes the probability density function, both assumed to be properly defined and continuous in domains specified later on (see, e.g., the assumptions (F.1-2) below). In this chapter, we are concerned with limit laws for the increments $\vartheta_n(\cdot)$ of the normed sample quantile process $b_n(\cdot)$, which are defined as follows. Given $0 < t < 1$ and a bandwidth $h \in (0, t \wedge (1 - t))$, we set

$$\vartheta_n(t, h; s) := b_n(t + hs) - b_n(t), \quad \text{for } s \in [-1, 1]. \quad (1.3)$$

We will let $h > 0$ vary in such a way that $h'_n \leq h \leq h''_n$, where $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ are two sequences of positive constants such that $0 < h'_n \leq h''_n < \infty$ and, for either choice of $h_n = h'_n$ or $h_n = h''_n$, the conditions (H.1-2-3) below are fulfilled by $\{h_n\}_{n \geq 1}$. We assume that

- (H.1) $h_n \downarrow 0$ and $nh_n \uparrow \infty$ as $n \uparrow \infty$;
(H.2) $\log(1/h_n)/\log \log n \rightarrow \infty$ as $n \rightarrow \infty$;
(H.3) $nh_n/\log n \rightarrow \infty$ as $n \rightarrow \infty$.

In addition, we will say that the sequences $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$, with $0 < h'_n \leq h''_n < \infty$, for $n \geq 1$, fulfill the assumption (H.4) if at least one of the conditions (H.4)(i), (H.4)(ii) below are satisfied. Set, for $x \in \mathbb{R}^+$, $\log_2 x := \log_+ \log_+ x$, and $\log_+ x := \log(x \vee e)$.

- (H.4)(i) $\frac{\sqrt{n}h'_n \log(1/h'_n)}{\log n \sqrt{\log_2 n}} \rightarrow \infty$ as $n \rightarrow \infty$;
(H.4)(ii) $\frac{\sqrt{h''_n \log(1/h''_n)}}{h'_n \log(1/h'_n)} = o\left(\frac{\sqrt{n}}{\log n}\right)$ as $n \rightarrow \infty$.

Remark 1.1.1. (i) Note that, under (H.1-2), the hypothesis (H.4)(i) is satisfied whenever (H.3'), introduced below, is satisfied.

- (H.3') $n^{1/2}h'_n/\log n \rightarrow \infty$ as $n \rightarrow \infty$.

Obviously (H.3') is a stronger condition than (H.3).

(ii) For the particular choices $h'_n = n^{-r}$ and $h''_n = n^{-s}$, for $n \geq 1$, with $0 < s \leq r < 1$, the hypothesis (H.4)(ii) is satisfied whenever $s \leq r < (1+s)/2$.

(iii) Our main results are established under the hypotheses (H.1-2-3-4). However, it is unlikely that (H.4) is necessary, and these results may still hold under the only hypotheses (H.1-2-3) (see Remark 1.4.2 in the sequel).

We now specify the range of t in (1.3). We fix $0 < t'_1 < t_1 < t_2 < t'_2 < 1$, and let $d := (t'_1 - t_1) \wedge (t'_2 - t_2)$,

$$u'_1 = Q(t'_1), \quad u'_2 = Q(t'_2) \quad \text{and} \quad u_1 = Q(t_1), \quad u_2 = Q(t_2). \quad (1.4)$$

We then assume that the following conditions hold.

- (F.1) F is twice continuously differentiable on $\tilde{J} := [u'_1, u'_2]$;
(F.2) $F' = f$ is strictly positive on \tilde{J} ;
(F.3) $\sup_{u'_1 < u < u'_2} (F(u)(1 - F(u))|f'(u)|)/f^2(u) \leq \gamma$ for some $\gamma > 0$;
(F.4) $h''_n < d$ for all $n \geq 1$.

Note that (F.4) ensures that $t + hs \in J := [t'_1, t'_2]$, uniformly over $t \in I := [t_1, t_2]$, $h'_n \leq h \leq h''_n$ and $|s| \leq 1$. It is noteworthy that the inequality in (F.3) is equivalent to

$$\sup_{t'_1 < t < t'_2} t(1-t) \frac{|f'(Q(t))|}{f^2(Q(t))} \leq \gamma. \quad (1.5)$$

Moreover, the fact that $q(t) = \frac{d}{dt}Q(t)$ exists, and defines a positive and continuous function on J is ensured by the conditions (F.1-2-3). This, in turn, implies that $u'_1 < u_1 < u_2 < u'_2$.

Some more notation is needed for the statement of our results. Denote by \mathbb{S}_0 , the, so-called, Strassen set, a variant of which having been introduced by Strassen (1964) in the framework of the law of the iterated logarithm for partial sums. Here, \mathbb{S}_0 is the unit ball of

the reproducing kernel Hilbert space pertaining to the *two-sided standard Wiener process* $\{W(s) : |s| \leq 1\}$. The latter process is conveniently defined by setting

$$W(s) = \begin{cases} W_1(s) & \text{for } s \geq 0, \\ W_2(-s) & \text{for } s < 0, \end{cases}$$

where W_1 and W_2 are independent standard Wiener processes. We have, namely,

$$\mathbb{S}_0 = \left\{ g \in AC(-1, 1), \quad g(0) = 0 \quad \text{and} \quad \int_{-1}^1 \dot{g}(s)^2 ds \leq 1 \right\},$$

where $AC(-1, 1)$ stands for the set of all absolutely continuous functions g on $[-1, 1]$, with Lebesgue derivative \dot{g} . We denote by $\mathcal{B}(-1, 1)$ the set of all bounded functions on $[-1, 1]$ and set, for any $g \in \mathcal{B}(-1, 1)$, $\|g\| := \sup_{-1 \leq s \leq 1} |g(s)|$. Finally, we set, for any $\varepsilon > 0$, $\mathbb{S}_0^\varepsilon = \{h \in \mathcal{B}(-1, 1) : \inf_{g \in \mathbb{S}_0} \|g - h\| < \varepsilon\}$. Our main result may now be stated as follows.

Theorem 1.1.1. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3-4), with $0 < h'_n \leq h''_n < \infty$. Then, under (F.1-2-3-4), we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{t_1 \leq t \leq t_2} \left(\inf_{g \in \mathbb{S}_0} \left\| \frac{\vartheta_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \right) \right\} = 0. \quad (1.6)$$

Moreover, for each pair of constants c_1, c_2 with $0 < c_1 < c_2 < 1$, we have, almost surely,

$$\forall g \in \mathbb{S}_0, \quad \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \inf_{c_1 \leq t \leq c_2} \left\| \frac{\vartheta_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \right\} = 0. \quad (1.7)$$

The proof of Theorem 1.1.1 is postponed until Section 1.4. A rough outline of our arguments is as follows. First, we establish, in Section 1.2, a version of this theorem for iid uniform (0,1) random variables corresponding to the case where $f(x) = \mathbb{I}_{[0,1]}(x)$ and $q(t) = \mathbb{I}_{[0,1]}(t)$ (see, e.g., Theorem 1.2.1 below). Then, we make use of a continuity argument to treat the general framework. In Section 1.3 we present some statistical applications of our results. Section 1.4 is devoted to the proofs of our theorems. Finally, in the Appendix, we provide details on a technical fact used in Section 1.4.

1.2 The Uniform Case

Let U_1, U_2, \dots , be iid uniform (0,1) random variables. In this context, we denote by $U_n(t) := n^{-1} \#\{U_i \leq t, 1 \leq i \leq n\}$, for $t \in \mathbb{R}$, the (right-continuous) empirical distribution function, and by $V_n(t) := \inf\{u \geq 0 : U_n(u) \geq t\}$, for $0 \leq t \leq 1$, with $V_n(t) = 0$ for $t \leq 0$ and $V_n(t) = V_n(1)$ for $t \geq 1$, the (left-continuous) empirical quantile function. Here, $\#E$ denotes the cardinality of E . We define the uniform *empirical process* based upon U_1, \dots, U_n by $\alpha_n(t) := n^{1/2} (U_n(t) - t)$ for $t \in \mathbb{R}$, and the corresponding uniform sample *quantile process*

$$\beta_n(t) := n^{1/2} (V_n(t) - t) \quad \text{for } t \in \mathbb{R}. \quad (1.8)$$

Note that, in this setup, the quantile process is equal to the normed empirical quantile process, since, as was already pointed out, $q(t) = \mathbb{I}_{[0,1]}(t)$. Define further, for any $0 \leq t \leq 1$ and $h \in (0, t \wedge (1 - t))$, the increment functions

$$\xi_n(h, t; s) := \alpha_n(t + hs) - \alpha_n(t), \quad (1.9)$$

$$\zeta_n(h, t; s) := \beta_n(t + hs) - \beta_n(t), \quad (1.10)$$

for $-1 \leq s \leq 1$. The limiting behavior of the maximal oscillations of these processes has been extensively investigated in the literature (see, e.g., Stute (1982b), Deheuvels et Mason (1990), Deheuvels et Mason (1992), Deheuvels (1992), Deheuvels et Einmahl (2000), Mason (2004), and the references therein).

We are now ready to state, in Theorem 1.2.1, the uniform version of our Theorem 1.1.1. It is noteworthy that Theorem 1.2.1 is somehow stronger than Theorem 1.1.1, in the sense that it holds on a larger interval (compare (1.6) and (1.11) below). Stronger results may be obtained in the general case by making sharper assumptions upon the distribution function F . This will be considered elsewhere.

Theorem 1.2.1. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3-4), with $0 < h'_n \leq h''_n < \infty$. Then we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{h \leq t \leq 1-h} \left(\inf_{g \in \mathbb{S}_0} \left\| \frac{\zeta_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \right) \right\} = 0. \quad (1.11)$$

Moreover, for any pair of constants c_1, c_2 with $0 < c_1 < c_2 < 1$, we have, almost surely,

$$\forall g \in \mathbb{S}_0, \quad \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \inf_{c_1 \leq t \leq c_2} \left\| \frac{\zeta_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \right\} = 0. \quad (1.12)$$

The following corollary of Theorem 1.2.1 will be instrumental in the proof of Theorem 1.1.1, postponed until Section 1.4.2.

Corollary 1.2.1. *Under the assumptions of Theorem 1.2.1, we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{h \leq t \leq 1-h} \left(\sup_{-1 \leq s \leq 1} \frac{|\zeta_n(h, t; s)|}{\sqrt{2h \log(1/h)}} \right) \right\} = 1. \quad (1.13)$$

The proof of Theorem 1.2.1 is postponed until Section 1.4.1. An outline of our forthcoming arguments is as follows. First, we will establish, in Proposition 1.4.1 below, a version of this result for $\xi_n(h, t; \cdot)$. Proposition 1.4.1 will be shown to follow from Theorem 1.1 of Varron (2005)). Given this first result, Theorem 1.2.1 is straightforward under (H.4)(i), and, to establish Theorem 1.2.1 under (H.4)(ii), we will base the remainder of our proof on a uniform-in-bandwidth Bahadur-Kiefer-type representation of ζ_n in terms of ξ_n (see, e.g., Bahadur (1966), Kiefer (1967), Kiefer (1972) and Deheuvels et Mason (1992)). This representation is captured in Lemma 1.4.1 in the sequel.

1.3 Some Applications

1.3.1 The k -Spacings

In this sub-section, we provide some consequences of the just-given Theorems 1.1.1 and 1.2.1. The details of the corresponding proofs are postponed until Section 1.4.

We first consider the uniform case and denote by $U_{(1)} \leq \dots \leq U_{(n)}$ the order statistics pertaining to a random sample U_1, \dots, U_n of iid uniform (0,1) random variables. The k -spacings of $U_{(1)} \leq \dots \leq U_{(n)}$ are then defined by

$$\Delta_{i,n}(k) := U_{(k+i)} - U_{(i)} \quad \text{for } i = 0, \dots, n+1-k, \quad (1.14)$$

where $U_{(0)} := 0$ and $U_{(n+1)} := 1$. For any integer $1 \leq d \leq n$, we set

$$\delta_n(d) := \max_{1 \leq k \leq d} \max_{0 \leq i \leq n+1-k} |\Delta_{i,n}(k) - k/n|. \quad (1.15)$$

The following Theorem 1.3.1 will be shown to follow from Theorem 1.2.1. Set $[x] \geq x > [x] - 1$ the ceiling function of $x \in \mathbb{R}$.

Theorem 1.3.1. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3-4), with $0 < h'_n \leq h''_n < \infty$. Then, we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \frac{\sqrt{n} \delta_n(\lceil nh \rceil)}{\sqrt{2h \log(1/h)}} = 1. \quad (1.16)$$

The proof of Theorem 1.3.1 is given in Section 1.4.4.

We now turn our attention to the case of possibly non-uniform random variables. Let $X_{(1)} \leq \dots \leq X_{(n)}$ be the order statistics pertaining to a random sample X_1, \dots, X_n , of iid random variables with common distribution function F , density function f and quantile function Q . Under the hypotheses (F.1-2-3), define the k -spacings of $X_{(1)} \leq \dots \leq X_{(n)}$ by setting

$$D_{i,n}(k) := X_{(k+i)} - X_{(i)} \quad \text{for } i = i_{1,k,n}, i_{1,k,n} + 1, \dots, i_{2,n}, \quad (1.17)$$

where $i_{1,k,n}$ and $i_{2,n}$ are respectively defined by

$$i_{1,k,n} := \min\left\{i : \frac{i+k}{n} \geq t_1\right\} \quad (1.18)$$

$$i_{2,n} := \max\left\{i : \frac{i}{n} < t_2\right\}. \quad (1.19)$$

For any integer $1 \leq d \leq n$, we set

$$d_n(d) := \max_{1 \leq k \leq d} \max_{i_{1,k,n} \leq i \leq i_{2,n}} \left| f(X_{(i)}) \left| D_{i,n}(k) - \frac{k}{nf(X_{(i)})} \right| \right|. \quad (1.20)$$

Theorem 1.3.2. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3-4), with $0 < h'_n \leq h''_n < \infty$. Then, under (F.1-2-3-4), we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \frac{\sqrt{n} d_n(\lceil nh \rceil)}{\sqrt{2h \log(1/h)}} = 1. \quad (1.21)$$

The proof of Theorem 1.3.2 is postponed until Section 1.4.3.

Remark 1.3.1. *Consider the sequences $i'_{1,n}$ and $i'_{2,k,n}$ defined as follows,*

$$i'_{1,n} := \min\left\{i : \frac{i}{n} \geq t'_1\right\}, \quad (1.22)$$

$$i'_{2,k,n} := \max\left\{i : \frac{i+k}{n} \leq t'_2\right\}. \quad (1.23)$$

It is noteworthy that the two following relations hold for every integer $0 \leq k \leq \lceil nh \rceil$, $h \in [h'_n, h''_n]$,

$$1 \leq i'_{1,n} \leq i_{1,k,n} \quad \text{and} \quad i_{2,n} \leq i'_{2,k,n} \leq n - k.$$

Keep in mind the difference between the definitions (1.17) and (1.14), with respect to the possible values of the index i . The reason of this modified range is that the conclusion of Theorem 1.2.1 is slightly stronger than that of Theorem 1.1.1, as was already mentioned.

1.3.2 A Law of the Logarithm for Nearest-Neighbor Density Estimators

In this sub-section, we show that the results of the Sub-section 1.3.1 imply a uniform-in-bandwidth law of the logarithm for a nearest-neighbor nonparametric density estimator. Let, as in Sub-section 1.3.1, $X_{(1)} \leq \dots \leq X_{(n)}$ be the order statistics pertaining to a random sample X_1, \dots, X_n of iid variables with distribution function F , density function $F' = f$ and quantile function Q . Define the random sequences

$$u_{1,n} := \max(Q_n(t_1), u'_1), \quad u_{2,n} := \min(Q_n(t_2), u'_2), \quad n \geq 1, \quad (1.24)$$

where Q_n stands, as in Sub-section 1.1.2, for the empirical quantile function. Note that, for n large enough, $u_{1,n} < u_{2,n}$ almost surely. Further introduce K , an arbitrary kernel on \mathbb{R} , that is a measurable function integrating to one on \mathbb{R} , and denote by $\{k'_n\}_{n \geq 1}$ and $\{k''_n\}_{n \geq 1}$ two sequences such that $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ fulfill the conditions (H.1-2-3-4), with $h'_n = k'_n/n$ and $h''_n = k''_n/n$, for $n \geq 1$. Select $k > 0$ such that $k \in [k'_n, k''_n]$. On the interval $[u_{1,n}, u_{2,n}]$, define the k nearest-neighbor empirical density function, based upon the kernel K and the sample X_1, \dots, X_n , by

$$\widehat{f}_{n,k}(x) := \frac{1}{nR_k(x)} \sum_{i=1}^n K\left(\frac{x - X_i}{R_k(x)}\right), \quad (1.25)$$

where $R_k(x) := \inf\{r > 0 : [x - r/2, x + r/2] \supset [k]$ elements of the sample $X_1, \dots, X_n\}$. This random function is often referred to as the *adaptive variable bandwidth of order k* . The following additional assumptions upon the kernel K will be needed to state our result concerning nearest-neighbor density estimators.

(K.A) K is of bounded variation on \mathbb{R} .

(K.B) K is compactly supported.

Theorem 1.3.3. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3-4), with $0 < h'_n \leq h''_n < \infty$. Set, for $n \geq 1$, $k'_n = nh'_n$ and $k''_n = nh''_n$. Then, under (F.1-2-3-4) and (K.A-B), we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{x \in [u_{1,n}, u_{2,n}]} \sup_{k \in [k'_n, k''_n]} \frac{\sqrt{k} |\widehat{f}_{n,k}(x) - \mathbb{E}\widehat{f}_{n,k}(x)|}{\sqrt{2f^2(x) \log(n/k)}} = \left\{ \int_{\mathbb{R}} K^2(t) dt \right\}^{1/2}. \quad (1.26)$$

Remark 1.3.2. *Note that, in the setting of density estimation, the hypothesis (H.4)(i) generally holds, since most of the bandwidth selection procedures lead to choices like $h'_n = c_1 n^{-\alpha}$, for a given $0 < c_1 < \infty$, with $\alpha < 1/2$.*

We will show, in the forthcoming Section 1.4.5, that Theorem 1.3.3 is a natural consequence of a combination of Corollary 2.1 of Varron (2006), with Theorem 1.3.2. We mention that Theorem 1.1 of Varron (2005) may be adapted likewise to obtain a similar result as that stated in Theorem 1.3.3, in the case of the usual Parzen-Rosenblatt (see, e.g., Parzen (1962) and Rosenblatt (1956)) kernel density estimator (see Facts 1.5.1 and 1.5.2 in the Appendix).

1.4 Proofs

1.4.1 Proof of Theorem 1.2.1

We first establish the analogue of Theorem 1.2.1 for ξ_n . Recall the definition (1.9) of ξ_n .

Proposition 1.4.1. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3), with $0 < h'_n \leq h''_n < \infty$. Then, with probability one,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{h \leq t \leq 1-h} \left(\inf_{g \in \mathbb{S}_0} \left\| \frac{\xi_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \right) \right\} = 0. \quad (1.27)$$

Moreover, for any pair of constants $0 < c_1 < c_2 < 1$, we have, almost surely,

$$\forall g \in \mathbb{S}_0, \quad \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \inf_{c_1 \leq t \leq c_2} \left\| \frac{\xi_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \right\} = 0. \quad (1.28)$$

Note that, the hypotheses (H.1-2-3) are sufficient in Proposition 1.4.1.

Proof. Set $\mathcal{G}_n := \{ \mathbb{I}_{[(t \wedge (t+v)), (t \vee (t+v))]} , h \in [h'_n, h''_n], h \leq t \leq 1-h, -h \leq v \leq h \}$ and

$$\mathcal{G} := \{ \mathbb{I}_{[x, y]}, 0 \leq x \leq 1, 0 \leq y \leq 1 \}. \quad (1.29)$$

It is well known that the class of all closed intervals in \mathbb{R} forms a *VC* class (see, e.g., van der Vaart et Wellner (1996)). Therefore, making use of the result of Exercise 9 on page 151 of van der Vaart et Wellner (1996), it is readily shown that \mathcal{G} constitutes a *VC*-subgraph class of functions (we refer to Section 2.6.2 in van der Vaart et Wellner (1996) for the definitions of *VC* classes of sets, and *VC*-subgraph classes of functions, see also the Annex of the present document). Given that $\mathcal{G}_n \subset \mathcal{G}$, we see that \mathcal{G}_n constitutes a *VC*-subgraph class of functions and satisfies the entropy condition (HK.III)(i) in the Appendix. Thus, (1.28) is a direct consequence of Theorem 1.1 of Varron (2005) (which is recalled in Fact 1.5.1 in the Appendix for convenience). The same arguments readily show that, for each $0 < \lambda < 1/2$ and $\varepsilon > 0$, there exists almost surely an $n(\varepsilon)$ such that, for all $n \geq n(\varepsilon)$,

$$\{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [\lambda, 1-\lambda] \} \subset \mathbb{S}_0^\varepsilon. \quad (1.30)$$

Our proof is completed by the observation that $\tilde{U}_i := [(U_i + \lambda) \text{ modulo } 1]$ and U_i are identically distributed. Thus, we have the distributional identities

$$\begin{aligned} & \{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [h, \lambda] \}_{n \geq 1} \\ & \stackrel{d}{=} \{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [\lambda, 2\lambda - h] \}_{n \geq 1}, \end{aligned}$$

and

$$\begin{aligned} & \{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [1-\lambda, 1-h] \}_{n \geq 1} \\ & \stackrel{d}{=} \{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [1-2\lambda, 1-\lambda-h] \}_{n \geq 1}. \end{aligned}$$

By combining these statements with (1.30), we obtain, in turn, that for each $\varepsilon > 0$, there exists almost surely an $n(\varepsilon)$, such that for all $n \geq n(\varepsilon)$,

$$\{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [0, \lambda] \} \subset \mathbb{S}_0^\varepsilon,$$

and

$$\{ \xi_n(h, t; \cdot) : h \in [h'_n, h''_n], t \in [1-\lambda, 1-h] \} \subset \mathbb{S}_0^\varepsilon.$$

The proof of proposition 1.4.1 is now complete. \square

Remark 1.4.1. *We note that Corollary 3 of Mason (2004), which is a d -variate extension of Theorem 3.1 of Deheuvels et Mason (1992), can be established through these arguments.*

The completion of the proof of Theorem 1.2.1 under the hypothesis (H.4)(i) is straightforward. Indeed, Kiefer (1967) showed that, almost surely as $n \rightarrow \infty$,

$$\sup_{0 \leq t \leq 1} |\alpha_n(t) + \beta_n(t)| = \mathcal{O}\left(\frac{(\log n)^{1/2}(\log_2 n)^{1/4}}{n^{1/4}}\right),$$

which implies that, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} \sup_{h \in [h'_n, h''_n]} \sup_{h \leq t \leq 1-h} \sup_{|s| \leq 1} \frac{\zeta_n(h, t; s)}{\sqrt{2h \log(1/h)}} &= \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{h \leq t \leq 1-h} \sup_{|s| \leq 1} \frac{\xi_n(h, t; s)}{\sqrt{2h \log(1/h)}} \right. \\ &\quad \left. + \mathcal{O}\left(\frac{(\log n)^{1/2}(\log_2 n)^{1/4}}{n^{1/4}(h \log(1/h))^{1/2}}\right) \right\}, \end{aligned} \quad (1.31)$$

and

$$\begin{aligned} \sup_{h \in [h'_n, h''_n]} \inf_{h \leq t \leq 1-h} \sup_{|s| \leq 1} \frac{\zeta_n(h, t; s)}{\sqrt{2h \log(1/h)}} &= \sup_{h \in [h'_n, h''_n]} \left\{ \inf_{h \leq t \leq 1-h} \sup_{|s| \leq 1} \frac{\xi_n(h, t; s)}{\sqrt{2h \log(1/h)}} \right. \\ &\quad \left. + \mathcal{O}\left(\frac{(\log n)^{1/2}(\log_2 n)^{1/4}}{n^{1/4}(h \log(1/h))^{1/2}}\right) \right\}. \end{aligned} \quad (1.32)$$

Since, under (H.4)(i),

$$\sup_{h \in [h'_n, h''_n]} \mathcal{O}\left(\frac{(\log n)(\log_2 n)^{1/2}}{n^{1/2}h \log(1/h)}\right) = o(1),$$

(1.31)–(1.32), when combined with Proposition 1.4.1, are enough to establish Theorem 1.2.1 under the hypothesis (H.4)(i).

To prove Theorem 1.2.1 under the hypothesis (H.4)(ii), we will work under the following notation. We will set $\|g\|_+ := \sup_{0 \leq s \leq 1} |g(s)|$ for the sup-norm of a function $g \in \mathcal{B}(0, 1)$, in contrast with $\|h\| = \sup_{-1 \leq s \leq 1} |g(s)|$, used when $g \in \mathcal{B}(-1, 1)$. To simplify matters, we will give proofs of our theorems with the formal replacement of $\|\cdot\|$ by $\|\cdot\|_+$. The technicalities needed to revert from $\|\cdot\|_+$ to $\|\cdot\|$ are straightforward, but lengthy, and will therefore be omitted.

The completion of the proof of Theorem 1.2.1 under the hypothesis (H.4)(ii) will require a uniform-in-bandwidth Bahadur-Kiefer-type representation of ζ_n in terms of ξ_n . The following two lemmas are oriented towards the aim of establishing a representation of the kind. Our forthcoming results mimick that obtained by Deheuvels et Mason (1992), in a slightly more general setup of varying bandwidths. We give the details of their proofs for the sake of completeness.

Lemma 1.4.1. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two sequences fulfilling the conditions (H.1-2-3) and (H.4)(ii). Then, for any $\lambda > 1$, we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{0 \leq t \leq 1 - \lambda h} \frac{\|\zeta_n(h, t; \cdot) + \xi_n(h, V_n(t); \cdot)\|_+}{\sqrt{2h \log(1/h)}} = 0. \quad (1.33)$$

Proof. Choose any $h \in [h'_n, h''_n]$, with $\lambda > 1$, $0 < t < 1 - \lambda h$ and $0 < s < 1$. Observe that, ultimately as $n \rightarrow \infty$,

$$\begin{aligned} \zeta_n(h, t; s) + \xi_n(h, V_n(t); s) &= \{\zeta_n(h, t; s) + \alpha_n(V_n(t + hs)) - \alpha_n(V_n(t))\} \\ &\quad + \{\alpha_n(V_n(t) + hs) - \alpha_n(V_n(t + hs))\}. \end{aligned} \quad (1.34)$$

Making use of the easily proven fact that $|U_n(V_n(t)) - t| \leq 1/n$ for any $n \geq 1$ and $0 \leq t \leq 1$, an application of the triangle inequality to the right-hand side of (1.34) establishes, in turn, that, uniformly over $h \in [h'_n, h''_n]$, $0 \leq t \leq 1 - \lambda h$ and $0 \leq s \leq 1$,

$$|\zeta_n(h, t; s) + \{\alpha_n(V_n(t + hs)) - \alpha_n(V_n(t))\}| \leq 2n^{-1/2}. \quad (1.35)$$

We invoke Theorem 1(III) in Mason (1984) to obtain that, whenever $\{h''_n\}_{n \geq 1}$ satisfies (H.1-2-3), we have

$$\lim_{n \rightarrow \infty} \sup_{0 \leq h \leq h''_n} \sup_{0 \leq t \leq 1-h} \frac{\sqrt{n}|V_n(t+h) - V_n(t) - h|}{\sqrt{2h''_n \log(1/h''_n)}} = 1 \text{ a.s. .}$$

Thus, for any $\varepsilon > 0$, we have, almost surely for all n sufficiently large, uniformly over $h \in [h'_n, h''_n]$, $0 \leq t \leq 1 - \lambda h$ and $0 \leq s \leq 1$,

$$|V_n(t + hs) - V_n(t) - hs| \leq v_n := (1 + \varepsilon)^2 (2n^{-1}h''_n \log(1/h''_n))^{1/2}. \quad (1.36)$$

We next observe that, if the sequence $\{h''_n\}_{n \geq 1}$ satisfies, ultimately in $n \rightarrow \infty$, the assumptions (H.1-2-3), then, such is also the case for $(v_n)_{n \geq 1}$ in (1.36). The first half of Proposition 1.4.1 implies therefore that

$$\limsup_{n \rightarrow \infty} \sup_{0 \leq v \leq v_n} \sup_{0 \leq V_n(u) \leq 1-v_n} \frac{|\alpha_n(V_n(u)) - \alpha_n(V_n(u) + v)|}{\sqrt{2v_n \log(1/v_n)}} \leq 1 \text{ a.s. .} \quad (1.37)$$

Observe that, since $\{h''_n, n \geq 1\}$ satisfies (H.1-2-3), we have

$$1/2 \leq \liminf_{n \rightarrow \infty} (\log(1/v_n)) / \log n \leq \limsup_{n \rightarrow \infty} (\log(1/v_n)) / \log n \leq 1. \quad (1.38)$$

By combining (1.36) with (1.37) and (1.38), we get that, for all n sufficiently large,

$$\begin{aligned} & \sup_{h \in [h'_n, h''_n]} \sup_{0 \leq t \leq 1-\lambda h} \sup_{0 \leq s \leq 1} n^{1/4} (h''_n \log(1/h''_n))^{-1/4} (\log n)^{-1/2} \\ & \times |\alpha_n(V_n(t + hs)) - \alpha_n(V_n(t) + hs)| \leq 2^{3/4}(1 + \varepsilon) \text{ a.s.} \end{aligned}$$

Since $\varepsilon > 0$ can be chosen arbitrarily small in the above inequality, we see that, almost surely,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{0 \leq t \leq 1-\lambda h} \sup_{0 \leq s \leq 1} n^{1/4} (h''_n \log(1/h''_n))^{-1/4} (\log n)^{-1/2} \\ & \times |\alpha_n(V_n(t + hs)) - \alpha_n(V_n(t) + hs)| \leq 2^{3/4}. \end{aligned} \quad (1.39)$$

Under the hypothesis (H.4)(ii), (1.39), when combined with (1.34) and (1.35), suffices to complete the proof of (1.33). \square

Remark 1.4.2. *It is likely that the following result holds : for any sequences $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ fulfilling the hypotheses (H.1-2-3),*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{0 \leq t \leq 1-h} \sup_{0 \leq s \leq 1} \frac{|V_n(t + hs) - V_n(t) - hs|}{\sqrt{2h \log(1/h)}} \leq 1 \text{ a.s. .} \quad (1.40)$$

If so, following the same lines as above, it can be shown that the result of Lemma 1.4.1, and, consequently, the results of Theorems 1.1.1, 1.3.1, 1.3.2 and 1.3.3, still hold under the only hypotheses (H.1-2-3).

Lemma 1.4.2. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two sequences, assumed, each, to fulfill the conditions (H.1-2-3-4). Then, for any fixed $\lambda > 1$, we have almost surely*

$$\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{1-\lambda h \leq t \leq 1-h} \frac{\|\zeta_n(h, t; \cdot) + \xi_n(h, 1-h; \cdot)\|_+}{\sqrt{2h \log(1/h)}} \leq 2\sqrt{\lambda-1}, \quad (1.41)$$

and for all n sufficiently large

$$V_n(1-\lambda h) < 1-h. \quad (1.42)$$

Proof. By setting, respectively, $t = h, s = -1$ in Proposition 1.4.1, and $t = 0, s = 1$ in Lemma 1.4.1, we obtain readily that $\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \pm \beta_n(h) / \sqrt{2h \log(1/h)} \leq 1$ almost surely. Thus, under (H.1-2-3-4), we have, almost surely,

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} (V_n(1-\lambda h) - (1-h)) / h = 1 - \lambda < 0. \quad (1.43)$$

Set $\omega_n(a) := \sup_{0 \leq t \leq 1-a} \{\|\xi_n(a, t; \cdot)\|_+\}$ for $0 \leq a \leq 1$. By combining (1.43) with the above Proposition 1.4.1, we obtain that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{1-\lambda h \leq t \leq 1-h} \frac{\|\xi_n(h, V_n(t); \cdot) - \xi_n(h, 1-h; \cdot)\|_+}{\sqrt{2h \log(1/h)}} \\ & \leq \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \frac{2\omega_n((\lambda-1)h)}{\sqrt{2h \log(1/h)}} = 2\sqrt{\lambda-1} \quad \text{a.s.} \end{aligned}$$

In view of (1.33), this last result readily yields (1.41), whereas (1.42) is a direct consequence of (1.43). \square

We are now ready to complete the proof of Theorem 1.2.1. To establish (1.11), we fix an $\varepsilon > 0$, and choose $\lambda = 1 + \frac{\varepsilon^2}{16}$. In view of (1.41), (1.42), Proposition 1.4.1 and (1.33), there exists almost surely an $N_\varepsilon < \infty$, such that, for all $n \geq N_\varepsilon$,

$$V_n(1-\lambda h) < 1-h \quad \text{uniformly over } h \in [h'_n, h''_n],$$

and

$$\sup_{h \in [h'_n, h''_n]} \left\{ \sup_{0 \leq t \leq 1-h} \left(\inf_{g \in \mathbb{S}_0} \left\| \frac{\xi_n(h, t; \cdot)}{2h \log(1/h)} - g \right\|_+ \right) \right\} < \varepsilon/2 \quad \text{a.s.},$$

$$\sup_{h \in [h'_n, h''_n]} \left\{ \sup_{0 \leq t \leq 1-\lambda h} \frac{\|\zeta_n(h, t; \cdot) + \xi_n(h, V_n(t); \cdot)\|_+}{\sqrt{2h \log(1/h)}} \right\} \leq \varepsilon/2 \quad \text{a.s.}, \quad (1.44)$$

$$\sup_{h \in [h'_n, h''_n]} \left\{ \sup_{1-\lambda h \leq t \leq 1-h} \frac{\|\zeta_n(h, t; \cdot) + \xi_n(h, 1-h; \cdot)\|_+}{\sqrt{2h \log(1/h)}} \right\} \leq \varepsilon/2 \quad \text{a.s.} \quad (1.45)$$

Therefore, for all $n \geq N_\varepsilon$, we have, with probability one,

$$\sup_{h \in [h'_n, h''_n]} \left\{ \sup_{0 \leq t \leq 1-h} \left(\inf_{g \in \mathbb{S}_0} \left\| \frac{\zeta_n(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\|_+ \right) \right\} < \varepsilon. \quad (1.46)$$

This last result suffices for the proof of the version of (1.11) obtained with $\|\cdot\|_+$ replacing $\|\cdot\|$. The proof of the version of (1.11) making use of the sup-norm $\|\cdot\|$ follows along the same lines and is therefore omitted.

To establish (1.12), we first select an arbitrary $g \in \mathbb{S}_0$ and fix a $\varepsilon > 0$. By (1.28), there exist almost surely an $n_\varepsilon^{(4)}$ and a sequence $t_n^{(1)} \in (1/4, 3/4)$, $n = 1, 2, \dots$, such that, for all $n \geq n_\varepsilon^{(4)}$,

$$\sup_{h \in [h'_n, h''_n]} \left\| \xi_n(h, t_n^{(1)}; \cdot) / \sqrt{2h \log(1/h)} - g \right\|_+ < \varepsilon/4. \quad (1.47)$$

Set now $t_n = U_n(t_n^{(1)})$ for $n \geq 1$. We have, uniformly over $h \in [h'_n, h''_n]$,

$$\begin{aligned} \left\| \frac{\zeta_n(h, t_n^{(1)}; \cdot)}{\sqrt{2h \log(1/h)}} + g \right\|_+ &\leq \frac{\|\zeta_n(h, t_n; \cdot) + \xi_n(h, V_n(t_n); \cdot)\|_+}{\sqrt{2h \log(1/h)}} \\ &+ \frac{\|-\xi_n(h, V_n(t_n); \cdot) + \xi_n(h, t_n^{(1)}; \cdot)\|_+}{\sqrt{2h \log(1/h)}} \\ &+ \left\| -\frac{\xi_n(h, t_n^{(1)}; \cdot)}{\sqrt{2h \log(1/h)}} + g \right\|_+. \end{aligned} \quad (1.48)$$

The Glivenko-Cantelli theorem, when combined with the definition $t_n = U_n(t_n^{(1)})$ of t_n , with $t_n^{(1)} \in (1/4, 3/4)$, readily implies that, almost surely for all n sufficiently large, $t_n \in (1/8, 7/8)$ and $V_n(t_n) \leq t_n^{(1)} < V_n(t_n + 1/n)$. This, in turn, entails that (see, e.g., Deheuvels (1982))

$$\limsup_{n \rightarrow \infty} |V_n(t_n) - t_n^{(1)}| (n/\log n) = 1 \quad \text{a.s.} \quad (1.49)$$

Set $\rho_n := \log n/n$ for $n \geq 1$. By Theorem 1(I) of Mason *et al.* (1983), it follows that (see, e.g., (2.17) in Deheuvels et Mason (1992))

$$\sup_{|t' - t''| \leq 2\rho_n} \sup_{0 \leq t' \leq 1 - \rho_n} \frac{n}{\log n} [U_n(t') - U_n(t'')] = \mathcal{O}(1) \quad \text{a.s. .}$$

By combining this last result with (1.49), we readily obtain that

$$\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \frac{\|-\xi_n(h, V_n(t_n); \cdot) + \xi_n(h, t_n^{(1)}; \cdot)\|_+}{\sqrt{2h \log(1/h)}} = 0 \quad \text{a.s. .} \quad (1.50)$$

By (1.44) and (1.50), there exists almost surely an $N'_\varepsilon < \infty$, such that, for all $n \geq N'_\varepsilon$, we have $1/8 < t_n < 7/8$, $t_n < 1 - \lambda h''_n$ and

$$\sup_{h \in [h'_n, h''_n]} \left\| \frac{\zeta_n(h, t_n^{(1)}; \cdot)}{\sqrt{2h \log(1/h)}} + g \right\|_+ < \varepsilon. \quad (1.51)$$

By choosing $\varepsilon > 0$ arbitrarily small in (1.51), we obtain readily (1.12), which completes the proof of the version of Theorem 1.2.1 pertaining to the sup-norm $\|\cdot\|_+$. The extension to the case of $\|\cdot\|$ is straightforward, and hence, omitted. \square

1.4.2 Proof of Theorem 1.1.1

Recall that $Q_n(t) = Q(V_n(t))$ for $0 < t < 1$. Keep in mind that the fact that $q(t) = \frac{d}{dt}Q(t)$ exists, and defines a positive and continuous function on $[t'_1, t'_2]$, is a consequence of (1.2)-(1.4), when combined with the assumptions (F.1-2-3). In view of the definitions (1.1)-(1.8) of b_n and β_n , this, when combined with Taylor's formula, entails, almost surely for all n

sufficiently large, the existence of a $\theta_{t,n} \in (t \wedge V_n(t), t \vee V_n(t))$ for each $t \in I = [t_1, t_2]$, such that

$$b_n(t) = \frac{n^{1/2}(Q(V_n(t)) - Q(t))}{q(t)} = n^{1/2}(V_n(t) - V(t)) \frac{q(\theta_{t,n})}{q(t)} = \beta_n(t) \frac{q(\theta_{t,n})}{q(t)}. \quad (1.52)$$

Here, the Glivenko-Cantelli theorem is enough to show that, almost surely, $|V_n(t) - t| \rightarrow 0$ uniformly over $t \in I = [t_1, t_2]$. Thus, with probability 1 for all large n , we have

$$(t \wedge V_n(t), t \vee V_n(t)) \subseteq J = [t'_1, t'_2] \quad \text{for all } t \in I.$$

Recall the definition (1.3) of $\vartheta_n(t, h, s) = b_n(t + hs) - b_n(t)$, and the definition (1.10) of $\zeta_n(h, t; s) = \beta_n(t + hs) - \beta_n(t)$. In view of (1.52), we may write

$$\vartheta_n(t, h; s) = \left\{ \beta_n(t + hs) - \beta_n(t) + \beta_n(t) \right\} \frac{q(\theta_{t+hs,n})}{q(t + hs)} - \beta_n(t) \frac{q(\theta_{t,n})}{q(t)},$$

whence

$$\begin{aligned} \vartheta_n(t, h; s) - \zeta_n(t, h; s) &= \zeta_n(t, h; s) \left(\frac{q(\theta_{t+hs,n})}{q(t + hs)} - 1 \right) \\ &\quad - \beta_n(t) \left(\frac{q(\theta_{t,n})}{q(t)} - \frac{q(\theta_{t+hs,n})}{q(t + hs)} \right) \\ &=: A_n(t, h; s) - B_n(t, h; s), \end{aligned} \quad (1.53)$$

with $\theta_{t+hs,n} \in [(t + hs) \wedge V_n(t + hs), (t + hs) \vee V_n(t + hs)] \subseteq J = [t'_1, t'_2]$, for all $t \in I$, $|s| \leq 1$ and $h \in [h'_n, h''_n]$, with probability 1 for n large enough (by the hypothesis (F.4)). To control both $A_n(t, h; s)$ and $B_n(t, h; s)$, we will make use of the following fact, due to Csörgő et Révész (1978) (see also Lemma 1.4.1 in Csörgő (1983)).

Fact 1.4.1. *Under (F1-2-3-4), we have, for every $y_1, y_2 \in (t'_1, t'_2)$,*

$$\frac{q(y_2)}{q(y_1)} = \frac{f(Q(y_1))}{f(Q(y_2))} \leq \left\{ \frac{(y_1 \vee y_2)}{(y_1 \wedge y_2)} \times \frac{1 - (y_1 \wedge y_2)}{1 - (y_1 \vee y_2)} \right\}^\gamma,$$

where $\gamma > 0$ is as in (F3).

First consider $B_n(t, h; s)$ in (1.53). We set

$$\begin{aligned} \tau_n(h, t; s) &:= \max \left\{ \left[(t \vee V_n(t)) - (t \wedge V_n(t)) \right], \right. \\ &\quad \left. \left[((t + hs) \vee V_n(t + hs)) - ((t + hs) \wedge V_n(t + hs)) \right] \right\} \\ &= \max \left\{ |V_n(t) - t|, |V_n(t + hs) - (t + hs)| \right\}, \end{aligned} \quad (1.54)$$

and

$$\tau_n := \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{t_1 \leq t \leq t_2} \left[\sup_{|s| \leq 1} \tau_n(h, t; s) \right] \right\}, \quad (1.55)$$

Noting that $\theta_{t,n} \vee t \leq t + \tau_n$ and $\theta_{t,n} \wedge t \geq t - \tau_n$, and making use of Fact 1.4.1, we get, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} \frac{q(\theta_{t,n})}{q(t)} &\leq \left(\frac{t + \tau_n}{t - \tau_n} \times \frac{1 - (t - \tau_n)}{1 - (t + \tau_n)} \right)^\gamma \\ &\leq \left(1 + \frac{2\tau_n}{t - t^2 - \tau_n + \tau_n^2} \right)^\gamma. \end{aligned} \quad (1.56)$$

Similarly, we readily obtain that, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} \frac{q(\theta_{t+hs,n})}{q(t+hs)} &\leq \left(1 + \frac{2\tau_n}{t+hs - (t+hs)^2 - \tau_n + \tau_n^2}\right)^\gamma & (1.57) \\ \text{and } \frac{q(\theta_{t+hs,n})}{q(t+hs)} &\geq \left(1 - \frac{2\tau_n}{t+hs - (t+hs)^2 - \tau_n + \tau_n^2}\right)^\gamma. \end{aligned}$$

By all this, setting $\Gamma = \lfloor \gamma \rfloor + 1$, we get, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} \frac{q(\theta_{t,n})}{q(t)} &\leq \left(1 + \frac{2\tau_n}{t - t^2 - \tau_n + \tau_n^2}\right)^\Gamma, \\ \frac{q(\theta_{t+hs,n})}{q(t+hs)} &\geq \left(1 - \frac{2\tau_n}{t+hs - (t+hs)^2 - \tau_n + \tau_n^2}\right)^\Gamma. \end{aligned}$$

Making use of the formula $x^\Gamma - y^\Gamma = (x - y)(\sum_{j=0}^{\Gamma-1} x^j y^{\Gamma-j})$, in combination with (1.57), and observing that, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} \frac{1}{t - t^2 - \tau_n + \tau_n^2} - \frac{1}{t+hs - (t+hs)^2 - \tau_n + \tau_n^2} &= \frac{hs(1 - 2t - hs)}{(t - t^2 - \tau_n + \tau_n^2)(t+hs - (t+hs)^2 - \tau_n + \tau_n^2)}, \end{aligned}$$

we readily conclude that, almost surely as $n \rightarrow \infty$,

$$\frac{q(\theta_{t,n})}{q(t)} - \frac{q(\theta_{t+hs,n})}{q(t+hs)} = \mathcal{O}(\tau_n), \quad (1.58)$$

uniformly over $t \in I$, $h \in [h'_n, h''_n]$ and $s \in [-1, 1]$. By the Chung (1949) law of the iterated logarithm [LIL] applied to the sup-norm of β_n , we observe that, almost surely as $n \rightarrow \infty$,

$$\tau_n \leq n^{-1/2} \sup_{0 \leq t \leq 1} |\beta_n(t)| = \mathcal{O}\left(\sqrt{\frac{\log_2 n}{n}}\right). \quad (1.59)$$

Then, by combining (H.1), (H.2), (1.53), (1.54), (1.55), (1.58), (1.59) and the Chung (1949) LIL, it is readily shown that, for any $\varepsilon > 0$, there exists almost surely an $N_{1,\varepsilon} < \infty$, such that for all $n \geq N_{1,\varepsilon}$, we have

$$\sup_{h \in [h'_n, h''_n]} \sup_{t_1 \leq t \leq t_2} \sup_{-1 \leq s \leq 1} \frac{|B_n(t, h; s)|}{\sqrt{2h \log(1/h)}} \leq \varepsilon. \quad (1.60)$$

We now turn our attention to $A_n(t, h; s)$ in (1.53). A similar argument based upon (1.57) shows that

$$\frac{q(\theta_{t+hs,n})}{q(t+hs)} - 1 \rightarrow 0, \quad \text{almost surely as } n \rightarrow \infty. \quad (1.61)$$

By combining (1.61) with the conclusion of Corollary 1.2.1, it is readily shown that, for each $\varepsilon > 0$, there exists almost surely an $N_{2,\varepsilon} < \infty$, such that, for all $n \geq N_{2,\varepsilon}$,

$$\sup_{h \in [h'_n, h''_n]} \sup_{t_1 \leq t \leq t_2} \sup_{-1 \leq s \leq 1} \frac{|A_n(t, h; s)|}{\sqrt{2h \log(1/h)}} \leq \varepsilon. \quad (1.62)$$

The proof of Theorem 1.1.1 is completed by combining (1.53) with (1.60) and (1.62).

1.4.3 Proof of Theorem 1.3.2

Recall the definition (1.1) of $b_n(t) = n^{1/2}(Q_n(t) - Q(t))/q(t)$, where $Q_n(t) = X_{(i)}$ for $(i-1)/n < t \leq i/n$. Thus, for any $0 < s, t < 1$, setting $i = \lceil nt \rceil$ and $j = \lceil ns \rceil$, where $\lceil u \rceil \geq u > \lceil u \rceil - 1$ is the ceiling function, we have $Q_n(s) = X_{(j)}$, $Q_n(t) = X_{(i)}$, and

$$\begin{aligned} b_n(s) - b_n(t) &= n^{1/2} \left\{ \frac{X_{(j)} - Q(s)}{q(s)} - \frac{X_{(i)} - Q(t)}{q(t)} \right\} \\ &= \frac{n^{1/2}}{q(t)} \left\{ X_{(j)} - X_{(i)} - (Q(s) - Q(t)) + (X_{(j)} - Q(s)) \left[\frac{q(t)}{q(s)} - 1 \right] \right\} \\ &= \frac{n^{1/2}}{q(t)} \left\{ \left(X_{(j)} - X_{(i)} - (Q(\frac{j}{n}) - Q(\frac{i}{n})) \right) + \left((Q(\frac{j}{n}) - Q(\frac{i}{n})) - (Q(s) - Q(t)) \right) \right. \\ &\quad \left. + (Q_n(s) - Q(s)) \left[\frac{q(t)}{q(s)} - 1 \right] \right\}. \end{aligned}$$

This allows us to write

$$b_n(s) - b_n(t) = \frac{n^{1/2}}{q(t)} \left[\left(X_{(j)} - X_{(i)} - Q(\frac{j}{n}) + Q(\frac{i}{n}) \right) + \varepsilon_n(t, s) \right], \quad (1.63)$$

where

$$\begin{aligned} \varepsilon_n(t, s) &:= \varepsilon_{1,n}(t, s) - \varepsilon_{2,n}(t, s), \\ \varepsilon_{1,n}(t, s) &:= (Q_n(s) - Q(s)) \left[\frac{q(t)}{q(s)} - 1 \right], \\ \varepsilon_{2,n}(t, s) &:= Q(s) - Q(t) - (Q(\frac{j}{n}) - Q(\frac{i}{n})). \end{aligned} \quad (1.64)$$

Since $(i-1)/n < t \leq i/n$ and $(j-1)/n < s \leq j/n$, the assumptions (F.1–2) readily imply that,

$$|Q(s) - Q(\frac{j}{n})| \leq |s - \frac{j}{n}| q(s_1) \leq \frac{q(s_1)}{n} \quad \text{for some } s_1 \in ((j-1)/n, j/n),$$

and, similarly,

$$|Q(t) - Q(\frac{i}{n})| \leq |t - \frac{i}{n}| q(t_1) \leq \frac{q(t_1)}{n} \quad \text{for some } t_1 \in ((i-1)/n, i/n).$$

It follows therefore that, for each specified $0 < \lambda < 1$, we have, uniformly over $h \in [h'_n, h''_n]$, ultimately as $n \rightarrow \infty$,

$$\sup_{\lambda \leq t \leq 1-\lambda} \sup_{-1 \leq u \leq 1} |\varepsilon_{2,n}(t, t+hu)| = \mathcal{O}\left(\frac{1}{n}\right) \text{ a.s.} \quad (1.65)$$

Moreover, under (F.3), it follows readily from Fact 1.4.1 that, for $u > 0$ and as $h \rightarrow 0$,

$$\begin{aligned} \frac{q(t)}{q(t+hu)} - 1 &\leq \left(\frac{t+hu}{t} \times \frac{1-t}{1-t-hu} \right)^\gamma - 1 \\ &= \left(1 + \frac{uh}{t[1-(t+uh)]} \right)^\gamma - 1 \\ &= \mathcal{O}(h^\gamma). \end{aligned}$$

Similarly, for $u < 0$, one can show that $\frac{q(t)}{q(t+hu)} - 1 = \mathcal{O}(h^\gamma)$ as $h \rightarrow 0$. We now combine these two last results with (1.52) and (1.59). We so obtain that, for each specified $0 < \lambda < 1$, we have, with probability 1, uniformly over $h \in [h'_n, h''_n]$ and ultimately as $n \rightarrow \infty$,

$$\sup_{\lambda \leq t \leq 1-\lambda} \sup_{-1 \leq u \leq 1} |\varepsilon_{1,n}(t, t+hu)| = \mathcal{O}\left(h^\gamma \sqrt{\frac{\log_2 n}{n}}\right). \quad (1.66)$$

Recall that $(i-1)/n < t \leq i/n$ in (1.63). A Taylor expansion based upon (F.1), together with the Chung (1949) LIL, as stated in (1.59), and (1.52), shows readily that, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} |f(Q(t)) - f(X_{(i)})| &= |f(Q(t)) - f(Q_n(t))| = \mathcal{O}\left(|Q(t) - Q_n(t)|\right) \\ &= \mathcal{O}\left(|Q(t) - Q\left(\frac{i}{n}\right)|\right) + \mathcal{O}\left(Q\left(\frac{i}{n}\right) - Q_n\left(\frac{i}{n}\right)\right) \\ &= \mathcal{O}\left(\sqrt{\frac{\log_2 n}{n}}\right) + \mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}\left(\sqrt{\frac{\log_2 n}{n}}\right). \end{aligned} \quad (1.67)$$

Set

$$d_n^{(1)}(d) := \max_{1 \leq k \leq d} \max_{i_{1,k,n} \leq i \leq i_{2,n}} |f(X_{(i)})| D_{i,n}(k) - \left(Q\left(\frac{i+k}{n}\right) - Q\left(\frac{i}{n}\right)\right),$$

where we recall that

$$\begin{aligned} i_{1,k,n} &= \min\left\{i : \frac{i+k}{n} \geq t_1\right\} \\ i_{2,n} &= \max\left\{i : \frac{i}{n} < t_2\right\}. \end{aligned}$$

Note that these definitions ensure that

$$(i_{1,k,n} + k - 1)/n < t_1 \leq (i_{1,k,n} + k)/n, \quad (1.68)$$

$$i_{2,n}/n < t_2 \leq (i_{2,n} + 1)/n. \quad (1.69)$$

Now, observe that, for all $h \in [h'_n, h''_n]$ and every $i_{1, \lceil nh \rceil, n} \leq i \leq i_{2,n}$,

$$\forall t \in \left(\frac{i-1}{n}, \frac{i}{n}\right], Q_n(t) = X_{(i)}, \quad (1.70)$$

$$\exists t \in \left(\frac{i-1}{n}, \frac{i}{n}\right], Q_n(t+h) = X_{(i+\lceil nh \rceil)}, \quad (1.71)$$

$$\nexists t \in \left(\frac{i-1}{n}, \frac{i}{n}\right], Q_n(t+h) = X_{(i+\lceil nh \rceil+1)}, \quad (1.72)$$

$$\exists t \in \left(\frac{i-1}{n}, \frac{i}{n}\right], Q_n(t-h) = X_{(i-\lceil nh \rceil)}, \quad (1.73)$$

$$\nexists t \in \left(\frac{i-1}{n}, \frac{i}{n}\right], Q_n(t-h) = X_{(i-\lceil nh \rceil-1)}. \quad (1.74)$$

In view of (1.63)–(1.74), the conclusion of Theorem 1.1.1 entails that

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \frac{|\sqrt{n} d_n^{(1)}(\lceil nh \rceil)|}{\sqrt{2h \log(1/h)}} = 1 \text{ a.s.} \quad (1.75)$$

To complete the proof of Theorem 1.3.2, first note that, under (F.1-2),

$$Q\left(\frac{i+k}{n}\right) - Q\left(\frac{i}{n}\right) = \frac{k}{n} q\left(\frac{i}{n} + h_1\right) \quad \text{for a given } h_1 \leq \frac{k}{n}. \quad (1.76)$$

Moreover, by (F.1-2), we have

$$\begin{aligned} f\left(Q\left(\frac{i}{n} + h_1\right)\right) &= f\left(Q\left(\frac{i}{n}\right) + \mathcal{O}(k/n)\right) \\ &= f\left(Q\left(\frac{i}{n}\right)\right) + \mathcal{O}(k/n). \end{aligned} \quad (1.77)$$

But, once again, a Taylor expansion based upon (F.1), together with the Chung (1949) LIL, as stated in (1.59), and (1.52), shows readily that, almost surely as $n \rightarrow \infty$,

$$\begin{aligned} f\left(Q\left(\frac{i}{n}\right)\right) &= f\left(Q_n\left(\frac{i}{n}\right) + \mathcal{O}\left(\sqrt{\frac{\log \log n}{n}}\right)\right) \\ &= f\left(X_{(i)}\right) + \mathcal{O}\left(\sqrt{\frac{\log \log n}{n}}\right). \end{aligned} \quad (1.78)$$

In view of (1.76), (1.77) and (1.78), we conclude by routine arguments that, under H .(1-2-3) and the assumptions above, (1.75) holds with the formal replacement of $d_n^{(1)}$ by d_n . This completes the proof of Theorem 1.3.2.

1.4.4 Proof of Theorem 1.3.1

Recall the definition (1.8) of β_n . Following the lines of the above-given proof of Theorem 1.3.2, we select $0 \leq s, t \leq 1$ and set $i = \lceil nt \rceil$ and $j = \lceil ns \rceil$. We then write

$$\beta_n(t) - \beta_n(s) = n^{1/2} \left(\left(U_{(j)} - U_{(i)} - \frac{j-i}{n} \right) + \frac{j-i}{n} - (t-s) \right), \quad (1.79)$$

and observe that, in (1.79),

$$\left| \frac{j-i}{n} - (t-s) \right| \leq \frac{1}{n}. \quad (1.80)$$

Looking carefully at the arguments used in the proof of Theorem 1.2.1, it is straightforward that the following results hold. Under the assumptions of Theorem 1.2.1, we have, almost surely,

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{0 \leq t \leq 1-h} \left(\sup_{0 \leq s \leq 1} \frac{|\zeta_n(h, t; s)|}{\sqrt{2h \log(1/h)}} \right) \right\} = 1, \quad (1.81)$$

and

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \sup_{h \leq t \leq 1} \left(\sup_{-1 \leq s \leq 0} \frac{|\zeta_n(h, t; s)|}{\sqrt{2h \log(1/h)}} \right) \right\} = 1. \quad (1.82)$$

In view of (1.79), (1.80), (1.81) and (1.82), we obtain the proof of (1.16) by similar arguments as in the proof of Theorem 1.3.2.

1.4.5 Proof of Theorem 1.3.3

The Parzen-Rosenblatt kernel estimator of the density function f (see, e.g., Parzen (1962) and Rosenblatt (1956)) is defined, for some kernel K fulfilling $(K.A-B)$, and a positive bandwidth h , by

$$\tilde{f}_{n,h}(x) := \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right).$$

Recall the notation $\tilde{J} = [u'_1, u'_2]$. An application of Theorem 1.1 of Varron (2005) (see Facts 1.5.1 and 1.5.2 in the Appendix) yields readily the following proposition whose proof is omitted.

Proposition 1.4.2. Denote by \mathcal{J} a sub-interval of $\tilde{\mathcal{J}}$ with non-empty interior. Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two non-random sequences fulfilling the conditions (H.1-2-3), with $0 < h'_n \leq h''_n < \infty$. Set, for $n \geq 1$, $k'_n = nh'_n$ and $k''_n = nh''_n$. Then, under (F.1-2-3-4) and (K.A-B), we have, almost surely,

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{x \in \mathcal{J}} \frac{\sqrt{nh} |\tilde{f}_{n,h}(x) - \mathbf{E}\tilde{f}_{n,h}(x)|}{\sqrt{2f(x) \log(1/h)}} = \left\{ \int_{\mathbf{R}} K^2(t) dt \right\}^{1/2}. \quad (1.83)$$

To complete the proof of Theorem 1.3.3, we select two sequences $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ fulfilling the conditions (H.1-2-3-4), and set $k'_n = nh'_n$ and $k''_n = nh''_n$. Given this notation, we claim that, almost surely as $n \rightarrow \infty$,

$$\sup_{x \in [u_{1,n}, u_{2,n}]} \sup_{k \in [k'_n, k''_n]} \left| R_k(x) - \frac{k}{nf(x)} \right| \rightarrow 0. \quad (1.84)$$

To see how (1.84) follows from Theorem 1.3.2, select $x \in [u_{1,n}, u_{2,n}]$ and let $j = j_k(x)$ be the smallest integer for which $X_{(j)} \geq x - R_k(x)/2$. Then, at least one of the relations

$$X_{(j)} = x - \frac{R_k(x)}{2}, \quad X_{(j+[k]-1)} = x + \frac{R_k(x)}{2}$$

holds. This naturally implies that

$$R_k(x) \leq \max \left\{ [X_{(j+[k])} - X_{(j)}], [X_{(j+[k]-1)} - X_{(j-1)}] \right\},$$

and

$$R_k(x) \geq [X_{(j+[k]-1)} - X_{(j)}].$$

Moreover, observe that, for all $x' \in [X_{(j+[k]-1)}, X_{(j+[k]+1)}]$, $f(x') = f(x) + o(1)$ almost surely as $n \rightarrow \infty$, for all $k \in [k'_n, k''_n]$. Putting all these results together, we can apply Theorem 1.3.2 and conclude to (1.84).

Introduce the modified sequences defined by

$$\tilde{h}'_n := \frac{h'_n}{2} \inf_{x \in \tilde{\mathcal{J}}} \frac{1}{f(x)} \quad \text{and} \quad \tilde{h}''_n := 2h''_n \sup_{x \in \tilde{\mathcal{J}}} \frac{1}{f(x)}. \quad (1.85)$$

We infer from (1.84) and (1.85) that, almost surely as $n \rightarrow \infty$,

$$\tilde{h}'_n \leq \inf_{x \in [u_{1,n}, u_{2,n}]} \inf_{k \in [k'_n, k''_n]} R_k(x) \leq \sup_{x \in [u_{1,n}, u_{2,n}]} \sup_{k \in [k'_n, k''_n]} R_k(x) \leq \tilde{h}''_n. \quad (1.86)$$

Recalling from (F.1-2) that f is bounded and positive on $\tilde{\mathcal{J}}$, with $\tilde{\mathcal{J}} \supset [u_{1,n}, u_{2,n}]$ for all $n \geq 1$, we see that $\{\tilde{h}'_n\}_{n \geq 1}$ [resp. $\{\tilde{h}''_n\}_{n \geq 1}$] fulfills (H.1-2-3-4), whenever such is the case for $\{h'_n\}_{n \geq 1}$ [resp. $\{h''_n\}_{n \geq 1}$]. Thus, it is straightforward that, under the hypotheses of Theorem 1.3.3,

$$\lim_{n \rightarrow \infty} \sup_{x \in [u_{1,n}, u_{2,n}]} \sup_{k \in [k'_n, k''_n]} \frac{\sqrt{k} |\hat{f}_{n,k}(x) - \mathbf{E}\hat{f}_{n,k}(x)|}{\sqrt{2f^2(x) \log(n/k)}} \leq \left\{ \int_{\mathbf{R}} K^2(t) dt \right\}^{1/2}. \quad (1.87)$$

Moreover, in view of (1.84), for any $k_n = nh_n \in [k'_n, k''_n]$, it holds that, almost surely as $n \rightarrow \infty$,

$$\inf_{x \in \tilde{\mathcal{J}}} \frac{1}{f(x)} \frac{h_n}{2} \leq \inf_{x \in [u_{1,n}, u_{2,n}]} R_{k_n}(x) \leq \sup_{x \in [u_{1,n}, u_{2,n}]} R_{k_n}(x) \leq 2h_n \sup_{x \in \tilde{\mathcal{J}}} \frac{1}{f(x)}. \quad (1.88)$$

We now compare (1.88) with the condition (B.1) in Deheuvels et Mason (2004). The fact that

$$\lim_{n \rightarrow \infty} \sup_{x \in [u_{1,n}, u_{2,n}]} \frac{\sqrt{k_n} |\widehat{f}_{n,k_n}(x) - \mathbb{E}\widehat{f}_{n,k_n}(x)|}{\sqrt{2f^2(x) \log(n/k_n)}} = \left\{ \int_{\mathbb{R}} K^2(t) dt \right\}^{1/2} \quad (1.89)$$

follows from Proposition 1.4.2, in view of (1.84) and (1.86), along the same lines as the first part of Theorem 1.2 in Deheuvels et Mason (2004) is shown to be a consequence of their Corollary 3.2. We omit the details of this book-keeping argument.

By combining (1.87) with (1.89), the proof of Theorem (1.3.3) is readily achieved.

1.5 Appendix

1.5.1 Useful facts

Here we present two necessary facts, that have been shown to be instrumental in our proofs.

Fix $d \geq 1$. Let $(\mathbf{Z}_i)_{i \geq 1}$ be a sequence of iid random variables defined on \mathbb{R}^d , and \mathcal{G} be a class of real Borel functions defined on \mathbb{R}^d . For each $h > 0$, $n \geq 1$ and $\mathbf{z} \in \mathbb{R}^d$, we set

$$G_n(K, h, \mathbf{z}) := \sum_{i=1}^n K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h^{1/d}}\right) - \mathbb{E}\left\{K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h^{1/d}}\right)\right\}.$$

Further set $I^d := [0, 1]^d$, $|u| := \max_{1 \leq i \leq d} |u_i|$ and

$$\mathcal{F} := \left\{ K(\lambda(\cdot - \mathbf{z})), \mathbf{z} \in \mathbb{R}^d, \lambda > 0, K \in \mathcal{G} \right\}.$$

Introduce the following assumptions on \mathcal{G} .

- (HK.I) (i) $\lim_{|u| \rightarrow 0} \sup_{K \in \mathcal{G}} \int_{\mathbb{R}^d} (K(\mathbf{x}) - K(\mathbf{x} + \mathbf{u}))^2 d\mathbf{x} = 0$.
(ii) $\lim_{\lambda \rightarrow 1} \sup_{K \in \mathcal{G}} \int_{\mathbb{R}^d} (K(\lambda \mathbf{x}) - K(\mathbf{x}))^2 d\mathbf{x} = 0$.
- (HK.II) (i) $\forall K \in \mathcal{G}, \sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| \leq 1$.
(ii) $\forall K \in \mathcal{G}, \forall x \in I^d |K(\mathbf{x})| = 0$.
- (HK.III) (i) $\exists C > 0, v > 0, \forall 0 < \varepsilon < 1, \mathcal{N}(\varepsilon, \mathcal{F}) \leq C\varepsilon^{-v}$.
(ii) \mathcal{F} is pointwise separable.

Here $\mathcal{N}(\varepsilon, \mathcal{F}) := \sup\{\mathcal{N}(\varepsilon, \mathcal{F}, \mathbb{L}_2(\mathbb{P})), \mathbb{P} \text{ probability measure}\}$ denotes the uniform covering number of \mathcal{F} for ε and the class of norms $\{\mathbb{L}_2(\mathbb{P})\}$, with \mathbb{P} varying in the set of all probability measures on \mathbb{R}^d (for more details we refer to van der Vaart et Wellner (1996)), pp 83-84). Let $\mathcal{B}(\mathcal{G})$ denote the set of all real bounded functions on \mathcal{G} , continuous with respect to pointwise convergence. Let $\mathbb{L}_2 \star (\mathcal{G})$ be the Hilbert subspace of $\mathbb{L}_2(\mathbb{R}^d, m)$ spanned by \mathcal{G} (Here, m denotes the Lebesgue measure on \mathbb{R}^d). For $f \in \mathbb{L}_2 \star (\mathbb{R}^d)$, set $\Psi_f(g) := (f, g)$, $g \in \mathcal{G}$ and $J(\Psi_f) := \int_{\mathbb{R}^d} f^2 dm$. For each $\Psi \in \mathcal{B}(\mathcal{G})$ of the form Ψ_f , $f \in \mathbb{L}_2 \star (\mathbb{R}^d)$, set $J(\Psi) = \infty$. Finally, set $\mathbb{K} := \{\Psi : J(\Psi) \leq 1\}$.

The following result is due to Varron (2005) (see also Varron (2006)).

Fact 1.5.1. *Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two sequences of positive constants, fulfilling the assumptions (H.1-2-3), and such that $0 < h'_n \leq h''_n < \infty$. Suppose that \mathbf{Z}_1 has a density*

function f , such that the following conditions hold. For some compact set $H \in \mathbb{R}^d$, there exists a $\varsigma > 0$ such that f is continuous and strictly positive on

$$H^\varsigma := \{\mathbf{x} \in \mathbb{R}^d : \inf_{\mathbf{z} \in H} \|\mathbf{x} - \mathbf{z}\|_{\mathbb{R}^d} \leq \varsigma\}.$$

Here, $\|\cdot\|_{\mathbb{R}^d}$ denotes the usual euclidian norm on \mathbb{R}^d . Then, if \mathcal{G} is a class of real Borel functions satisfying (HK.I-II-III), we have almost surely,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\mathbf{z} \in H} \sup_{h \in [h'_n, h''_n]} \inf_{\Psi \in \mathbb{K}} \left\| \frac{G_n(\cdot, \mathbf{z}, h)}{\sqrt{2f(\mathbf{z})nh \log(1/h)}} - \Psi \right\|_{\mathcal{G}} &= 0, \\ \forall \Psi \in \mathbb{K} \liminf_{n \rightarrow \infty} \inf_{\mathbf{z} \in H} \left\| \frac{G_n(\cdot, \mathbf{z}, h)}{\sqrt{2f(\mathbf{z})nh \log(1/h)}} - \Psi \right\|_{\mathcal{G}} &= 0, \end{aligned}$$

where $\|\psi\|_{\mathcal{G}} = \sup_{K \in \mathcal{G}} |\psi(K)|$.

A direct consequence of this result is as follows (see, e.g., Varron (2006)).

Fact 1.5.2. *Let K be a measurable kernel of bounded variation, with compact support. Whenever the sequences $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ both fulfills the assumptions (H.1-2-3), we have, almost surely,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in H} \frac{\sqrt{nh} |f_n(K, \mathbf{x}, h) - \mathbb{E}f_n(K, \mathbf{x}, h)|}{\sqrt{2f(\mathbf{x}) \log(1/h)}} = \sqrt{\int_{\mathbb{R}^d} K^2 dm},$$

where $f_n(K, \mathbf{x}, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\mathbf{Z}_i - \mathbf{x}}{h^{1/d}}\right) - \mathbb{E}\left\{K\left(\frac{\mathbf{Z}_i - \mathbf{x}}{h^{1/d}}\right)\right\}$.

Note that Fact 1.5.2 can be derived from Theorem 3.1.2 presented in **Chapter 3**.

Chapitre 2

Loi fonctionnelle uniforme du logarithme pour les incréments du processus empirique avec censure.

2.1 Présentation générale- Principaux résultats

2.1.1 Introduction- Notations

Dans ce chapitre, nous présentons des lois fonctionnelles uniformes pour les incréments du processus empirique de Kaplan-Meier définis en (2.5). Ces lois constituent une extension du théorème 1.2 de Deheuvels et Einmahl (2000) ; en effet, nous montrons que leur résultat principal (qui constitue une version, adaptée à la convergence uniforme sur un intervalle, d'un théorème de Deheuvels et Einmahl (1996) écrit dans le cadre de la convergence ponctuelle des estimateurs) reste valable *uniformément en fonction du choix de la fenêtre* $h \in [h'_n, h''_n]$, où h'_n et h''_n vérifient tous deux les conditions [CRS] (i.e., les conditions (H.1-2-3) rappelées ci-dessous), avec $0 < h'_n \leq h''_n < \infty$. Nous déduisons de ce théorème des résultats de convergence uniforme, relativement à la localisation et à la taille de fenêtre h , pour certaines fonctionnelles du processus empirique de Kaplan-Meier. Nous traiterons, en particulier, le cas des estimateurs de la *densité de survie*, définie en (2.2), et du *taux de mortalité* (ou de *panne*), défini en (2.15).

Nous supposerons disposer de copies aléatoires, indépendantes et de même loi, (Y_i, C_i) , $i = 1, 2, \dots$, du couple aléatoire générique (Y, C) . Ici, Y désigne la variable aléatoire d'intérêt, supposée positive, et représentant une *durée de survie*, et C la variable aléatoire, supposée positive elle aussi, représentant un *temps de censure*. Nous supposerons de plus disposer, sous l'hypothèse générale d'indépendance entre Y et C , de l'échantillon observé (Z_i, δ_i) , $i = 1, \dots, n$, $n \geq 1$, où

$$Z_i := Y_i \wedge C_i \quad \text{et} \quad \delta_i := \mathbb{1}_{\{Y_i \leq C_i\}}.$$

Ici, la notation $\mathbb{1}_E$ définit la fonction indicatrice de l'événement E . Désignons par $F(x) = \mathbb{P}(Y \leq x)$, $G(x) = \mathbb{P}(C \leq x)$ et $H(x) = \mathbb{P}(Z \leq x)$, les versions continues à droite des fonctions de répartition de X , Y et Z . Notons que, dans la suite du présent exposé, F , G et H ne seront pas nécessairement supposées continues. D'une manière générale, nous poserons, pour toute fonction R (et en particulier dans le cas de F , G , ...),

$$R_-(x) := \lim_{\varepsilon \downarrow 0} R(x - \varepsilon) \quad \text{et} \quad R_+(x) := \lim_{\varepsilon \downarrow 0} R(x + \varepsilon),$$

lorsque les limites correspondantes existent. Au vu de ces notations et définitions, on a $F_+(x) = F(x)$, $G_+(x) = G(x)$ et $H_+(x) = H(x)$. Pour toute fonction de répartition L , continue à droite, c'est à dire, telle que $L(x) = L_+(x)$, on définit le point supérieur de la distribution correspondante par

$$T_L = \sup\{t : L(t) < 1\},$$

et on étend la définition de L à l'infini, en posant $L_-(\infty) = \lim_{x \rightarrow \infty} L(x)$. On supposera que $F_-(\infty) = 1$, mais, par contre, on n'exclura pas le cas où la distribution de C est *défective*, au sens que $G_-(\infty) = 1 - \mathbb{P}(C = \infty) \leq 1$, l'inégalité pouvant être stricte. En particulier, si $\mathbb{P}(C = \infty) = 1$, alors $G(x) = 0$ pour tout $x < \infty$ et la censure est inactive car $Z = Y \wedge C = Y$ presque sûrement. Nos hypothèses de travail permettent, en englobant ce cas de figure, de prendre en compte la situation *non censurée* où F_n^* (défini par (4.4) ci-dessous) coïncide avec la fonction de répartition empirique classique.

On supposera par la suite que $\Theta := \min(T_F, T_G) > 0$. Pour $0 \leq z \leq \Theta$, les estimateurs de Kaplan-Meier (voir Kaplan et Meier (1958)) de F et de G sont définis par

$$\begin{aligned} F_n^*(z) &= 1 - \prod_{\substack{i: Z_i \leq z \\ 1 \leq i \leq n}} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)} \right)^{\delta_i}, \\ G_n^*(z) &= 1 - \prod_{\substack{i: Z_i \leq z \\ 1 \leq i \leq n}} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)} \right)^{1 - \delta_i}, \end{aligned} \quad (2.1)$$

où $N_n(x) = \sum_{i=1}^n \mathbb{1}_{\{Z_i \geq x\}}$. On utilise ici les conventions $\prod_{\emptyset} = 1$ et $0^0 = 1$.

Dans la suite de ce chapitre, on supposera que $h > 0$ est tel que $h'_n \leq h \leq h''_n$. Ici, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ sont deux suites de constantes réelles positives, vérifiant $0 < h'_n \leq h''_n < \infty$. Nous supposons que $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ sont telles que, pour chacun des choix $h_n = h'_n$ et $h_n = h''_n$, la suite $\{h_n\}_{n \geq 1}$ vérifie les hypothèses (H.1-2-3) ci-dessous.

(H.1) $h_n \downarrow 0$ et $nh_n \uparrow \infty$ lorsque $n \uparrow \infty$;

(H.2) $\log(1/h_n) / \log \log n \rightarrow \infty$ lorsque $n \rightarrow \infty$;

(H.3) $nh_n / \log n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

Afin de définir un estimateur à noyau $f_{n,h}^*$ de la densité de survie, on se donne une *fenêtre* $h > 0$ et on introduit un *noyau* K , fonction mesurable réelle de variable réelle, vérifiant les hypothèses suivantes.

(K1) K est une fonction à variation bornée sur \mathbb{R} .

(K2) Il existe une constante $0 < T < \infty$, telle que $K(u) = 0$ pour $|u| \geq T/2$.

(K3) $\int_{-\infty}^{\infty} K(u) du = 1$

Nous travaillerons sous les hypothèses de régularité suivantes, portant sur F et G . Soient des constantes a, a', b et b' telles que $0 < a' < a < b < b' < \Theta$. Soit, de plus, $H^{(1)}(b) = \mathbb{P}(Z \leq b, \delta = 1)$ (voir (2.17) plus bas). On suppose que

(F1) $F(0) = G(0) = 0$;

(F2) F et G sont continues sur $[a', b']$;

(F3) $f = (d/dx)F(x)$ existe, est continue et strictement positive sur $[a', b']$;

(F4) $h''_n \leq [(b' - b) \wedge (1 - H^{(1)}(b))]$, pour tout $n \geq 1$.

Notons que l'hypothèse (F4) est introduite ici pour des raisons similaires à celles qui nous avaient poussés à introduire l'hypothèse (F.4) dans le **chapitre 1**.

Pour tout $x \in \mathbb{R}$, l'estimateur à noyau $f_{n,h}^*(x)$ de $f(x) = (d/dx)F(x)$, associé au noyau K et à la fenêtre h (voir Watson et Leadbetter (1964a), Watson et Leadbetter (1964b), Tanner et Wong (1983), Deheuvels et Einmahl (1996), Deheuvels et Einmahl (2000)), est défini par

$$f_{n,h}^*(x) = \int_{-\infty}^{\infty} h^{-1}K\left(\frac{t-x}{h}\right)dF_n^*(t). \quad (2.2)$$

Pour tout $x \in \mathbb{R}$, on pose

$$\widehat{\mathbb{E}}f_{n,h}^*(x) = \int_{-\infty}^{\infty} h^{-1}K\left(\frac{t-x}{h}\right)dF(t). \quad (2.3)$$

Remarquons que dans le cas non censuré, $\widehat{\mathbb{E}}f_{n,h}^*(x) = \mathbb{E}f_{n,h}^*(x)$, où \mathbb{E} désigne l'espérance usuelle. Cependant, $\widehat{\mathbb{E}}f_{n,h}^*(x)$ et $\mathbb{E}f_{n,h}^*(x)$ sont généralement distinctes dans le cas censuré strict (correspondant à $\mathbb{P}(C < T_F) > 0$).

Par analogie au cas non censuré, on peut définir le processus empirique de Kaplan-Meier a_n^* par

$$a_n^*(t) := n^{1/2}(F_n^*(t) - F(t)), \quad (2.4)$$

pour $0 \leq t \leq \Theta$. Nous introduisons ensuite les incréments de ce processus, en posant

$$\eta_n^*(h, t; s) := a_n^*(t + hs) - a_n^*(t), \quad (2.5)$$

pour $0 \leq t \leq \Theta$, $0 \leq s \leq 1$ et $0 \leq h \leq \Theta - t$.

Les définitions (2.2), (2.3) et (2.5), compte tenu des hypothèses (K1-2-3), permettent d'écrire

$$\begin{aligned} f_{n,h}^*(x) - \widehat{\mathbb{E}}f_{n,h}^*(x) &= h^{-1} \int_{-T/2}^{T/2} K(u)d\{F_n^*(x + hu) - F_n^*(x) - F(x + hu) + F(x)\} \\ &= -h^{-1} \int_{-T/2}^{T/2} \{F_n^*(x + hu) - F_n^*(x) - F(x + hu) + F(x)\}dK(u) \\ &= -h^{-1}n^{-1/2} \int_{-T/2}^{T/2} \eta_n^*(h, x; u)dK(u). \end{aligned} \quad (2.6)$$

Cette relation va nous permettre d'obtenir des résultats de convergence pour l'estimateur de la densité de survie à partir des résultats obtenus sur le processus η_n^* . Une illustration de ce principe sera donnée dans la démonstration du théorème 2.1.6 ci-après.

Dans la suite de ce chapitre, on désignera par Ψ une fonction fixée, continue et strictement positive sur $[a', b']$. On supposera disposer d'un estimateur Ψ_n de Ψ tel que

$$(C1) \quad \sup_{a \leq x \leq b} |\Psi_n(x)/\Psi(x) - 1| \rightarrow 0, \text{ presque sûrement lorsque } n \rightarrow \infty.$$

Comme ailleurs dans le présent mémoire, nous désignerons, respectivement, par $(\mathcal{B}[0, 1], \mathcal{U})$ et $(AC[0, 1], \mathcal{U})$ l'ensemble des fonctions ℓ bornées et l'ensemble des fonctions ℓ absolument continues sur $[0, 1]$, munis, tous deux, de la topologie uniforme \mathcal{U} , définie par la norme uniforme $\|\ell\| = \sup_{0 \leq s \leq 1} |\ell(s)|$. Pour tout $\ell \in \mathcal{B}[0, 1]$, nous posons

$$|\ell|_H = \begin{cases} \left\{ \int_0^1 \ell^2(s)ds \right\}^{1/2}, & \text{si } \ell \in AC[0, 1] \text{ et } \ell(0) = 0, \\ \infty, & \text{sinon,} \end{cases}$$

où, comme ailleurs dans cette thèse, pour chaque $\ell \in AC[0, 1]$, $\dot{\ell} = (d/ds)\ell$ représente la dérivée de Lebesgue de ℓ .

Nous introduisons les ensembles de fonctions

$$\mathbb{S}_\eta = \{\ell \in AC[0, 1] : \ell(0) = 0, |\ell|_H^2 \leq \eta\}, \quad (2.7)$$

pour tout $\eta > 0$. Notons que $\mathbb{S} := \mathbb{S}_1$ est l'ensemble de Strassen (voir Strassen (1964)) introduit dans le **chapitre 1** (lorsque l'on considérait des fonctions de $AC[-1, 1]$; il était alors noté \mathbb{S}_0). Par ailleurs, $\mathbb{S}_\eta = \eta^{1/2}\mathbb{S}$, avec $\gamma\mathcal{G} := \{\gamma\ell : \ell \in \mathcal{G}\}$.

2.1.2 Résultats de la littérature

Après avoir introduit, dans le §2.1.1, les définitions, notations et hypothèses nécessaires, nous pouvons maintenant faire une brève synthèse des principaux résultats de convergence, connus à ce jour, pour les incréments du processus de Kaplan-Meier, définis, plus haut, par (2.5). Soit une suite de constantes positives $\{h_n\}_{n \geq 1}$. Au vu de (2.5), on définit, pour $x_0 \in (a, b)$ et $n \geq 1$, le sous-ensemble de fonctions

$$\mathcal{E}_n^*(\Psi_n) := \left\{ \frac{\pm \eta_n^*(h_n, x_0; I)}{\sqrt{2h_n \log_2 n}} \left(\Psi_n(x_0) \frac{1 - G(x_0)}{f(x_0)} \right)^{1/2} \right\} \subseteq \mathcal{B}[0, 1], \quad (2.8)$$

où, d'une manière générale $\log_2(v) := \log_+ \log_+(v)$ et $\log_+(v) = \log(v \vee e)$, pour $v \in \mathbb{R}$. On définit également le sous-ensemble de fonctions

$$\mathcal{L}_n^*(\Psi_n) := \left\{ \frac{\pm \eta_n^*(h_n, x; I)}{\sqrt{2h_n \log(1/h_n)}} \left(\Psi_n(x) \frac{1 - G(x)}{f(x)} \right)^{1/2} : a \leq x \leq b \right\} \subseteq \mathcal{B}[0, 1]. \quad (2.9)$$

Le théorème ci-dessous est une version simplifiée du théorème 1.2 de Deheuvels et Einmahl (1996).

Théorème 2.1.1. *Soit $\{h_n\}_{n \geq 1}$ une suite de constantes vérifiant les conditions (H.1-2-3). On suppose vérifiées les conditions (F1-2-3-4) et (C1). Alors la suite de fonctions $(\mathcal{E}_n^*(\Psi_n), n \geq 1)$ est presque sûrement relativement compacte dans $\mathcal{B}[0, 1]$, avec, comme ensemble limite, l'ensemble \mathbb{S}_{M_0} , où $M_0 = \Psi(x_0)$.*

En utilisant la relation (2.6), on déduit du théorème 2.1.1 le résultat suivant pour l'estimateur f_{n, h_n}^* de la densité de survie f (selon la définition (2.2)).

Théorème 2.1.2. *Soit $\{h_n\}_{n \geq 1}$ une suite vérifiant les conditions (H.1-2-3). On suppose vérifiées les conditions (F1-2-3-4), (K1-2-3) et (C1). Alors, pour tout $x_0 \in [a, b]$ fixé, avec probabilité 1,*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left\{ \frac{nh_n}{2 \log_2 n} \right\}^{1/2} \pm (f_{n, h_n}^*(x_0) - \widehat{\mathbf{E}} f_{n, h_n}^*(x_0)) \left\{ \Psi_n(x_0) \times \frac{1 - G(x_0)}{f(x_0)} \right\}^{1/2} \\ = \{\Psi(x_0)\}^{1/2} \left\{ \int_{-\infty}^{\infty} K^2(u) du \right\}^{1/2}. \end{aligned} \quad (2.10)$$

Dans Deheuvels et Einmahl (2000), il est montré que l'on peut obtenir des résultats similaires (mais avec des constantes de normalisation différentes), pour la convergence *uniforme en $x \in [a, b]$* de $f_{n, h_n}^*(x)$. Deheuvels et Einmahl établissent ainsi les deux théorèmes suivants (le théorème 2.1.4 étant, là encore, obtenu comme conséquence directe du théorème 2.1.3, à travers la relation (2.6)).

Théorème 2.1.3. Soit $\{h_n\}_{n \geq 1}$ une suite de constantes positives vérifiant les conditions (H.1-2-3). On suppose vérifiées les conditions (F1-2-3-4) et (C1). Alors la suite $(\mathcal{L}_n^*(\Psi_n), n \geq 1)$ est presque sûrement relativement compacte dans $\mathcal{B}[0, 1]$, avec comme ensemble limite l'ensemble \mathbb{S}_M , où $M = \sup_{a \leq x \leq b} \Psi(x)$.

Théorème 2.1.4. Soit $\{h_n\}_{n \geq 1}$ une suite de constantes positives vérifiant les conditions (H.1-2-3). On suppose vérifiées les conditions (F1-2-3-4), (K1-2-3) et (C1). Alors, avec probabilité 1,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{a \leq x \leq b} \pm (f_{n,h_n}^*(x) - \widehat{\mathbb{E}} f_{n,h_n}^*(x)) \left\{ \Psi_n(x) \times \frac{1-G(x)}{f(x)} \right\}^{1/2} \\ &= \sup_{a \leq x \leq b} \{ \Psi(x) \}^{1/2} \left\{ \int_{-\infty}^{\infty} K^2(u) du \right\}^{1/2}. \end{aligned} \quad (2.11)$$

2.1.3 Nouveaux résultats

Dans cette partie, nous présentons des généralisations nouvelles des théorèmes 2.1.3 et 2.1.4. Nous montrons, en effet, que ces résultats demeurent valables, *uniformément en* $h \in [h'_n, h''_n]$. Ici, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ désignent des suites de constantes, vérifiant les conditions (H.1-2-3) et telles que $0 < h'_n \leq h''_n < \infty$, $n \geq 1$. En sus de ces premiers résultats, nous établissons un théorème de convergence dans le même esprit pour l'estimateur du taux de mortalité (ou de panne).

Introduisons le sous-ensemble de fonctions de $\mathcal{B}[0, 1]$ défini par

$$\begin{aligned} \mathcal{K}_n^*(\Psi_n) &:= \left\{ \frac{\pm \eta_n^*(h, x; I)}{\sqrt{2h \log(1/h)}} \left(\Psi_n(x) \frac{1-G(x)}{f(x)} \right)^{1/2} \right. \\ &\quad \left. : a \leq x \leq b, h \in [h'_n, h''_n] \right\} \subseteq \mathcal{B}[0, 1]. \end{aligned} \quad (2.12)$$

Notre premier théorème est le suivant.

Théorème 2.1.5. Soit deux suites de constantes positives, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$, vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n$. Sous les conditions (F1-2-3-4) et (C1), la suite $(\mathcal{K}_n^*(\Psi_n), n \geq 1)$ est presque sûrement relativement compacte dans $\mathcal{B}[0, 1]$ (muni de la topologie de la convergence uniforme), avec comme ensemble limite l'ensemble \mathbb{S}_M , où $M = \sup_{a \leq x \leq b} \Psi(x)$.

Remarque 2.1.1. Dans le cas non censuré, correspondant à $\mathbb{P}(C < T_F) = 0$, $G(x) = 0$ pour tout $x \in [a, b]$. Le théorème 2.1.5 se réduit alors à une version du théorème 1.1 de Varron (2005).

Notre deuxième théorème, énoncé ci-dessous, est dans le même esprit que le précédent.

Théorème 2.1.6. Soit $\{h'_n\}_{n \geq 1}$, $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives, vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. On suppose vérifiées les conditions (F1-2-3-4), (K1-2-3) et (C1). Alors, avec probabilité un,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \frac{nh}{2 \log(1/h)} \right\}^{1/2} \sup_{a \leq x \leq b} \pm (f_{n,h}^*(x) - \widehat{\mathbb{E}} f_{n,h}^*(x)) \left\{ \Psi_n(x) \times \frac{1-G(x)}{f(x)} \right\}^{1/2} \\ &= \sup_{a \leq x \leq b} \{\Psi(x)\}^{1/2} \left\{ \int_{-\infty}^{\infty} K^2(u) du \right\}^{1/2}. \end{aligned} \quad (2.13)$$

Pour énoncer le résultat suivant, nous nous plaçons sous les hypothèses de régularité (F1-2-3-4). Celles-ci nous permettent de définir *le taux de mortalité*, ou *taux de panne*, ou encore, *taux de hasard instantané* (nous utiliserons dans cette thèse l'une ou l'autre de ces appellations) λ associé à F . Ce dernier est défini, pour tout $x < T_F$, par

$$\lambda(x) = \frac{f(x)}{1-F(x)}. \quad (2.14)$$

Cette quantité joue, notamment, un rôle central dans le modèle semi-paramétrique de Cox (très utilisé en épidémiologie). L'estimateur non-paramétrique $\lambda_{n,h}^*$ de λ , associé à l'estimateur de Kaplan-Meier, est défini par

$$\lambda_{n,h}^*(x) = \frac{f_{n,h}^*(x)}{1-F_n^*(x)}, \quad (2.15)$$

où $f_{n,h}^*$ et F_n^* sont les estimateurs respectifs de f et F , définis en (2.2) et (2.1). Dans la suite de ce chapitre, nous établirons le théorème suivant pour décrire la convergence de $\lambda_{n,h}^*(x)$.

Théorème 2.1.7. Soit $\{h'_n\}_{n \geq 1}$, $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. On suppose vérifiées les conditions (F1-2-3-4), (K1-2-3) et (C1). Alors, avec probabilité un,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \frac{nh}{2 \log(1/h)} \right\}^{1/2} \sup_{a \leq x \leq b} \pm \left(\lambda_{n,h}^*(x) - \frac{\widehat{\mathbb{E}} f_{n,h}^*(x)}{1-F(x)} \right) \\ & \times \left\{ \Psi_n(x) \times \frac{1-H(x)}{\lambda(x)} \right\}^{1/2} = \sup_{a \leq x \leq b} \{\Psi(x)\}^{1/2} \left\{ \int_{-\infty}^{\infty} K^2(u) du \right\}^{1/2}. \end{aligned}$$

La suite du présent chapitre est consacrée aux démonstrations des théorèmes 2.1.5, 2.1.6 et 2.1.7. Nous verrons, en particulier, que le théorème 2.1.7 peut se déduire du théorème 2.1.6 par un raisonnement direct. Comme nous l'avons déjà évoqué, le théorème 2.1.6 peut être déduit du théorème 2.1.5, en faisant une utilisation adéquate de la relation (2.6). L'essentiel des difficultés se concentre donc dans la démonstration du théorème 2.1.5. Pour établir celle-ci, nous procéderons de la manière suivante. L'idée sous-jacente de cette démonstration est, d'une certaine manière, semblable à celle que nous avons utilisée au **chapitre 1**, pour établir le théorème 1.1.1 traitant les incréments du processus des quantiles). Pour l'établir, nous utiliserons les résultats sur le processus empirique uniforme obtenus dans le **chapitre 1**, et rappelés dans les faits 2.2.2 et 2.2.3, ceux-ci étant associés avec des lemmes d'approximation convenables. Ces derniers nous permettront de contrôler l'écart entre les oscillations du processus de Kaplan-Meier et celles du processus empirique uniforme.

2.2 Preuves des théorèmes 2.1.5, 2.1.6 et 2.1.7

2.2.1 Quelques notations supplémentaires

Observons, tout d'abord, que, sous l'hypothèse générale d'indépendance entre Y et C , on peut décomposer $H(x)$ de la manière suivante,

$$\begin{aligned} H(x) &= 1 - (1 - F(x))(1 - G(x)) \\ &= H^{(1)}(x) + H^{(0)}(x), \end{aligned} \quad (2.16)$$

où

$$H^{(1)}(x) := \mathbb{P}(Z \leq x, \delta = 1) = \int_0^x (1 - G_-(t))dF(t) = H_+^{(1)}(x), \quad (2.17)$$

et

$$H^{(0)}(x) := \mathbb{P}(Z \leq x, \delta = 0) = \int_0^x (1 - F(t))dG(t) = H_+^{(0)}(x). \quad (2.18)$$

Nous adopterons la notation

$$p = \mathbb{P}(\delta = 1) = \int_0^\infty (1 - G_-(t))dF(t) = H_-^{(1)}(\infty) = 1 - H_-^{(0)}(\infty). \quad (2.19)$$

Dans le cadre de nos hypothèses de travail (celles-ci permettant à la distribution de C d'être défective), p peut prendre n'importe quelle valeur dans $(0, 1]$, le cas *non censuré* correspondant à la valeur $p = 1$. Conformément aux définitions (2.17) et (2.18), $H^{(1)}(x)$ [resp. $H^{(0)}(x)$] croît de 0 à p [resp. $1 - p$] lorsque x croît de 0 à Θ . On peut donc définir les fonctions de quantile associées respectivement à $H^{(1)}$ et $H^{(0)}$ comme suit.

$$Q^{(1)}(s) = \inf\{x : H^{(1)}(x) \geq s\} \text{ pour } 0 < s < p, \quad (2.20)$$

$$Q^{(0)}(s) = \inf\{x : H^{(0)}(x) \geq s\} \text{ pour } 0 < s < 1 - p. \quad (2.21)$$

Pour tout $n \geq 1$, et après avoir introduit les quantités

$$\begin{aligned} N_n(x) &= \sum_{i=1}^n \mathbb{1}_{\{Z_i \geq x\}} = n(1 - H_{n-}(x)), \\ N_{n-}(x) &= \sum_{i=1}^n \mathbb{1}_{\{Z_i > x\}}, \end{aligned} \quad (2.22)$$

nous définissons les contreparties empiriques de H , $H^{(1)}$ et $H^{(0)}$ par

$$H_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{Z_i \leq x\}} = H_n^{(1)}(x) + H_n^{(0)}(x) = 1 - n^{-1}N_{n-}(x), \quad (2.23)$$

$$\begin{aligned} H_n^{(1)}(x) &= n^{-1} \sum_{i=1}^n \delta_i \mathbb{1}_{\{Z_i \leq x\}}, \quad \text{et} \\ H_n^{(0)}(x) &= n^{-1} \sum_{i=1}^n (1 - \delta_i) \mathbb{1}_{\{Z_i \leq x\}}. \end{aligned} \quad (2.24)$$

Soit Λ_n^* la contrepartie empirique de la fonction de taux de mortalité cumulé, définie par

$$\Lambda_n^*(x) = \int_0^x \frac{1}{1 - H_{n-}(u)} dH_n^{(1)}(u) = \Lambda_{n+}^*(x) \text{ pour } x \geq 0. \quad (2.25)$$

Rappelons que, pour $x \geq 0$, la fonction de taux de mortalité cumulé peut être définie par

$$\begin{aligned}\Lambda(x) &= \int_0^x \frac{1}{1 - F_-(u)} dF(u) = \int_0^x \frac{1 - G_-(u)}{1 - H_-(u)} dF(u) \\ &= \int_0^x \frac{1}{1 - H_-(u)} dH^{(1)}(u).\end{aligned}\quad (2.26)$$

Les estimateurs non-paramétriques de Kaplan-Meier, F_n^* et G_n^* , de F et G , respectivement, basés sur l'échantillon $\{(Z_i, \delta_i) : 1 \leq i \leq n\}$ peuvent s'exprimer de la manière suivante (voir, par exemple, Shorack et Wellner (1986) p.295). On a, pour tout $x \geq 0$,

$$\begin{aligned}F_n^*(x) &= 1 - \prod_{\substack{i: Z_i, n \leq x \\ 1 \leq i \leq n}} \left(1 - \frac{\delta_{i,n}}{n - i + 1}\right) = \int_0^x (1 - F_{n-}^*(u)) d\Lambda_n^*(u) \\ &= \int_0^x \frac{1 - F_{n-}^*(u)}{1 - H_{n-}(u)} dH_n^{(1)}(u) = \int_0^x \frac{1}{1 - G_{n-}^*(u)} dH_n^{(1)}(u).\end{aligned}\quad (2.27)$$

Et, de façon analogue,

$$\begin{aligned}G_n^*(x) &= 1 - \prod_{\substack{i: Z_i, n \leq x \\ 1 \leq i \leq n}} \left(1 - \frac{1 - \delta_{i,n}}{n - i + 1}\right) \\ &= \int_0^x \frac{1}{1 - F_n^*(u)} dH_n^{(0)}(u).\end{aligned}\quad (2.28)$$

Introduisons les processus empiriques, $a_n^* = a_{n,+}^*$ et $b_n^* = b_{n,+}^*$, associés à ces quantités, et définis, pour tout $x \in [a', b']$, par

$$\begin{aligned}a_{n\pm}^*(x) &:= n^{1/2}(F_{n\pm}^*(x) - F_{\pm}(x)) \text{ et} \\ b_{n\pm}^*(x) &:= n^{1/2}(G_{n\pm}^*(x) - G_{\pm}(x)).\end{aligned}\quad (2.29)$$

Dans le même esprit, posons

$$\mathcal{H}_n^{(j)}(x) := n^{1/2}(H_n^{(j)}(x) - H^{(j)}(x)) \quad \text{pour } j = 0, 1 \text{ et } x \in [a', b']. \quad (2.30)$$

2.2.2 Un résultat d'approximation utile

Observons, en premier lieu, que l'on peut décomposer $a_n^*(x)$ de la façon suivante. Pour tout $x \in [a', b']$, nous avons la relation

$$\begin{aligned}a_n^*(x) &= n^{1/2}(F_n^*(x) - F(x)) = n^{1/2} \left\{ \int_0^x dF_n^*(u) - \int_0^x dF(u) \right\} \\ &= n^{1/2} \left\{ \int_0^x \frac{1}{1 - G_{n-}^*(u)} dH_n^{(1)}(u) - \int_0^x \frac{1}{1 - G_{n-}^*(u)} dH^{(1)}(u) \right. \\ &\quad \left. + \int_0^x \frac{1 - G_-(u)}{1 - G_{n-}^*(u)} dF(u) - \int_0^x \frac{1 - G_{n-}^*(u)}{1 - G_{n-}^*(u)} dF(u) \right\} \\ &= \int_0^x \frac{1}{1 - G_{n-}^*(u)} d\mathcal{H}_n^{(1)}(u) + \int_0^x \frac{b_{n-}^*(u)}{1 - G_{n-}^*(u)} dF(u) =: a_n'^*(x) - a_n''^*(x).\end{aligned}\quad (2.31)$$

La formule (2.31) nous permet d'écrire que

$$\eta_n^*(h, t; s) = \{a_n'^*(t + hs) - a_n'^*(t)\} - \{a_n''^*(t + hs) - a_n''^*(t)\}. \quad (2.32)$$

Le lemme suivant a été obtenu par Deheuvels et Einmahl (2000). Grâce à celui-ci, nous pourrons montrer par la suite que, sous les conditions (H.1-2-3), les oscillations du processus $a_n''^*$ sont négligeables par rapport à celles de $a_n'^*$.

Lemme 2.2.1. *Soit R fixé, tel que $0 < R < \Theta$. Supposons qu'il existe une version $f(x) = (d/dx)F(x)$, de la dérivée de Lebesgue de F , uniformément bornée sur $[0, R]$. Alors il existe une constante $C_1(R) < \infty$, telle que, presque sûrement pour tout n suffisamment grand, on ait, uniformément en $0 \leq s, t \leq R$,*

$$\frac{|a_n''^*(t) - a_n''^*(s)|}{|t - s|} = \frac{1}{|t - s|} \left| \int_s^t \frac{b_{n-}^*(u)}{1 - G_{n-}^*(u)} dF(u) \right| \leq C_1(R)(\log_2(n))^{1/2}. \quad (2.33)$$

Démonstration. Omise. \square

Pour prendre la mesure de l'amplitude des oscillations du processus $a_n'^*$, il sera commode d'introduire la quantité $A_{n,1}$, définie par

$$\begin{aligned} A_{n,1}(s, t) &:= a_n'^*(s) - a_n'^*(t) - \frac{1}{1 - G_-(s)} \int_s^t d\mathcal{H}_n^{(1)}(u) \\ &= \int_s^t \left(\frac{1}{1 - G_{n-}^*(u)} - \frac{1}{1 - G_-(u)} \right) d\{\mathcal{H}_n^{(1)}(u) - \mathcal{H}_n^{(1)}(s)\} \\ &= \left(\frac{1}{1 - G_{n-}^*(t)} - \frac{1}{1 - G_-(s)} \right) \{\mathcal{H}_n^{(1)}(t) - \mathcal{H}_n^{(1)}(s)\} \\ &\quad - \int_s^t \{\mathcal{H}_n^{(1)}(u) - \mathcal{H}_n^{(1)}(s)\} d\left\{ \frac{1}{1 - G_{n-}(u)} \right\}. \end{aligned} \quad (2.34)$$

Par la suite, nous travaillerons dans l'espace de probabilité invoqué dans le fait suivant, établi par Deheuvels et Einmahl (1996).

Fait 2.2.1. *Sur un espace de probabilité convenablement élargi $(\Omega, \mathcal{A}, \mathbb{P})$, il est possible de définir $\{Y_n : n \geq 1\}$ et $\{C_n : n \geq 1\}$, conjointement à une suite i.i.d. $\{U_n : n \geq 1\}$ de variables aléatoires, de loi uniforme sur $(0, 1)$, telle que les propriétés suivantes soient vérifiées. Pour $n \geq 1$ et $s \in \mathbb{R}$, soit*

$$\begin{aligned} U_n(s) &= n^{-1} \sum_{i=1}^n \mathbb{I}_{\{U_i \leq s\}}, \text{ et} \\ \alpha_n(s) &= n^{1/2}(U_n(s) - s). \end{aligned} \quad (2.35)$$

On a, presque sûrement,

$$H_n^{(1)}(x) = U_n(H^{(1)}(x)) \quad \text{pour } 0 < H^{(1)}(x) < p, \quad (2.36)$$

et

$$H_n^{(0)}(x) = U_n(H^{(0)}(x) + p) - U_n(p) \quad \text{pour } 0 < H^{(0)}(x) < 1 - p. \quad (2.37)$$

Pour la suite de notre démonstration, nous aurons besoin des notations additionnelles suivantes. Posons

$$\begin{aligned} w_n^{(1)}(h) &:= \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \left| \mathcal{H}_n^{(1)}(t) - \mathcal{H}_n^{(1)}(s) \right| \\ &= \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \left| \alpha_n(H^{(1)}(t)) - \alpha_n(H^{(1)}(s)) \right|, \end{aligned} \quad (2.38)$$

et

$$\varpi_n^{(1)} := \sup_{h \in [h'_n, h''_n]} \frac{w_n^{(1)}(h)}{\tau_h}, \quad (2.39)$$

avec

$$\tau_h := \sqrt{2h \log(1/h)}. \quad (2.40)$$

Nous pouvons maintenant énoncer un lemme permettant d'évaluer le comportement uniforme asymptotique de $A_{n,1}(s, t)$.

Lemme 2.2.2. *Sous les conditions (F1-2-4), il existe une suite de constantes $C_{4,n}$, telle que $C_{4,n} \rightarrow 0$ lorsque $n \rightarrow \infty$, et une constante $C_5 > 0$, de telle sorte que, presque sûrement pour tout n suffisamment grand,*

$$\sup_{h \in [h'_n, h''_n]} \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \tau_h^{-1} |A_{n,1}(s, t)| \leq \varpi_n^{(1)} \times \left\{ C_5 \sqrt{\frac{\log_2(n)}{n}} + C_{4,n} \right\}. \quad (2.41)$$

Démonstration. D'après la loi du logarithme itéré de Földes et Rejtő (1981), il existe une constante $C_5 \in \mathbb{R}$, telle que, presque sûrement pour tout n suffisamment grand,

$$\sup_{a \leq t \leq b'} \left| \frac{1}{1 - G_{n-}^*(t)} - \frac{1}{1 - G_-(t)} \right| \leq \frac{C_5}{3} \sqrt{\frac{\log_2(n)}{n}}. \quad (2.42)$$

De plus, d'après la condition de continuité (F2), portant sur G , nous pouvons écrire, au vu de (F1-4), que

$$C_{4,n} := 2 \sup_{h \in [h'_n, h''_n]} \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \left| \frac{1}{1 - G_-(t)} - \frac{1}{1 - G_-(s)} \right| \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty. \quad (2.43)$$

Ainsi, en combinant la relation (2.39) avec les relations (2.42) et (2.43), nous obtenons que

$$\begin{aligned} & \sup_{h \in [h'_n, h''_n]} \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \tau_h^{-1} \left| \frac{1}{1 - G_{n-}^*(t)} - \frac{1}{1 - G_-(s)} \left\{ \mathcal{H}_n^{(1)}(t) - \mathcal{H}_n^{(1)}(s) \right\} \right| \\ & \leq \varpi_n^{(1)} \sup_{h \in [h'_n, h''_n]} \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \left| \frac{1}{1 - G_{n-}^*(t)} - \frac{1}{1 - G_-(t)} + \frac{1}{1 - G_-(t)} - \frac{1}{1 - G_-(s)} \right| \\ & \leq \varpi_n^{(1)} \left\{ \frac{C_5}{3} \sqrt{\frac{\log_2(n)}{n}} + \frac{C_{4,n}}{2} \right\}. \end{aligned} \quad (2.44)$$

Par ailleurs, nous pouvons écrire que

$$\begin{aligned} & \sup_{h \in [h'_n, h''_n]} \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \tau_h^{-1} \left| \int_s^t \left\{ \mathcal{H}_n^{(1)}(u) - \mathcal{H}_n^{(1)}(s) \right\} d \left\{ \frac{1}{1 - G_{n-}^*(u)} \right\} \right| \\ & \leq \varpi_n^{(1)} \sup_{h \in [h'_n, h''_n]} \sup_{\substack{a \leq s, t \leq b' \\ |t-s| \leq h}} \left| \frac{1}{1 - G_{n-}^*(t)} - \frac{1}{1 - G_{n-}^*(s)} \right| \\ & \leq \varpi_n^{(1)} \times \left\{ \frac{2C_5}{3} \sqrt{\frac{\log_2(n)}{n}} + \frac{C_{4,n}}{2} \right\}. \end{aligned} \quad (2.45)$$

La conclusion du lemme 2.2.2 s'obtient alors immédiatement en combinant (2.34) avec (2.44) et (2.45). \square

Considérons maintenant la quantité $\eta_n^{*(1)}(h, t; s)$, définie par

$$\begin{aligned} \eta_n^{*(1)}(h, t; s) &:= \frac{1}{1 - G_-(t)} \left\{ \mathcal{H}_n^{(1)}(t + hs) - \mathcal{H}_n^{(1)}(t) \right\} \\ &= \frac{1}{1 - G_-(t)} \left\{ \alpha_n(H^{(1)}(t + hs)) - \alpha_n(H^{(1)}(t)) \right\}. \end{aligned} \quad (2.46)$$

On déduit des lemmes 2.2.1 et 2.2.2 le résultat d'approximation suivant. Dans ce dernier, les constantes C_5 et $C_{4,n}$ sont définies, comme dans le lemme 2.2.2, et la constante $C_1(R)$ est définie, comme dans le lemme 2.2.1, pour un $0 < R < \Theta$ fixé.

Lemme 2.2.3. *Sous les conditions (F1-2-3-4), il existe une suite de constantes $C_{4,n}$, telle que $C_{4,n} \rightarrow 0$ lorsque $n \rightarrow \infty$, et des constantes C_5 et $C_1 = C_1(b)$, telles que, presque sûrement pour tout n suffisamment grand,*

$$\begin{aligned} &\sup_{h \in [h'_n, h''_n]} \sup_{a \leq t \leq b} \tau_h^{-1} \|\eta_n^*(h, t; I) - \eta_n^{*(1)}(h, t; I)\| \\ &\leq C_1 \sqrt{\frac{h''_n \log_2(n)}{2 \log(1/h'_n)}} + \varpi_n^{(1)} \left(C_5 \sqrt{\frac{\log_2(n)}{n}} + C_{4,n} \right). \end{aligned} \quad (2.47)$$

Démonstration. En vertu des définitions (2.5), (2.31), (2.34) et (2.46), nous pouvons écrire que

$$\begin{aligned} \sup_{h \in [h'_n, h''_n]} \sup_{a \leq t \leq b} \|\eta_n^*(h, t; I) - \eta_n^{*(1)}(h, t; I)\| &\leq \sup_{h \in [h'_n, h''_n]} \sup_{a \leq t \leq b} \sup_{0 \leq s \leq 1} |a_n^{''*}(t + hs) - a_n^{''*}(t)| \\ &\quad + \sup_{h \in [h'_n, h''_n]} \sup_{a \leq t \leq b} \sup_{0 \leq s \leq 1} |A_{n,1}(t + hs, t)|. \end{aligned} \quad (2.48)$$

La conclusion du lemme est alors obtenue, en faisant un usage combiné des lemmes 2.2.1 et 2.2.2. \square

Pour établir le théorème 2.1.5, nous allons combiner le résultat d'approximation établi dans le lemme 2.2.3, avec des résultats connus de convergence uniforme du processus empirique uniforme (faits 2.2.2 et 2.2.3 ci-dessous).

2.2.3 Preuve du théorème 2.1.5

Citons, dans un premier temps, deux corollaires de la proposition 1.4.1 du **chapitre 1**. Posons, pour $h > 0$,

$$w_n(h) := \sup_{\substack{0 \leq s, t \leq 1 \\ |t-s| \leq h}} |\alpha_n(t) - \alpha_n(s)|, \quad (2.49)$$

Fait 2.2.2. *Soit deux suites de constantes positives, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$, vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. Pour tout $\gamma > 0$, on a, avec probabilité un,*

$$\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \tau_h^{-1} w_n(\gamma h) = \gamma^{1/2}. \quad (2.50)$$

D'autre part, rappelons la notation utilisée pour les incréments du processus empirique. On pose

$$\xi_n(h, t; s) = \alpha_n(t + hs) - \alpha_n(t). \quad (2.51)$$

Fait 2.2.3. *Soit deux suites de constantes positives, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. Pour tout couple de constantes réelles (c_1, c_2) tel que $0 \leq c_1 < c_2 \leq 1 - h''_n$, et tout $\gamma > 0$, on a presque sûrement,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{c_1 \leq t \leq c_2} \inf_{g \in \mathbb{S}_\gamma} \left\| \frac{\xi_n(\gamma h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| = 0, \quad (2.52)$$

et

$$\forall g \in \mathbb{S}_\gamma, \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \inf_{c_1 \leq t \leq c_2} \left\| \frac{\xi_n(\gamma h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| = 0. \quad (2.53)$$

En combinant le fait 2.2.2 et le lemme 2.2.3, on obtient le résultat suivant.

Lemme 2.2.4. *Soit deux suites de constantes positives, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$, vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F1-2-3-4), on a, avec probabilité 1,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \tau_h^{-1} \sup_{a \leq t \leq b} \|\eta_n^*(h, t; I) - \eta_n^{*(1)}(h, t; I)\| = 0. \quad (2.54)$$

Démonstration. Soit

$$D = \sup_{a \leq x \leq b'} \frac{dH^{(1)}(x)}{dx} = \sup_{a \leq x \leq b'} f(x)(1 - G(x)).$$

On a, uniformément en $a \leq s, t \leq b'$,

$$|H^{(1)}(t) - H^{(1)}(s)| \leq D|t - s|. \quad (2.55)$$

Ainsi, au vu de (2.38) et (2.49), on a, à partir d'un certain rang $[apcr]$, pour $h \in [h'_n, h''_n]$,

$$w_n^{(1)}(h) \leq w_n(Dh), \quad (2.56)$$

et donc, en utilisant le fait 2.2.2, et en rappelant la définition (2.39) de $\varpi_n^{(1)}$, on a, presque sûrement,

$$\limsup_{n \rightarrow \infty} \varpi_n^{(1)} \leq D^{1/2}. \quad (2.57)$$

Par ailleurs, d'après le lemme 2.2.3,

$$C_5 \sqrt{\frac{\log_2(n)}{n}} + C_{4,n} \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty. \quad (2.58)$$

Enfin, d'après les hypothèses (H.1) et (H.2) sur les suites $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$,

$$\sqrt{\frac{h''_n \log_2(n)}{2 \log(1/h'_n)}} = o(h''_n^{1/2}) \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty. \quad (2.59)$$

La preuve est finalisée en combinant le lemme 2.2.3 à ces trois derniers résultats. \square

Soit $N \geq 1$ fixé. Considérons maintenant la discrétisation de l'intervalle $[a, b]$ suivante. Pour $1 \leq i \leq N$, soit $t_{i,N} = a + (i-1)(b-a)N^{-1}$ et $\gamma_{i,N} = f(t_{i,N})(1 - G(t_{i,N}))$. On pose

$$\begin{aligned} \eta_{n,N}^{\star(1)}(h, t; s) &:= \frac{1}{1 - G(t)} \left\{ a_n \left(H^{(1)}(t) + shf(t_{i,N})(1 - G(t_{i,N})) \right) - a_n \left(H^{(1)}(t) \right) \right\} \\ &= \frac{1}{1 - G(t)} \xi_n(\gamma_{i,N}h, H^{(1)}(t); s). \end{aligned} \quad (2.60)$$

Le lemme suivant montre que l'on peut approcher $\eta_n^{\star(1)}$ par $\eta_{n,N}^{\star(1)}$ avec un ordre suffisant pour nos besoins.

Lemme 2.2.5. *Soit deux suites de constantes, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$, vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F1-2-3-4), pour tout $\varepsilon > 0$, il existe presque sûrement un $N_0 = N_0(\varepsilon) < \infty$, tel que, pour tout $N \geq N_0$,*

$$\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \tau_h^{-1} \sup_{a \leq t \leq b} \|\eta_{n,N}^{\star(1)}(h, t; I) - \eta_n^{\star(1)}(h, t; I)\| \leq \varepsilon. \quad (2.61)$$

Démonstration. Soit

$$e_N = \max_{1 \leq i \leq N} \left(\sup_{h \in [h'_n, h''_n]} \sup_{t_{i,N} \leq t \leq t_{i+1,N} + h} \left| f(t)(1 - G(t)) - f(t_{i,N})(1 - G(t_{i,N})) \right| \right). \quad (2.62)$$

Pour tout $1 \leq i \leq N$, $h \in [h'_n, h''_n]$, $t \in [t_{i,N}, t_{i+1,N}]$ et $s \in [0, 1]$, il existe un $t_1 \in [t_{i,N}, t_{i+1,N} + h]$, tel que, à partir d'un certain rang,

$$\begin{aligned} &|H^{(1)}(t + hs) - \{H^{(1)}(t) + hsf(t_{i,N})(1 - G(t_{i,N}))\}| \\ &\leq hs |f(t_1)(1 - G(t_1)) - f(t_{i,N})(1 - G(t_{i,N}))| \leq e_N h. \end{aligned}$$

Au vu des définitions (2.46), (2.49) et (2.60), et en utilisant le résultat du fait 2.2.2, cette dernière inégalité implique que, presque sûrement,

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \tau_h^{-1} \sup_{a \leq t \leq b} \|\eta_{n,N}^{\star(1)}(h, t; I) - \eta_n^{\star(1)}(h, t; I)\| \\ &\leq \sup_{a \leq t \leq b} \left\{ \frac{1}{1 - G(t)} \right\} \limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \tau_h^{-1} w_n(e_N h) = \left\{ \frac{1}{1 - G(b)} \right\} e_N^{1/2}. \end{aligned} \quad (2.63)$$

Eu égard à la définition (2.62) de e_N , compte tenu de l'hypothèse (H.1), de la définition de b , ainsi que des hypothèses (F2-3) de continuité de f et G , il est possible de rendre le terme du membre de droite aussi petit que l'on veut en prenant N suffisamment grand. Ceci permet de conclure la démonstration du lemme 2.2.5. \square

Soit $R(\cdot)$, une fonction réelle, continue et strictement positive sur $[a', b']$. On pose

$$S_{i,N} = \inf_{t_{i,N} \leq t \leq t_{i+1,N}} \frac{R(t)}{1 - G(t)} \quad \text{et} \quad R_{i,N} = \sup_{t_{i,N} \leq t \leq t_{i+1,N}} \frac{R(t)}{1 - G(t)}. \quad (2.64)$$

Définissons, de plus,

$$\begin{aligned} \mathcal{M}_{N,1}^{1/2} &:= \max_{1 \leq i \leq N} R_{i,N} \gamma_{i,N}^{1/2} =: \max_{1 \leq i \leq N} \mathcal{M}_{i,N,1}^{1/2}, \\ \mathcal{M}_{N,2}^{1/2} &:= \max_{1 \leq i \leq N} S_{i,N} \gamma_{i,N}^{1/2} =: \max_{1 \leq i \leq N} \mathcal{M}_{i,N,2}^{1/2} \quad \text{et} \end{aligned} \quad (2.65)$$

$$\mathcal{M}^{1/2} := \sup_{a \leq t \leq b} \left\{ \frac{R(t)}{1 - G(t)} \right\} \sqrt{f(t)(1 - G(t))} = \sup_{a \leq t \leq b} R(t) \sqrt{\frac{f(t)}{1 - G(t)}}.$$

De manière évidente, on a, lorsque $N \rightarrow \infty$, pour $k = 1, 2$,

$$|\mathcal{M}_{N,k} - \mathcal{M}| \rightarrow 0. \quad (2.66)$$

Lemme 2.2.6. *Soit deux suites de constantes, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$, vérifiant les conditions (H.1-2-3), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F1-2-3-4), on a, presque sûrement,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{a \leq t \leq b} \inf_{g \in \mathbb{S}_{\mathcal{M}}} \left\| R(t) \frac{\eta_{n,N}^{*(1)}(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| = 0, \quad (2.67)$$

et

$$\forall g \in \mathbb{S}_{\mathcal{M}}, \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \inf_{a \leq t \leq b} \left\| R(t) \frac{\eta_{n,N}^{*(1)}(h, t; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| = 0. \quad (2.68)$$

Démonstration. Soit $\varepsilon > 0$. Pour tout $1 \leq i \leq N$ fixé, on pose $c_1 = H^{(1)}(t_{i,N})$ et $c_2 = H^{(1)}(t_{i+1,N})$. D'après la définition (2.17) et sous les hypothèses (F2-3), la fonction $H^{(1)}$ est strictement croissante sur $[a, b]$. On a donc, sous (F4),

$$0 \leq c_1 < c_2 \leq H^{(1)}(b) \leq 1 - h''_n.$$

Soit $\rho = \max_{1 \leq i \leq N} R_{i,N}$. D'après la relation (2.60), on a, pour tout $h \in [h'_n, h''_n]$ et tout $t_{i,N} \leq t \leq t_{i+1,N}$,

$$\mathcal{K}_{n,N,i}^{*(1)}(h, t; I) := R(t) \tau_h^{-1} \eta_{n,N}^{*(1)}(h, t; I) = \frac{R(t)}{1 - G(t)} \tau_h^{-1} \xi_n(\gamma_{i,N} h, H^{(1)}(t); I). \quad (2.69)$$

Ainsi, d'après le résultat (2.52), portant sur le processus empirique « classique » (i.e. dans le cas non censuré), il existe presque sûrement un $n_0 = n(\varepsilon, i, N)$ tel que, pour tout $n \geq n_0$,

$$\sup_{h \in [h'_n, h''_n]} \sup_{c_1 \leq u \leq c_2} \inf_{g \in \mathbb{S}_{\gamma_{i,N}}} \left\| \frac{\xi_n(\gamma_{i,N} h, u; \cdot)}{\sqrt{2h \log(1/h)}} - g \right\| \leq \varepsilon/(2\rho). \quad (2.70)$$

On a donc, presque sûrement,

$$\sup_{h \in [h'_n, h''_n]} \sup_{t_{i,N} \leq t \leq t_{i+1,N}} \inf_{g \in \mathbb{S}_{\mathcal{M}_{i,N,1}}} \left\| \mathcal{K}_{n,N,i}^{*(1)}(h, t; \cdot) - g \right\| \leq \varepsilon/2. \quad (2.71)$$

Désignons par κ la fonction qui, à tout $t \in [a, b]$ associe l'indice entier i tel que $t \in [t_{i,N}, t_{i+1,N}]$. On définit, pour tout $t \in [a, b]$,

$$\mathcal{K}_{n,N}^{*(1)}(h, t; I) := \mathcal{K}_{n,N,\kappa(t)}^{*(1)}(h, t; I) \text{ lorsque } t_{i,N} \leq t \leq t_{i+1,N}. \quad (2.72)$$

Comme $\bigcup_{i=1}^N \mathbb{S}_{\mathcal{M}_{i,N,1}} \subseteq \mathbb{S}_{\mathcal{M}_{N,1}}$, on en conclut que, presque sûrement,

$$\sup_{h \in [h'_n, h''_n]} \sup_{a \leq t \leq b} \inf_{g \in \mathbb{S}_{\mathcal{M}_{N,1}}} \left\| \mathcal{K}_{n,N}^{*(1)}(h, t; \cdot) - g \right\| \leq \varepsilon/2, \quad (2.73)$$

ce qui, au vu de (2.66), implique (2.67).

Pour établir (2.68), on sélectionne $g_i \in \mathbb{S}_{\gamma_{i,N}}$, pour un indice i tel que $1 \leq i \leq N$, et on fixe un $\varepsilon > 0$. De par le résultat (2.53), il existe, presque sûrement, un $t_0 \in [t_{i,N}, t_{i+1,N}]$, tel que

$$\sup_{h \in [h'_n, h''_n]} \left\| \frac{\xi_n(\gamma_{i,N} h, t_0; \cdot)}{\sqrt{2h \log(1/h)}} - g_i \right\| \leq \varepsilon/(2\rho), \quad (2.74)$$

et donc

$$\sup_{h \in [h'_n, h''_n]} \left\| S_{i,N} \frac{\xi_n(\gamma_{i,N} h, t_0; \cdot)}{\sqrt{2h \log(1/h)}} - S_{i,N} g_i \right\| \leq \varepsilon/2. \quad (2.75)$$

Comme, de plus, pour tout $g \in \mathbb{S}_{\mathcal{M}_{N,2}}$, il existe un indice $1 \leq j \leq N$, tel que $g = S_{j,N} \cdot g_j$ avec $g_j \in \mathbb{S}_{\gamma_j, N}$, on déduit de (2.75) que, pour tout $g \in \mathbb{S}_{\mathcal{M}_{N,2}}$, il existe presque sûrement un $t \in [a, b]$, tel que

$$\sup_{h \in [h'_n, h''_n]} \left\| \frac{\mathcal{K}_{n,N}^{*(1)}(h, t; I)}{\sqrt{2h \log(1/h)}} - g \right\| \leq \varepsilon/2. \quad (2.76)$$

Grâce à ce résultat et à (2.66), on obtient, naturellement, (2.68). \square

Pour achever la démonstration du théorème 2.1.5, observons, en premier lieu, que le choix de la fonction R défini par

$$R(t) = \{\Psi(t)(1 - G(t))/f(t)\}^{1/2},$$

conduit à la relation

$$\sup_{a \leq t \leq b} R(t) = \sup_{a \leq t \leq b} (\Psi(t))^{1/2}.$$

Ainsi, compte tenu des définitions (2.12) et (2.65), le résultat du théorème 2.1.5, correspondant au cas '++', est obtenu par une combinaison directe des lemmes 2.2.4, 2.2.5 et 2.2.6. La preuve dans le cas '--', obtenue en suivant le même principe, est omise.

2.2.4 Preuve du théorème 2.1.6

La preuve du théorème 2.1.6 repose principalement sur le théorème 2.1.5, grâce à l'égalité (2.6), et en faisant usage de la proposition suivante (cf. Deheuvels et Einmahl (2000)). Soit $(\mathcal{E}, \mathcal{T})$ un ensemble \mathcal{E} de fonctions, muni d'une topologie métrisable \mathcal{T} .

Proposition 2.2.1. *Soit \mathcal{B}_n une suite de fonctions de $(\mathcal{E}, \mathcal{T})$, presque sûrement relativement compacte, et ayant comme ensemble limite \mathcal{B} . Soit de plus $\Gamma : \mathcal{E} \rightarrow \mathbb{R}$ une fonctionnelle \mathcal{T} -continue. Alors on a*

$$\lim_{n \rightarrow \infty} \left\{ \sup_{\ell \in \mathcal{B}_n} \Gamma(\ell) \right\} = \sup_{\ell \in \mathcal{B}} \Gamma(\ell). \quad (2.77)$$

Démonstration. Omise. \square

En notant de plus que, pour tout $\gamma > 0$,

$$\begin{aligned} \sup_{\ell \in \mathbb{S}_\gamma} \left\{ - \int_{-T/2}^{T/2} \ell(u) dK(u) \right\} &= \sup_{\ell \in \mathbb{S}_\gamma} \left\{ - \int_{-T/2}^{T/2} \dot{\ell}(u) K(u) du \right\} \\ &= \left\{ \gamma \int_{-T/2}^{T/2} K^2(u) du \right\}^{1/2}, \end{aligned}$$

la conclusion du théorème 2.1.6 s'obtient par des arguments analytiques de routine.

2.2.5 Preuve du théorème 2.1.7

Posons

$$\Psi_{n,1} = \Psi_n \{(1 - F_n^*)^2 / (1 - F)^2\}.$$

On constate alors que

$$\begin{aligned} (f_{n,h}^* - \widehat{\mathbb{E}}f_{n,h}^*(x)) \left\{ \Psi_n(x) \frac{1 - G(x)}{f(x)} \right\}^{1/2} = \\ \left(\frac{f_{n,h}^*}{1 - F_n^*} - \frac{\widehat{\mathbb{E}}f_{n,h}^*(x)}{1 - F_n^*} \right) \left\{ \Psi_{n,1}(x) \frac{1 - H(x)}{\lambda(x)} \right\}^{1/2}. \end{aligned} \quad (2.78)$$

Compte tenu de cette relation, et en vertu du théorème 2.1.6, le théorème 2.1.7 s'obtient comme conséquence directe de la loi du logarithme itéré pour le processus a_n^* (voir Földes et Rejtő (1981)), jointe au fait – impliqué par (H.1-2) – que, pour tout $h \in [h'_n, h''_n]$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1/2} (\log_2 n)^{1/2} \left\{ \frac{nh}{\log(1/h)} \right\} \\ = \lim_{n \rightarrow \infty} \left\{ \frac{h \log_2 n}{\log(1/h)} \right\}^{1/2} = 0. \end{aligned} \quad (2.79)$$

Chapitre 3

Lois du logarithme uniformes pour les estimateurs non paramétriques à noyau.

3.1 Introduction-Résultats

Dans ce chapitre nous établissons un ensemble de résultats de convergence uniforme pour des estimateurs à noyau. Pour obtenir les théorèmes que nous allons énoncer, nous allons, à nouveau, faire usage de propriétés, convenablement sélectionnées, du processus empirique. Dans le cas présent, c'est le processus empirique multivarié, indexé par des classes de fonctions, qui va nous servir de base de travail. Les innovations que nous apporterons par la suite constituent des généralisations non triviales de résultats dus à Einmahl et Mason (2000) et Deheuvels et Mason (2004). Plus particulièrement, nous montrerons dans ce qui suit que les théorèmes de convergence présentés dans ces derniers articles demeurent valables, *uniformément en* $h \in [h'_n, h''_n]$, où $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ sont deux suites de constantes vérifiant les conditions [CRS]. Nos travaux complètent également ceux de Einmahl et Mason (2005), au sens que les lois du logarithme présentées par ces derniers auteurs, qui sont établies pour des fenêtres h variant dans des intervalles tels que $[\frac{c \log n}{n}; 1]$, ne fournissent pas de constante limite. Ceci restreint sérieusement leur intérêt pratique, notamment pour la construction de *bandes de confiance* simultanées presque sûres (voir le §4.4 dans le **chapitre 4** qui suit).

3.1.1 Notations et hypothèses

Soit $(\mathbf{X}_1, \mathbf{Z}_1), (\mathbf{X}_2, \mathbf{Z}_2), \dots$, une suite i.i.d. de répliques aléatoires du *couple aléatoire générique* $(\mathbf{X}, \mathbf{Z}) \in \mathbb{R}^d \times \mathbb{R}^q$, $d, q \geq 1$. Considérons une classe \mathcal{F} de fonctions mesurables, définies sur \mathbb{R}^q et à valeur réelle. Cette classe sera supposée *séparable point par point* (en anglais *pointwise measurable*¹; voir l'appendice de ce chapitre, ainsi que la p.110 de van der Vaart et Wellner (1996)) et de type *VC subgraph class* (voir l'annexe du présent mémoire, ainsi que le §2.6.2 de van der Vaart et Wellner (1996)). Etant donnée une fonction $\psi \in \mathcal{F}$, nous nous intéressons à la densité $f_{\mathbf{X}}(\mathbf{x})$ de \mathbf{X} en \mathbf{x} , ainsi qu'à l'espérance conditionnelle de $\psi(\mathbf{Z})$ sachant que $\mathbf{X} = \mathbf{x}$,

$$m_\psi(\mathbf{x}) = \mathbb{E}(\psi(\mathbf{Z}) \mid \mathbf{X} = \mathbf{x}), \quad (3.1)$$

¹Ce type d'hypothèse nous permet d'éviter les problèmes de mesurabilité, sans avoir recours aux « mesure externes » (voir van der Vaart et Wellner (1996) par exemple)

si tant est que cette fonction existe. Le problème de l'estimation de m_ψ et $f_{\mathbf{X}}$ a été abondamment traité dans la littérature des vingt dernières années. Pour plus de détails, nous renvoyons le lecteur à Collomb (1981), Prakasa Rao (1983), Devroye et Györfi (1985), Silverman (1986), Bosq et Lecoutre (1987), Nadaraya (1989), Roussas (1990), Härdle (1990), Scott (1992), Wand et Jones (1995), Eggermont et La Riccia (2001), Györfi *et al.* (2002), ainsi qu'aux références citées par ces auteurs.

En vue de l'étude des estimateurs à noyau de la régression et de la densité (définis en (3.11) et (3.13) ci-dessous), nous introduisons le processus suivant. Pour toute fonction $\psi \in \mathcal{F}$, et tout couple de fonctions continues, c_ψ et d_ψ , définies sur un compact I de \mathbb{R}^d , nous posons, pour tout $\mathbf{x} \in I$ et tout $h > 0$,

$$W_{n,h}(\mathbf{x}, \psi) = \sum_{j=1}^n \left(c_\psi(\mathbf{x})\psi(\mathbf{Z}_j) + d_\psi(\mathbf{x}) \right) K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h}\right) - n\mathbb{E}\left\{ \left(c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}) \right) K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right) \right\}. \quad (3.2)$$

Introduisons le processus empirique Υ_n basé sur les observations $(\mathbf{X}_1, \mathbf{Z}_1), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$, et indexé par un ensemble de fonctions \mathcal{G} . Pour $g \in \mathcal{G}$, $\Upsilon_n(g)$ est défini par

$$\Upsilon_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(\mathbf{X}_i, \mathbf{Z}_i) - \mathbb{E}g(\mathbf{X}_i, \mathbf{Z}_i)). \quad (3.3)$$

Posons, pour toute fonction mesurable ψ , pour tout $\mathbf{x} \in I$, et pour tout $h > 0$,

$$\eta_{\mathbf{x},h}(\mathbf{u}, \mathbf{v}) = \left(c_\psi(\mathbf{x})\psi(\mathbf{v}) + d_\psi(\mathbf{x}) \right) K\left(\frac{\mathbf{x} - \mathbf{u}}{h}\right), \quad \text{pour } \mathbf{u}, \mathbf{v} \in \mathbb{R}^d \times \mathbb{R}^q. \quad (3.4)$$

On a alors la relation

$$W_{n,h}(\mathbf{x}, \psi) = n^{1/2}\Upsilon_n(\eta_{\mathbf{x},h}). \quad (3.5)$$

Ainsi, l'étude de $W_{n,h}$, et, par conséquent, celle des estimateurs à noyau introduits dans la suite de ce chapitre, peut se ramener à l'étude du processus empirique Υ_n . Nous verrons l'illustration de ce principe au travers des démonstrations présentées tout au long de ce chapitre.

Avant de développer le détail de nos résultats, commençons par présenter nos hypothèses de travail, et notamment celles portant sur la distribution de (\mathbf{X}, \mathbf{Z}) . Soit I une boule de \mathbb{R}^d d'intérieur non vide. On supposera qu'il existe un $\alpha > 0$ tel que (\mathbf{X}, \mathbf{Z}) ait une densité jointe $f_{\mathbf{X},\mathbf{Z}}$ par rapport à la mesure de Lebesgue sur $I^\alpha \times \mathbb{R}^q$, avec

$$I^\alpha := \{\mathbf{x} \in \mathbb{R}^d : \inf_{\mathbf{y} \in I} \|\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^d} \leq \alpha\}.$$

Dans cette formule, $\|\cdot\|_{\mathbb{R}^d}$ désigne la norme euclidienne usuelle sur \mathbb{R}^d . En notant $f_{\mathbf{X}}$ la densité marginale de \mathbf{X} sur I^α , on supposera vérifiées certaines conditions de régularité, parmi les suivantes.

(F.I) Pour tout $\mathbf{x} \in I^\alpha$, $\lim_{\mathbf{x}' \rightarrow \mathbf{x}; \mathbf{x}' \in I^\alpha} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}', \mathbf{z}) = f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ pour presque tout $\mathbf{z} \in \mathbb{R}^q$;

(F.II) $f_{\mathbf{X}}$ est continue et strictement positive sur I^α ;

(F.III) \mathcal{F} est bornée (au sens que la classe \mathcal{F} possède une fonction enveloppe Ψ uniformément bornée, et mesurable, telle que $\Psi(\mathbf{z}) \geq \sup_{\psi \in \mathcal{F}} \psi(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^q$).

Sous les hypothèses (*F.I-II-III*), la fonction de régression, $m_\psi(\mathbf{x})$, est définie, pour $\mathbf{x} \in I^\alpha$, par

$$m_\psi(\mathbf{x}) = \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^q} \psi(\mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) d\mathbf{z} = \frac{r_\psi(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}, \quad (3.6)$$

avec

$$r_\psi(\mathbf{x}) = \int_{\mathbb{R}^q} \psi(\mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) d\mathbf{z}. \quad (3.7)$$

On introduit, par ailleurs, la variance conditionnelle $\sigma_\psi^2(\mathbf{x})$ de $\psi(\mathbf{Z})$, sachant que $\mathbf{X} = \mathbf{x}$. Sous (*F.I-II-III*), cette fonction de $\mathbf{x} \in I^\alpha$ est définie par

$$\sigma_\psi^2(\mathbf{x}) = \text{Var}(\psi(\mathbf{Z}) \mid \mathbf{X} = \mathbf{x}) = \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^q} \left(\psi(\mathbf{z}) - m_\psi(\mathbf{x}) \right)^2 f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) d\mathbf{z}. \quad (3.8)$$

Notons que, sous les hypothèses (*F.I-II-III*), les fonctions m_ψ , r_ψ et σ_ψ , introduites ci-dessus, sont continues sur I^α (voir, par exemple, le §A.3 de Deheuvels et Mason (2004) pour la démonstration de ce résultat).

Introduisons, par ailleurs, la famille de fonctions

$$\mathcal{M} := \{m_\psi/f_{\mathbf{X}} : \psi \in \mathcal{F}\}. \quad (3.9)$$

Dans les théorèmes qui suivent, concernant les estimateurs de la fonction de régression, nous ferons usage de l'hypothèse suivante sur \mathcal{M} , définie par (3.9).

(*F.IV*) L'ensemble de fonctions \mathcal{M} est relativement compact, pour la topologie engendrée par la norme uniforme sur I^α .

Remarquons que, pour $q = 1$, l'hypothèse (*F.IV*) est notamment vérifiée lorsque $\mathcal{F} = \{\mathbb{1}_{(-\infty, t]}, t \in \mathbb{R}\}$ (voir pp. 6-7 dans Einmahl et Mason (2000)). Cette propriété nous permettra de décrire, par la suite, sans difficulté, comme conséquence de résultats généraux, la convergence uniforme en $t \in \mathbb{R}$ de l'estimateur de la fonction de répartition conditionnelle $F(t|\cdot) := \mathbb{P}(Z \leq t|\cdot)$, lorsque $\mathbf{Z} = Z$ est une variable aléatoire réelle (voir le corollaire 4.3.1 au **chapitre 4** pour un exemple dans le cas *censuré*).

En préalable à l'énoncé des définitions de l'estimateur à noyau de $m_\psi(\mathbf{x})$, dit de Nadaraya-Watson (Nadaraya (1964), Watson (1964)), et de l'estimateur à noyau de $f_{\mathbf{X}}(\mathbf{x})$, dit d'Akaike-Parzen-Rosenblatt (Akaike (1954), Rosenblatt (1956), Parzen (1962)), nous allons préciser les notations et hypothèses qui seront imposées sur le noyau K . Soit

$$\mathcal{K} := \{K(\lambda(\cdot - \mathbf{z})), \mathbf{z} \in \mathbb{R}^d, \lambda > 0\}.$$

Introduisons les hypothèses (*K.I-II-III*) ci-dessous.

$$(K.I) \quad (i) \quad \lim_{|\mathbf{u}| \rightarrow 0} \int_{\mathbb{R}^d} (K(\mathbf{x}) - K(\mathbf{x} + \mathbf{u}))^2 d\mathbf{x} = 0.$$

$$(ii) \quad \lim_{\lambda \rightarrow 1} \int_{\mathbb{R}^d} (K(\lambda \mathbf{x}) - K(\mathbf{x}))^2 d\mathbf{x} = 0.$$

$$(K.II) \quad (i) \quad \text{Il existe un } 0 < \kappa < \infty \text{ tel que } K(\mathbf{s}) = 0, \text{ pour } |\mathbf{s}| \geq \frac{\kappa}{2}.$$

$$(ii) \quad \text{Il existe une constante } C_K \text{ telle que } \sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| \leq C_K.$$

$$(K.III) \quad (i) \quad \exists C > 0, v > 0, \forall 0 < \varepsilon < 1, \mathcal{N}(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-v}.$$

$$(ii) \quad \mathcal{K} \text{ est séparable point par point.}$$

Ici, et par la suite, nous poserons $|\mathbf{s}| = \max_{1 \leq j \leq d} |s_j|$ pour tout $\mathbf{s} \in \mathbb{R}^d$. De plus,

$$\mathcal{N}(\epsilon, \mathcal{K}) := \sup \left\{ \mathcal{N}(\epsilon, \mathcal{K}, L_2(\mathbb{P})), \mathbb{P} \text{ mesure de probabilité} \right\}, \quad (3.10)$$

représente le *nombre de recouvrement uniforme* de \mathcal{K} , relatif à $\epsilon > 0$, et à la classe de normes $\{L_2(\mathbb{P})\}$. Dans (3.10), \mathbb{P} varie dans l'ensemble des mesures de probabilité sur \mathbb{R}^d . Nous renvoyons à la définition A.1 de l'annexe du présent mémoire, pour une définition précise du *nombre de recouvrement* d'un espace métrique. Les ensembles de fonctions vérifiant la condition (K.III)(i) sont dits former des *classes à nombre de recouvrement polynomial*. La caractérisation de ces ensembles de fonctions n'est pas chose aisée. Par contre, les familles de fonctions de type *Vapnik-Červonenkis subgraph class* (en abrégé, *classes de fonctions VC*), qui sont un cas particulier des *classes à nombre de recouvrement polynomial*, sont relativement simples à caractériser. Nous renvoyons une nouvelle fois à l'annexe de ce mémoire pour plus de détails sur le sujet.

Les hypothèses (K.I-III) impliquent que la quantité $\int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t}$ soit finie, mais permettent à celle-ci de prendre la valeur 0. Pour éviter les difficultés technique inhérentes à ce dernier cas, nous supposons, par la suite, que K satisfait la condition additionnelle

$$(K.IV) \quad \int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1.$$

Remarque 3.1.1. *Soit un noyau K de la forme $K(\mathbf{x}) = \phi(P(\mathbf{x}))$, où P est un polynôme de d variables réelles, et où ϕ désigne une fonction réelle à variation bornée. On peut alors vérifier que les hypothèses (K.I) et (K.III) sont satisfaites pour de tels noyaux (ceci justifie, a posteriori, l'énoncé du théorème 0.3.1 du chapitre introductif).*

Avec ces notations, et pour un choix de fenêtre $h > 0$ approprié, on considère les estimateurs à noyau de $f_{\mathbf{X}}(\mathbf{x})$, $r_{\psi}(\mathbf{x})$, $m_{\psi}(\mathbf{x})$ et $\sigma_{\psi}^2(\mathbf{x})$ définis, respectivement, par :

$$f_{\mathbf{X};n}(\mathbf{x}; h) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right); \quad (3.11)$$

$$r_{\psi;n}(\mathbf{x}; h) = \frac{1}{nh^d} \sum_{i=1}^n \psi(\mathbf{Z}_i) K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right); \quad (3.12)$$

$$m_{\psi;n}(\mathbf{x}; h) = \begin{cases} \frac{r_{\psi;n}(\mathbf{x}; h)}{f_{\mathbf{X};n}(\mathbf{x}; h)} = \frac{\sum_{i=1}^n \psi(\mathbf{Z}_i) K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}, \text{ pour } f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n \psi(\mathbf{Z}_i), \text{ pour } f_{\mathbf{X};n}(\mathbf{x}; h) = 0; \end{cases} \quad (3.13)$$

$$\sigma_{\psi;n}^2(\mathbf{x}; h) = \begin{cases} \frac{\sum_{i=1}^n \left(\psi(\mathbf{Z}_i) - m_{\psi;n}(\mathbf{x}; h)\right)^2 K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}, \text{ pour } f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n \left\{ \psi(\mathbf{Z}_i) - \frac{1}{n} \sum_{j=1}^n \psi(\mathbf{Z}_j) \right\}^2, \text{ pour } f_{\mathbf{X};n}(\mathbf{x}; h) = 0. \end{cases} \quad (3.14)$$

Nous étudierons la convergence vers 0 de ces estimateurs, convenablement centrés par des *facteurs de centrage* convenables. Ces derniers, définis ci-dessous, ne coïncident pas nécessairement avec l'espérance des estimateurs. Nous concentrerons donc notre intérêt sur les *termes aléatoires*, correspondant aux estimateurs, centrés par leurs facteurs de centrage respectifs. Mentionnons, au passage, que l'ordre de grandeur asymptotique des *termes de biais*, définis par les différences entre les facteurs de centrage et les fonctions à estimer, peut être négligé par rapport à l'ordre de grandeur des *termes aléatoires*, moyennant certaines hypothèses additionnelles (cf. l'appendice du **chapitre 4**). Nous définissons, respectivement, les facteurs de centrage des estimateurs en (3.11), (3.12) et (3.13), par

$$\tilde{f}_{\mathbf{X}}(\mathbf{x}, h) := \mathbb{E} f_{\mathbf{X};n}(\mathbf{x}; h) = \mathbb{E} \left\{ \frac{1}{h^d} K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right\}; \quad (3.15)$$

$$\tilde{r}_{\psi}(\mathbf{x}, h) := \mathbb{E} r_{\psi;n}(\mathbf{x}; h) = \mathbb{E} \left\{ \frac{1}{h^d} \psi(\mathbf{Z}) K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right\}; \quad (3.16)$$

$$\widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h) = \begin{cases} \frac{\tilde{r}_{\psi}(\mathbf{x}, h)}{\tilde{f}_{\mathbf{X}}(\mathbf{x}, h)} = \frac{\mathbb{E} \left\{ \psi(\mathbf{Z}) K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right\}}{\mathbb{E} \left\{ K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right\}}, & \text{pour } \tilde{f}_{\mathbf{X}}(\mathbf{x}, h) \neq 0, \\ \mathbb{E}(\psi(\mathbf{Z})), & \text{pour } \tilde{f}_{\mathbf{X}}(\mathbf{x}, h) = 0. \end{cases} \quad (3.17)$$

Remarque 3.1.2. *Remarquons que l'égalité $\widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h) = \mathbb{E} m_{\psi;n}(\mathbf{x}; h)$ n'a, en général, pas lieu. Cependant, sous certaines hypothèses de régularité, la différence entre ces deux termes déterministes peut devenir négligeable lorsque $h \rightarrow 0$ et $nh^d \rightarrow \infty$ (voir, par exemple, (1.16) dans Deheuvels et Mason (2004)).*

Dans le présent chapitre, nous supposons que la *fenêtre* h est telle que $h \in [h'_n, h''_n]$. Ici, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ désignent deux suites de constantes réelles positives vérifiant les conditions (H.I-III-III) ci-dessous. Celles-ci sont exprimées en fonction d'une suite générique, $\{h_n\}_{n \geq 1}$, de réels positifs.

- (H.I) $h_n \downarrow 0$, $0 < h_n < 1$, et $nh_n^d \uparrow \infty$;
- (H.II) $nh_n^d / \log n \rightarrow \infty$ lorsque $n \rightarrow \infty$;
- (H.III) $\log(1/h_n) / \log \log n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

Dans la suite, nous ferons usage d'une fonction auxiliaire $\{\Theta(\mathbf{x}) : \mathbf{x} \in I\}$, supposée continue et strictement positive sur I . Nous supposons également disposer d'un estimateur de cette fonction, noté $\{\Theta_n(\mathbf{x}) : \mathbf{x} \in I\}$, pour $n = 1, 2, \dots$. On suppose que cet estimateur converge vers $\Theta(\cdot)$ sur I , au sens que, presque sûrement,

$$(\Theta.1) \quad \lim_{n \rightarrow \infty} \sup_{\mathbf{x} \in I} \left| \frac{\Theta_n(\mathbf{x})}{\Theta(\mathbf{x})} - 1 \right| = 0.$$

3.1.2 Résultats

Nous sommes maintenant en mesure d'énoncer les résultats originaux présents dans ce chapitre, qui complètent les travaux de Einmahl et Mason (2000), et de Deheuvels et Mason (2004). Il est à noter que notre théorème 3.1.2, ci-dessous, constitue une version améliorée du corollaire 2.1 de Varron (2006). La méthode dont nous faisons l'usage permet, ainsi, de compléter ces derniers résultats. Remarquons cependant que la généralisation que nous

apportons à ces travaux n'est que partielle, dans la mesure où Varron (2006) obtient, en sus des résultats que nous généralisons, des lois fonctionnelles de type Strassen (du même type que celles que nous avons présenté pour les incréments du processus des quantiles dans le **chapitre 1** et pour le processus empirique en données censurées dans le **chapitre 2**). Des lois fonctionnelles de ce type auraient pu être obtenues dans le cadre de notre étude, en combinant les deux méthodologies. Ceci, néanmoins, ne sera pas fait ici.

Théorème 3.1.1. *Soient $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.I-II-III-IV), (K.I-II-III-IV) et $(\Theta.1)$, on a, presque sûrement,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \frac{nh^d}{2 \log(1/h^d)} \right\}^{1/2} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \pm \Theta_n(\mathbf{x}) \left\{ m_{\psi;n}(\mathbf{x}; h) - \widehat{\mathbf{E}} m_{\psi;n}(\mathbf{x}; h) \right\} \\ &= \left\{ \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\Theta^2(\mathbf{x}) \sigma_\psi^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right\}^{1/2}. \end{aligned} \quad (3.18)$$

Nous énonçons ci-dessous une version de ce théorème, établie pour l'estimateur à noyau de la densité.

Théorème 3.1.2. *Soient $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.II), (K.I-II-III-IV) et $(\Theta.1)$, on a, presque sûrement,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \left\{ \frac{nh^d}{2 \log(1/h^d)} \right\}^{1/2} \sup_{\mathbf{x} \in I} \pm \Theta_n(\mathbf{x}) \left\{ f_{\mathbf{X};n}(\mathbf{x}; h) - \mathbf{E} f_{\mathbf{X};n}(\mathbf{x}; h) \right\} \\ &= \left\{ \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t} \sup_{\mathbf{x} \in I} \left(\Theta^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \right) \right\}^{1/2}. \end{aligned} \quad (3.19)$$

La démonstration des théorèmes 3.1.1 et 3.1.2 est immédiate au vu des résultats du §3.1.4 ci-dessous.

3.1.3 Un résultat utile.

Nous montrons, dans ce qui suit, que les théorèmes 3.1.1 et 3.1.2 sont des conséquences directes (grâce, notamment, au lemme 3.3.1, énoncé dans la suite du présent chapitre) d'un résultat général, énoncé dans le théorème 3.1.3 ci-dessous, et décrivant les oscillations d'une version du processus empirique multivarié, indexé par une famille de fonctions convenablement choisie. Ce dernier processus est défini en (3.2), et (3.20) ci-dessous.

Les notations suivantes seront utiles. Introduisons deux familles de fonctions, continues en $x \in I^\alpha$, et indexées par $\psi \in \mathcal{F}$, dénotées par

$$\mathcal{F}_C := \{c_\psi(x) : \psi \in \mathcal{F}\} \quad \text{et} \quad \mathcal{F}_D := \{d_\psi(x) : \psi \in \mathcal{F}\}.$$

Par la suite, les ensembles de fonctions \mathcal{F}_C et \mathcal{F}_D seront supposés relativement compacts pour la topologie de la convergence uniforme sur I^α . Le théorème d'Arzelà-Ascoli montre que cette hypothèse est équivalente au fait que ces ensembles de fonctions soient uniformément équicontinus et bornés sur I^α .

Soit $\psi \in \mathcal{F}$, et les fonctions $c_\psi \in \mathcal{F}_C$ et $d_\psi \in \mathcal{F}_D$ qui lui correspondent dans \mathcal{F}_C et \mathcal{F}_D . Posons, pour tout $\mathbf{x} \in I^\alpha$, et pour tout $h > 0$,

$$W_{n,h}(\mathbf{x}, \psi) = \sum_{j=1}^n \left(c_\psi(\mathbf{x})\psi(\mathbf{Z}_j) + d_\psi(\mathbf{x}) \right) K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h}\right) - n\mathbb{E}\left\{ \left(c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}) \right) K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right) \right\}. \quad (3.20)$$

Nous pouvons maintenant énoncer, dans le théorème 3.1.3 ci-dessous, un résultat technique, qui nous sera particulièrement utile pour démontrer les autres résultats de ce chapitre.

Théorème 3.1.3. *Soient $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.I-II-III) et (K.I-II-III-IV), on a, presque sûrement,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \frac{\sup_{\mathbf{x} \in I} \pm W_{n,h}(\mathbf{x}, \psi)}{\sqrt{2nh^d \log(1/h^d)}} = \sup_{\psi \in \mathcal{F}} \sigma(\psi), \quad (3.21)$$

où

$$\sigma^2(\psi) = \sup_{\mathbf{x} \in I} \mathbb{E}\left\{ \left(c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}) \right)^2 \mid \mathbf{X} = \mathbf{x} \right\} f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t}. \quad (3.22)$$

La démonstration du théorème 3.1.3 sera établie dans le §3.2 qui suit.

3.1.4 Corollaires

Nous présentons, dans ce paragraphe, trois corollaires du théorème 3.1.3. Ces derniers, mis ensemble, permettent d'établir les théorèmes 3.1.1 et 3.1.2, puisque l'hypothèse (Θ.1) assure que, presque sûrement lorsque $n \rightarrow \infty$,

$$\sup_{\mathbf{x} \in I} |\Theta_n(\mathbf{x}) - \Theta(\mathbf{x})| \rightarrow 0.$$

La démonstration du corollaire 3.1.3 est présentée dans la section 3.3. Les démonstrations des corollaires 3.1.1 et 3.1.2 sont directes en utilisant des arguments similaires, mais plus simples, et sont omises.

Corollaire 3.1.1. *Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.I-II-III) et (K.I-II-III-IV), on a, presque sûrement,*

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \{\hat{r}_{\psi;n}(\mathbf{x}, h) - \mathbf{E}r_{\psi;n}(\mathbf{x}, h)\}}{\sqrt{2 \log(1/h^d)}} = \tau_I, \quad (3.23)$$

où

$$\tau_I = \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \left\{ \tau_\psi^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2}, \quad (3.24)$$

et

$$\tau_\psi^2(\mathbf{x}) = \mathbb{E}(\psi^2(\mathbf{Z}) \mid \mathbf{X} = \mathbf{x}), \quad \text{pour } \mathbf{x} \in I \text{ et } \psi \in \mathcal{F}. \quad (3.25)$$

Corollaire 3.1.2. Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.II) et (K.I-II-III-IV), on a, presque sûrement,

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \{\widehat{f}_{\mathbf{X};n}(\mathbf{x}, h) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x}, h)\}}{\sqrt{2 \log(1/h^d)}} = \varrho_I, \quad (3.26)$$

où

$$\varrho_I = \sup_{\mathbf{x} \in I} \left\{ f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2}. \quad (3.27)$$

Corollaire 3.1.3. Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.I-II-III-IV) et (K.I-II-III-IV), on a, presque sûrement,

$$\lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \{\widehat{m}_{\psi;n}(\mathbf{x}, h) - \widehat{\mathbb{E}}m_{\psi;n}(\mathbf{x}, h)\}}{\sqrt{2 \log(1/h^d)}} = \sigma_I, \quad (3.28)$$

où

$$\sigma_I = \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \left\{ \frac{\sigma_{\psi}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2}, \quad (3.29)$$

et

$$\sigma_{\psi}^2(\mathbf{x}) = \text{Var}(\psi(\mathbf{Z}) \mid \mathbf{X} = \mathbf{x}), \quad \text{pour } \mathbf{x} \in I \text{ et } \psi \in \mathcal{F}. \quad (3.30)$$

3.2 Démonstration du théorème 3.1.3

Dans un premier temps, nous nous plaçons dans le cas simple où $\mathcal{F} = \{\psi\}$, pour une fonction fixée ψ , définie sur \mathbb{R}^q , à valeur réelle, mesurable et uniformément bornée. Nous emprunterons dans une large mesure les méthodes utilisées par Einmahl et Mason (2000), Einmahl et Mason (2005), Deheuvels et Mason (2004) et Varron (2006). Nous verrons au §3.2.2 comment étendre ces résultats au cas général.

3.2.1 Le cas $\mathcal{F} = \{\psi\}$

Pour établir notre résultat dans ce cas simplifié, où $\mathcal{F} = \{\psi\}$, ψ étant une fonction définie sur \mathbb{R}^q , mesurable, à valeur réelle et uniformément bornée, nous procéderons en deux étapes, reprenant la méthodologie utilisée dans les articles de Einmahl et Mason (2000) et Deheuvels et Mason (2004). Dans un premier temps, nous chercherons à obtenir une borne majorante de la limite supérieure de la suite évoquée dans le membre de gauche de (3.31) ci-dessous. Pour ce faire, nous utiliserons tout d'abord une technique de discrétisation, pour appliquer l'inégalité maximale de Bernstein (énoncée dans le fait 3.4.1 dans l'appendice de ce chapitre). Ensuite, nous évaluerons les oscillations entre les valeurs de la discrétisation. Pour cela, nous utiliserons une inégalité issue des travaux de Talagrand (1994) et Einmahl et Mason (2000), et « résumée » par Varron (2006) (voir le fait 3.4.3 cité dans l'appendice de ce chapitre; voir également Mason (2004)). Dans un second temps, nous montrerons que cette borne est optimale, au sens qu'elle correspond à la borne minorante de la limite inférieure, obtenue pour cette même suite.

Sans perte de généralité, nous supposerons que $\kappa = 1$ dans l'hypothèse (K.II)(i). Nous supposerons, de plus, que la fonction mesurable ψ est telle que $\sigma(\psi) > 0$ (où la définition de $\sigma(\psi)$ est rappelée au (3.32) ci-dessous). Compte tenu de ces notations et hypothèses, notre but est d'établir le résultat suivant.

Théorème 3.2.1. Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (F.I-II-III), (K.I-II-III-IV), on a, avec presque sûrement,

$$\lim \left\{ \sup_{h \in [h'_n, h''_n]} \left\{ 2nh^d \log(1/h^d) \right\}^{-1/2} \sup_{\mathbf{x} \in I} \pm W_{n,h}(\mathbf{x}, \psi) \right\} = \sigma(\psi), \quad (3.31)$$

où

$$\sigma^2(\psi) = \sup_{\mathbf{x} \in I} \mathbf{E} \left\{ (c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}))^2 \mid \mathbf{X} = \mathbf{x} \right\} f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t}. \quad (3.32)$$

Obtention de la borne supérieure

Dans cette première partie, en nous plaçant sous les hypothèses du théorème 3.2.1, nous allons montrer que, presque sûrement,

$$\forall \varepsilon > 0, \limsup_{n \rightarrow \infty} \left\{ \sup_{h \in [h'_n, h''_n]} \frac{\sup_{\mathbf{x} \in I} |W_{n,h}(\mathbf{x}, \psi)|}{\sqrt{2nh^d \log(1/h^d)}} \right\} \leq (1 + 2\varepsilon)\sigma(\psi). \quad (3.33)$$

Il sera utile de rappeler la définition du processus empirique Υ_n basé sur les observations $(\mathbf{X}_1, \mathbf{Z}_1), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$ et indexé par une classe de fonctions \mathcal{G} . Pour $g \in \mathcal{G}$, ce dernier est défini par

$$\Upsilon_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(g(\mathbf{X}_i, \mathbf{Z}_i) - \mathbf{E} g(\mathbf{X}_i, \mathbf{Z}_i) \right). \quad (3.34)$$

Pour toute fonction φ à valeur réelle, définie sur un ensemble B , nous utiliserons la notation $\|\varphi\|_B = \sup_{\mathbf{x} \in B} |\varphi(\mathbf{x})|$, et dans les cas particuliers où $B = \mathbf{R}^m, m \geq 1$, nous écrirons $\|\varphi\|_B = \|\varphi\|$. Notons, au passage, que le caractère borné de ψ suppose l'existence d'une constante $0 < M_\psi < \infty$ telle que $\|\psi\| \leq M_\psi$.

Pour toute famille \mathcal{G} de fonctions mesurables, notons

$$\|n^{1/2}\Upsilon_n\|_{\mathcal{G}} := \sup_{g \in \mathcal{G}} |n^{1/2}\Upsilon_n(g)|.$$

Pour tout $\mathbf{x} \in I^\alpha$, et tout $h \in [h'_n, h''_n]$, posons

$$\eta_{\mathbf{x},h,\psi}(\mathbf{u}, \mathbf{v}) = (c_\psi(\mathbf{x})\psi(\mathbf{v}) + d_\psi(\mathbf{x}))K\left(\frac{\mathbf{x} - \mathbf{u}}{h}\right), \quad \text{pour } \mathbf{u}, \mathbf{v} \in \mathbf{R}^d \times \mathbf{R}^q. \quad (3.35)$$

Nous constatons alors que

$$W_{n,h}(\mathbf{x}, \psi) = n^{1/2}\Upsilon_n(\eta_{\mathbf{x},h,\psi}). \quad (3.36)$$

Grâce à la relation (3.36), et au prix de quelques manipulations mathématiques, nous allons maintenant pouvoir déduire (3.33) de résultats connus de la théorie des processus empiriques.

Fixons $\varepsilon > 0$ dans (3.33). Introduisons des constantes $\gamma > 0$, $\delta \in (0, \alpha/4)$, et $\lambda \in (1, (1 + 2\gamma)^{1/d})$, qu'on précisera, par la suite, en fonction de $\varepsilon > 0$, pour les besoins de notre démonstration. Pour tout entier $k \geq 0$, posons

$$n_k := \lfloor (1 + \gamma)^k \rfloor, \quad (3.37)$$

où, comme ailleurs dans cette thèse, $\lfloor u \rfloor \leq u < \lfloor u \rfloor + 1$ désigne la partie entière (inférieure) de u .

Pour chaque entier $k \geq 1$, et n tel que $n_{k-1} \leq n \leq n_k$, considérons l'intervalle

$$[h'_{n_k}, h''_{n_{k-1}}] \supset [h'_n, h''_n].$$

On en établit la discrétisation suivante. Soit, tout d'abord,

$$R_k := \left\lfloor \frac{\log(h''_{n_{k-1}}/h'_{n_k})}{\log \lambda} \right\rfloor + 1. \quad (3.38)$$

On pose, ensuite,

$$h'_{n_k, R_k} := h''_{n_{k-1}} \quad \text{et} \quad h'_{n_k, \ell} := \lambda^\ell h'_{n_k} \quad \text{pour} \quad \ell = 0, \dots, R_k - 1. \quad (3.39)$$

Pour chaque entier ℓ vérifiant $0 \leq \ell \leq R_k$, nous incluons maintenant I dans une union de J_ℓ hyper-cubes *disjoints*, d'arête $d^{-1} \delta h'_{n_k, \ell}$. Ces hyper-cubes sont notés

$$\Gamma_{k, \ell, j} := \left\{ x_{k, \ell, j} +]0, d^{-1/2} \delta h'_{n_k, \ell}]^d \right\} \quad \text{pour} \quad 1 \leq j \leq J_\ell, \quad (3.40)$$

et vérifient les relations suivantes, pour tout entier $k \geq 1$ et tout entier $\ell = 0, \dots, R_k$ (il est utile de rappeler ici que, par hypothèse, $0 < \delta < \alpha/4$ et $h'_n < 1$ pour $n \geq 1$).

$$I \subset \bigcup_{j=1}^{J_\ell} \Gamma_{k, \ell, j} \subset I^{\alpha/2}. \quad (3.41)$$

Comme la construction (3.40) impose que les $\Gamma_{k, \ell, j}$, $1 \leq j \leq J_\ell$, sont disjoints, on constate qu'il existe une constante $C := C(\delta)$, dépendant uniquement de $\delta > 0$ et de I , telle que

$$J_\ell \leq \frac{C}{h'_{n_k, \ell}{}^d}, \quad \text{pour} \quad k \geq 1 \quad \text{et} \quad 0 \leq \ell \leq R_k. \quad (3.42)$$

Dans la suite, il sera commode, pour tout entier $k \geq 1$, de désigner par N_k l'ensemble d'indices défini par

$$N_k := \{n_{k-1} + 1, \dots, n_k\}, \quad (3.43)$$

lorsque $n_{k-1} < n_k$, et par $N_k := \emptyset$ pour les valeurs de k telles que $n_{k-1} = n_k$. On observera que, quel que soit le choix initial de $\gamma > 0$, l'ensemble N_k est non vide pour tout entier k suffisamment grand.

Notre démonstration commence par la décomposition suivante, avec $k \geq 1$ fixé suffisam-

ment grand pour que $N_k \neq \emptyset$.

$$\begin{aligned}
\mathbb{P}_k &:= \mathbb{P} \left[\max_{n \in N_k} \sup_{h \in [h'_n, h''_n]} \frac{\sup_{\mathbf{x} \in I} |W_{n,h}(\mathbf{x}, \psi)|}{\sqrt{2nh^d \log(1/h^d)}} \geq (1 + 2\varepsilon)\sigma(\psi) \right] \\
&= \mathbb{P} \left[\max_{n \in N_k} \sup_{h \in [h'_n, h''_n]} \frac{\sup_{\mathbf{x} \in I} |\Upsilon_n(\eta_{\mathbf{x},h})|}{\sqrt{2h^d \log(1/h^d)}} \geq (1 + 2\varepsilon)\sigma(\psi) \right] \\
&\leq \mathbb{P} \left[\max_{n \in N_k} \sup_{h \in [h'_{n_k}, h''_{n_k-1}]} \frac{\sup_{\mathbf{x} \in I} |\Upsilon_n(\eta_{\mathbf{x},h})|}{\sqrt{2h^d \log(1/h^d)}} \geq (1 + 2\varepsilon)\sigma(\psi) \right] \\
&\leq \mathbb{P} \left[\max_{n \in N_k} \max_{0 \leq l \leq R_k} \max_{1 \leq j \leq J_l} \frac{|W_{n,h'_{n_k,l}}(\mathbf{x}_{k,l,j}, \psi)|}{\sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)}} \geq (1 + \varepsilon)\sigma(\psi) \right] \\
&\quad + \mathbb{P} \left[\max_{n \in N_k} \max_{\substack{0 \leq l \leq R_k - 1 \\ 1 \leq j \leq J_l}} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h'_{n_k,l}, h'_{n_k,l+1}]}} \left| \frac{\Upsilon_n(\eta_{\mathbf{x},h})}{\sqrt{2h^d \log(1/h^d)}} \right. \right. \\
&\quad \left. \left. - \frac{W_{n,h'_{n_k,l}}(\mathbf{x}_{k,l,j}, \psi)}{\sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)}} \right| \geq \varepsilon\sigma(\psi) \right] \\
&:= \mathbb{P}_{1,k} + \mathbb{P}_{2,k}.
\end{aligned} \tag{3.44}$$

Nous allons montrer que \mathbb{P}_k est le terme général d'une série sommable afin d'appliquer le lemme de Borel-Cantelli. Au vu de (3.44), nous procéderons en deux étapes. Dans un premier temps, nous montrerons que, pour tout $\gamma > 0$ suffisamment petit, tout $\lambda > 1$ et tout $0 < \delta < \alpha/4$, $\mathbb{P}_{1,k}$ est le terme général d'une série sommable (partie discrétisation). Ensuite, nous montrerons que, pour un certain choix de $\gamma > 0$, $\lambda > 1$ et $\delta > 0$ suffisamment petits, $\mathbb{P}_{2,k}$ est, là encore, le terme général d'une série sommable (partie évaluation des oscillations).

Evaluation de $\mathbb{P}_{1,k}$: discrétisation. Fixons $\gamma > 0$. Pour tout $k \geq 1$, $0 \leq l \leq R_k$, $1 \leq j \leq J_l$, et $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^q$, on définit

$$g_{k,l,j}(\mathbf{u}, \mathbf{v}) := \eta_{\mathbf{x}_{k,l,j}, h'_{n_k,l}, \psi}(\mathbf{u}, \mathbf{v}). \tag{3.45}$$

Introduisons de plus, pour tout $k \geq 1$, la classe de fonctions définies sur $\mathbb{R}^d \times \mathbb{R}^q$,

$$\mathcal{G}_k := \{g_{k,l,j} : 0 \leq l \leq R_k, 1 \leq j \leq J_l\}.$$

D'après (3.35) et (3.45), il vient, pour tout $0 \leq l \leq R_k$, $1 \leq j \leq J_l$ et $h \in [h'_{n_k}, h''_{n_k-1}]$,

$$\|g_{k,l,j}\| + \|\eta_{\mathbf{x},h,\psi}\| \leq 2\{\|c_\psi\| \|\psi\| + \|d_\psi\|\} \|K\| =: M_1, \tag{3.46}$$

où $\|\psi\| \leq M_\psi < \infty$ sous (F.III).

Proposition 3.2.1. *On suppose satisfaites les conditions du théorème 3.2.1. Si (3.40) est vérifiée avec $0 < \delta < \alpha/4$, alors on a, presque sûrement, pour tout $\varepsilon > 0$,*

$$\limsup_{k \rightarrow \infty} \left\{ \max_{\substack{0 \leq l \leq R_k \\ 1 \leq j \leq J_l}} \frac{\max_{n \in N_k} |n^{1/2} \Upsilon_n(g_{k,l,j})|}{\sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)}} \right\} \leq (1 + \varepsilon)\sigma(\psi). \tag{3.47}$$

Démonstration. D'après l'inégalité (3.42), on a la majoration suivante

$$\begin{aligned} \mathbb{P}_{1,k} &\leq \sum_{l=0}^{R_k} \frac{C(\delta)}{h_{n_k,l}^d} \max_{1 \leq j \leq J_l} \mathbb{P} \left[\frac{\max_{n \in N_k} |n^{1/2} \Upsilon_n(g_{k,l,j})|}{\sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)}} \geq (1 + \varepsilon) \sigma(\psi) \right] \\ &=: \sum_{l=0}^{R_k} \frac{C(\delta)}{h_{n_k,l}^d} \max_{1 \leq j \leq J_l} \mathbb{P}_{1,k,l,j}. \end{aligned} \quad (3.48)$$

Nous allons chercher à appliquer l'inégalité maximale de Bernstein pour évaluer le terme $\mathbb{P}_{1,k,l,j}$. Pour ce faire, nous disposons de la majoration (3.46). De plus, d'après (3.35) et (3.45)

$$\text{Var}(g_{k,l,j}(\mathbf{X}, \mathbf{Z})) \leq \mathbb{E} \left[g_{k,l,j}^2(\mathbf{X}, \mathbf{Z}) \right] = \mathbb{E} \left[(c_\psi(\mathbf{x}_{k,l,j}) \psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{k,l,j}))^2 K^2 \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h_{n_k,l}'} \right) \right].$$

Sous l'hypothèse $(K.II)(i)$, en utilisant un argument de conditionnement classique, il apparaît que le terme de droite vaut

$$\begin{aligned} &\int_{|\mathbf{x}_{k,l,j} - \mathbf{t}| \leq h_{n_k,l}'/2} \mathbb{E} \left[(c_\psi(\mathbf{x}_{k,l,j}) \psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{k,l,j}))^2 \mid \mathbf{X} = \mathbf{t} \right] f_{\mathbf{X}}(\mathbf{t}) K^2 \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{t}}{h_{n_k,l}'} \right) d\mathbf{t} \\ &\leq \int_{|\mathbf{x}_{k,l,j} - \mathbf{t}| \leq h_{n_k,l}'/2} \sigma^2(\psi) \frac{1}{\int_{\mathbb{R}^d} K^2(\mathbf{u}) d\mathbf{u}} [1 + o(1)] K^2 \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{t}}{h_{n_k,l}'} \right) d\mathbf{t} \\ &= \frac{\sigma^2(\psi) h_{n_k,l}^d}{\int_{\mathbb{R}^d} K^2(\mathbf{u}) d\mathbf{u}} \int_{|\mathbf{t}| \leq 1/2} K^2(\mathbf{t}) d\mathbf{t} + o(h_{n_k,l}^d) \\ &= \sigma^2(\psi) h_{n_k,l}^d + o(h_{n_k,l}^d), \end{aligned} \quad (3.49)$$

où l'on a utilisé les hypothèses $(F.II)$ et $(K.II)(i)$, la continuité de la fonction $\mathbf{t} \rightarrow \mathbb{E}[(c_\psi(\mathbf{x}_{k,l,j}) \psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{k,l,j}))^2 \mid \mathbf{X} = \mathbf{t}] f_{\mathbf{X}}(\mathbf{t})$ et le changement de variable classique $\mathbf{u} = (\mathbf{x}_{k,l,j} - \mathbf{t})/h_{n_k,l}'$. On en déduit que, à partir d'un certain rang $[apcr]$,

$$\max_{\substack{0 \leq l \leq R_k \\ 1 \leq j \leq J_l}} \max_{n \in N_k} \text{Var}(g_{k,l,j}(\mathbf{X}, \mathbf{Z})) \leq \sigma^2(\psi) (1 + \varepsilon) h_{n_k,l}^d. \quad (3.50)$$

On peut alors appliquer l'inégalité maximale de Bernstein (fait 3.4.1 en appendice de ce chapitre), avec la constante $M = M_1$, aux variables :

$$U_i = g_{k,l,j}(\mathbf{X}_i, \mathbf{Z}_i) - \mathbb{E} g_{k,l,j}(\mathbf{X}_i, \mathbf{Z}_i), \quad i = 1, \dots, n.$$

On obtient

$$\begin{aligned} \mathbb{P}_{1,k,l,j} &= \mathbb{P} \left\{ \max_{n \in N_k} |n^{1/2} \Upsilon_n(g_{k,l,j})| \geq \sigma(\psi) (1 + \varepsilon) \sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)} \right\} \\ &\leq 2 \exp \left(- \frac{2\sigma(\psi)^2 (1 + \varepsilon) h_{n_k,l}^d \log(1/h_{n_k,l}^d) n_k}{2n_k \sigma^2(\psi) h_{n_k,l}^d + \frac{2M_1 \sigma(\psi)}{3} \sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)}} \right), \end{aligned}$$

et comme, d'après l'hypothèse $(H.II)$, $n h_{n_k,l}^d \gg \log(1/h_{n_k,l}^d)$, on obtient, $apcr$,

$$\begin{aligned}\mathbb{P}_{1,k,l,j} &\leq 2 \exp \left[-\log(1/h_{n_k,l}^{\prime d})^{1+\varepsilon/2} \right] \\ &= 2h_{n_k,l}^{\prime d(1+\varepsilon/2)}.\end{aligned}$$

En réinjectant cette majoration dans (3.48), il vient

$$\begin{aligned}\mathbb{P}_{1,k} &\leq 2C(\delta) \sum_{l=0}^{R_k} h_{n_k,l}^{\prime d\varepsilon/2} \\ &= 2C(\delta) h_{n_k}^{\prime d\varepsilon/2} \sum_{l=0}^{R_k} \lambda^{ld\varepsilon/2} \\ &\leq \frac{2C(\delta) h_{n_k}^{\prime d\varepsilon/2}}{\lambda^{d\varepsilon/2} - 1} \lambda^{d(R_k+1)\varepsilon/2}.\end{aligned}\tag{3.51}$$

D'après la définition (3.38), on a $R_k \leq \frac{\log(h_{n_{k-1}}''/h_{n_k}')}{\log(\lambda)} + 1$, si bien que

$$\mathbb{P}_{1,k} \leq \frac{2C(\delta)\lambda^{d\varepsilon}}{\lambda^{d\varepsilon/2} - 1} h_{n_{k-1}}''^{d\varepsilon/2}.\tag{3.52}$$

Enfin, l'hypothèse (H.III) implique que $\forall \eta > 0$, $\sum_{k \geq 1} h_{n_{k-1}}''^\eta < \infty$, ce qui conduit au résultat (3.47) *via* le lemme de Borel-Cantelli. \square

Remarque 3.2.1. *C'est grâce à l'utilisation de l'inégalité de Bernstein qu'il nous est possible d'obtenir la constante associée aux différentes lois du logarithme établies dans ce chapitre. Les outils (et notamment le fait 3.4.3 de l'appendice) utilisés par la suite permettent d'obtenir ces mêmes lois, mais sans la constante. Ces résultats se démontrent alors beaucoup plus facilement : il n'est plus nécessaire de procéder à l'étape de discrétisation de l'intervalle I . C'est la méthodologie utilisée dans l'article de Einmahl et Mason (2005), auquel nous renvoyons pour un exemple d'utilisation.*

Evaluation de $\mathbb{P}_{2,k}$: évaluation des oscillations. Soit $\varepsilon_1 = \varepsilon/2$. On a

$$\begin{aligned}\mathbb{P}_{2,k} &= \mathbb{P} \left[\max_{n \in N_k} \max_{\substack{0 \leq l \leq R_k - 1 \\ 1 \leq j \leq J_l}} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h_{n_k,l}', h_{n_k,l+1}']}} \left| \frac{W_{n,h}(\mathbf{x}, \psi)}{\sqrt{2nh^d \log(1/h^d)}} \right. \right. \\ &\quad \left. \left. - \frac{W_{n,h_{n_k,l}'}(\mathbf{x}_{k,l,j}, \psi)}{\sqrt{2n_k h_{n_k,l}^{\prime d} \log(1/h_{n_k,l}^{\prime d})}} \right| \geq 2\varepsilon_1 \sigma(\psi) \right] \\ &\leq \mathbb{P} \left[\max_{n \in N_k} \max_{\substack{0 \leq l \leq R_k - 1 \\ 1 \leq j \leq J_l}} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h_{n_k,l}', h_{n_k,l+1}']}} \left| \frac{W_{n,h}(\mathbf{x}, \psi) - W_{n,h_{n_k,l}'}(\mathbf{x}_{k,l,j}, \psi)}{\sqrt{2n_k h_{n_k,l}^{\prime d} \log(1/h_{n_k,l}^{\prime d})}} \right| \geq \varepsilon_1 \sigma(\psi) \right] \\ &\quad + \mathbb{P} \left[\max_{n \in N_k} \max_{\substack{0 \leq l \leq R_k - 1 \\ 1 \leq j \leq J_l}} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h_{n_k,l}', h_{n_k,l+1}']}} B_{n,k,l,h} \left| \frac{W_{n,h}(\mathbf{x}, \psi)}{\sqrt{2n_k h_{n_k,l}^{\prime d} \log(1/h_{n_k,l}^{\prime d})}} \right| \geq \varepsilon_1 \sigma(\psi) \right] \\ &\quad \text{avec } B_{n,k,l,h} = \left| \sqrt{\frac{2n_k h_{n_k,l}^{\prime d} \log(1/h_{n_k,l}^{\prime d})}{2nh^d \log(1/h^d)}} - 1 \right| \\ &=: \mathbb{P}_{2,1,k} + \mathbb{P}_{2,2,k}.\end{aligned}\tag{3.53}$$

En vue de l'évaluation de $\mathbb{P}_{2,2,k}$, remarquons dans un premier temps que, pour tout $n \in N_k$ et tout $h \in [h'_{n_k,l}, h'_{n_k,l+1}]$,

$$\begin{aligned} B_{n,k,l,h} &\leq \left| \sqrt{\frac{\log(1/h'_{n_k,l})}{\log(1/h^d)}} - 1 \right| \sqrt{\frac{n_k h'_{n_k,l}}{n h^d}} + \left| \sqrt{\frac{n_k h'_{n_k,l}}{n h^d}} - 1 \right| \\ &\leq \sqrt{\frac{\log(1/h'_{n_k,l})}{\log(1/\lambda^d h'_{n_k,l})}} - 1 \left| \sqrt{\frac{\lambda^d n_k}{n}} + \left| \sqrt{\frac{\lambda^d n_k}{n}} - 1 \right| \right|. \end{aligned} \quad (3.54)$$

Or, $n_k/n_{k-1} \rightarrow 1 + \gamma$ lorsque $k \rightarrow \infty$, et on a donc, *apcr*, $1 \leq n_k/n_{k-1} \leq 1 + 2\gamma$. En rappelant que $\lambda \leq (1 + 2\gamma)^{1/d}$, on en déduit que $1 \leq \sqrt{\lambda n_k/n_{k-1}} \leq (1 + 2\gamma)$. Par ailleurs, $1 \leq \log(1/h'_{n_k,l})/\log(1/\lambda^d h'_{n_k,l}) = 1 + \log(\lambda^d)/\log(1/\lambda^d h'_{n_k,l}) \leq 1 + \log(\lambda^d)/\log(1/\lambda^d h'_{n_{k-1}})$. En d'autres termes, on a $\lim_{k \rightarrow \infty} \max_{0 \leq l \leq R_k} \left| \log(1/h'_{n_k,l})/\log(1/\lambda^d h'_{n_k,l}) - 1 \right| = 0$, ou encore, *apcr*, $\max_{0 \leq l \leq R_k} \left| \log(1/h'_{n_k,l})/\log(1/\lambda^d h'_{n_k,l}) - 1 \right| \leq \gamma$. En combinant ces différents résultats, il s'ensuit que, pour $n \in N_k$, $0 \leq l \leq R_k$, $1 \leq j \leq J_l$ et $h \in [h'_{n_k,l}, h'_{n_k,l+1}]$, on a $B_{n,k,l,h} \leq (1 + 2\gamma)\gamma + 2\gamma$. Ainsi, il existe un $\gamma_\varepsilon > 0$ tel que pour tout $0 < \gamma < \gamma_\varepsilon$,

$$\liminf_{k \rightarrow \infty} \min_{n \in N_k} \inf_{h \in [h'_n, h''_n]} \varepsilon_1 \sigma(\psi) B_{n,k,l,h}^{-1} \geq 2 \left(1 + \sqrt{\frac{2}{A_2}} \right) D_1(\nu) \sigma(\psi), \quad (3.55)$$

où A_2 et $D_1(\nu)$ sont les constantes évoquées dans le fait 3.4.3.

Revenant sur le terme $\mathbb{P}_{2,2,k}$, on a la majoration suivante,

$$\begin{aligned} \mathbb{P}_{2,2,k} &= \mathbb{P} \left[\max_{n \in N_k} \max_{\substack{0 \leq l \leq R_k - 1 \\ 1 \leq j \leq J_l}} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h'_{n_k,l}, h'_{n_k,l+1}]}} B_{n,k,l,h} \left| \frac{W_{n,h}(\mathbf{x}, \psi)}{\sqrt{2n_k h'_{n_k,l} \log(1/h'_{n_k,l})}} \right| \geq \varepsilon_1 \sigma(\psi) \right] \\ &\leq \sum_{l=0}^{R_k-1} \sum_{j=1}^{J_l} \mathbb{P}_{2,2,k,l,j}, \end{aligned} \quad (3.56)$$

où, pour $0 \leq l \leq R_k - 1$ et $1 \leq j \leq J_l$,

$$\mathbb{P}_{2,2,k,l,j} := \mathbb{P} \left[\max_{n \in N_k} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h'_{n_k,l}, h'_{n_k,l+1}]}} B_{n,k,l,h} \left| \frac{W_{n,h}(\mathbf{x}, \psi)}{\sqrt{2n_k h'_{n_k,l} \log(1/h'_{n_k,l})}} \right| \geq \varepsilon_1 \sigma(\psi) \right].$$

Introduisons alors la classe de fonctions

$$\bar{\mathcal{F}}_{k,l,j} := \left\{ (c_\psi(\mathbf{x}) + d_\psi(\mathbf{x})\psi) K \left(\frac{\cdot - \mathbf{x}}{h} \right), \mathbf{x} \in \Gamma_{k,l,j}, h \in [h'_{n_k,l}, h'_{n_k,l+1}] \right\}. \quad (3.57)$$

Pour tout $0 \leq l \leq R_k$, $1 \leq j \leq J_l$ fixés, en suivant la même méthodologie qu'en (3.49), on obtient

$$\begin{aligned} \text{Var} \left[(c_\psi(\mathbf{x}) + d_\psi(\mathbf{x}))\psi(\mathbf{Z}) K \left(\frac{\mathbf{X} - \mathbf{x}}{h} \right) \right] &\leq \lambda^d \sigma(\psi)^2 h'_{n_k,l} + o(h'_{n_k,l}) \\ &\leq 2\sigma(\psi)^2 h'_{n_k,l} + o(h'_{n_k,l}) \\ &\leq 4\sigma(\psi)^2 h'_{n_k,l} \text{ apcr}, \end{aligned} \quad (3.58)$$

et ce, uniformément en $h \in [h'_{n_k,l}, h'_{n_k,l+1}]$ et $\mathbf{x} \in \Gamma_{k,l,j}$. Par ailleurs, pour tout $g \in \bar{\mathcal{F}}_{k,l,j}$ $\|g\| \leq M_1$ (où M_1 est définie en (3.46)).

Nous allons chercher à appliquer le fait 3.4.3 de l'appendice. Par la suite, on dira qu'un ensemble \mathcal{S} , de fonctions définies sur un ensemble \mathcal{X} , vérifie la condition $[\mathcal{E}]$ si :

(i) il existe une fonction enveloppe G , mesurable, à valeur réelle, satisfaisant

$$G(x) \geq \sup_{g \in \mathcal{S}} |g(x)|, \quad \text{pour tout } x \in \mathcal{X};$$

(ii) pour des constantes $C_0 > 0$ et $\nu > 0$ convenables, la condition d'entropie

$$N(\varepsilon, \mathcal{S}) \leq C_0 \varepsilon^{-\nu}, \quad 0 < \varepsilon \leq 0,$$

est vérifiée.

D'après les conditions (K.II)(ii) et (K.III)(i), la classe $\mathcal{K} = \{K(\lambda(\cdot - \mathbf{z})), \mathbf{z} \in \mathbb{R}^d, \lambda > 0\}$ vérifie la condition $[\mathcal{E}]$. Considérons alors la classe de fonctions de $\mathbf{v} \in \mathbb{R}^q$

$$\mathcal{F}_1 := \{(a\psi(\mathbf{v}) + b) : |a| \leq C, |b| \leq C\},$$

où $C > 0$ borne uniformément les deux fonctions c_ψ et d_ψ sur I^α . D'après le lemme 2.6.15, p.146 de van der Vaart et Wellner (1996), \mathcal{F}_1 est une classe de fonctions VC , et, par suite (voir proposition A.3.1 en annexe du mémoire), satisfait la condition $[\mathcal{E}]$ (puisque la fonction ψ est supposée uniformément bornée). En utilisant le résultat du lemme A.4.1 (voir annexe du mémoire), il s'ensuit que la classe de fonctions de $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^q$

$$\{(a\psi(\mathbf{v}) + b)K(\lambda(\mathbf{u} + z)) : \lambda > 0, z \in \mathbb{R}^d, |a| \leq C, |b| \leq C\}$$

vérifie la condition $[\mathcal{E}]$. Il en est finalement de même pour la classe de fonctions de $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^q$ définie par

$$\begin{aligned} \mathcal{G}' &= \{(a\psi(\mathbf{v}) + b)K(\lambda(\mathbf{u} + z)) - (a'\psi(\mathbf{v}) + b')K(\lambda'(\mathbf{u} + z')) : \\ &\quad \lambda > 0, \lambda' > 0, z \in \mathbb{R}^d, z' \in \mathbb{R}^d, |a| \leq C, |a'| \leq C, |b| \leq C, |b'| \leq C\}. \end{aligned} \quad (3.59)$$

Par ailleurs, pour tout $0 \leq l \leq R_k$ et $1 \leq j \leq J_l$, $\overline{\mathcal{F}}_{k,l,j} \subseteq \mathcal{G}'$, et donc $\overline{\mathcal{F}}_{k,l,j}$ vérifie la condition $[\mathcal{E}]$. De plus, comme la classe \mathcal{K} est supposée séparable point par point, la continuité des fonctions c_ψ , d_ψ assure cette propriété aux classes \mathcal{G}' et $\overline{\mathcal{F}}_{k,l,j}$.

On peut donc appliquer le fait 3.4.3 avec $\tau = 2\sigma(\psi)$ (cf. (3.58)) et $\rho = \tau\sqrt{2/A_2}$, qui, combiné à (3.55), entraîne

$$\begin{aligned} \mathbb{P}_{2,2,k,l,j} &\leq \mathbb{P} \left[\max_{n \in N_k} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h'_{n_k,l}, h'_{n_k,l+1}]}} \left| \frac{W_{n,h}(\mathbf{x}, \psi)}{\sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)}} \right| \geq 2 \left(1 + \sqrt{\frac{2}{A_2}}\right) D_1(\nu) \sigma(\psi) \right] \\ &\leq \mathbb{P} \left[\max_{1 \leq n \leq n_k} \|n^{1/2} \Upsilon_n\|_{\overline{\mathcal{F}}_{k,l,j}} \geq D_1(\nu) (\tau + \rho) \sqrt{2n_k h_{n_k,l}^d \log(1/h_{n_k,l}^d)} \right] \\ &\leq 4 \exp \left[-A_2 \left(\frac{\rho}{\tau}\right)^2 \log(1/h_{n_k,l}^d) \right] \\ &= 4h_{n_k,l}^{2d}. \end{aligned} \quad (3.60)$$

En répétant les étapes menant à (3.51) et (3.52), il vient, au vu de (3.56) et (3.60),

$$\begin{aligned} \mathbb{P}_{2,2,k} &\leq 4C(\delta) \sum_{l=0}^{R_k-1} h_{n_k,l}^d \\ &\leq 4C(\delta) h_{n_k}^d \sum_{l=0}^{R_k-1} \lambda^{ld} \\ &\leq \frac{4C(\delta)}{\lambda^d - 1} h_{n_{k-1}}^d. \end{aligned} \quad (3.61)$$

D'après (H.III), on conclut que $\sum_{k \geq 1} \mathbb{P}_{2,2,k} < \infty$.

Pour l'évaluation de $\mathbb{P}_{2,1,k}$, on effectue la décomposition suivante, analogue à celle faite pour $\mathbb{P}_{2,2,k}$.

$$\begin{aligned} \mathbb{P}_{2,1,k} &= \mathbb{P} \left[\max_{n \in N_k} \max_{\substack{0 \leq l \leq R_k - 1 \\ 1 \leq j \leq J_l}} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h'_{n_k,l}, h'_{n_k,l+1}]}} \left| \frac{W_{n,h}(\mathbf{x}, \psi) - W_{n,h'_{n_k,l}}(\mathbf{x}_{k,l,j}, \psi)}{\sqrt{2n_k h'^d_{n_k,l} \log(1/h'^d_{n_k,l})}} \right| \geq \varepsilon_1 \sigma(\psi) \right] \\ &\leq \sum_{l=0}^{R_k-1} \sum_{j=1}^{J_l} \mathbb{P}_{2,1,k,l,j}, \end{aligned} \quad (3.62)$$

où, pour $0 \leq l \leq R_k - 1$ et $1 \leq j \leq J_l$,

$$\mathbb{P}_{2,1,k,l,j} := \mathbb{P} \left[\max_{n \in N_k} \sup_{\substack{\mathbf{x} \in \Gamma_{k,l,j} \\ h \in [h'_{n_k,l}, h'_{n_k,l+1}]}} \left| \frac{W_{n,h}(\mathbf{x}, \psi) - W_{n,h'_{n_k,l}}(\mathbf{x}_{k,l,j}, \psi)}{\sqrt{2n_k h'^d_{n_k,l} \log(1/h'^d_{n_k,l})}} \right| \geq \varepsilon_1 \sigma(\psi) \right].$$

Introduisons maintenant la classe de fonctions

$$\overline{\mathcal{F}}'_{k,l,j} := \left\{ g_{k,l,j} - \eta_{\mathbf{x},h,\psi}, \mathbf{x} \in \Gamma_{k,l,j}, h \in [h'_{n_k,l}, h'_{n_k,l+1}] \right\}, \quad (3.63)$$

avec $g_{k,l,j}$ défini en (3.45). Pour évaluer $\mathbb{P}_{2,1,k,l,j}$, l'idée va être d'appliquer une nouvelle fois le fait 3.4.3 de l'appendice. Pour ce faire, on peut noter que, pour tout $g \in \overline{\mathcal{F}}'_{k,l,j}$, $\|g\| \leq M_1$, où M_1 est défini en (3.46). Cherchons alors à obtenir une majoration de $\sup_{g \in \overline{\mathcal{F}}'_{k,l,j}} \text{Var}(g(\mathbf{X}, \mathbf{Z}))$. Premièrement, d'après le quadrillage de I sélectionné (voir (3.40)),

$$\sup_{\mathbf{x} \in \Gamma_{k,l,j}} \|\mathbf{x} - \mathbf{x}_{k,l,j}\| \leq \delta h'_{n_k,l}.$$

Pour toute fonction ϕ définie sur I^α et à valeur réelle, on définit le module de continuité de ϕ par

$$\omega_\phi(\delta) := \sup \left\{ |\phi(\mathbf{x}) - \phi(\mathbf{y})| : \|\mathbf{x} - \mathbf{y}\| \leq \delta, \text{ et } \mathbf{x}, \mathbf{y} \in I^\alpha \right\}, \quad \delta \geq 0. \quad (3.64)$$

Soit par ailleurs $\beta := \|\psi\|^2 + 1 < \infty$. En utilisant l'inégalité classique $(a+b)^2 \leq 2(a^2 + b^2)$, et d'après (3.35) et (3.45), on a, pour tout $k \geq 1$, tout $0 \leq l \leq R_k$ et tout $1 \leq j \leq J_l$,

$$\begin{aligned} &\mathbb{E} \left\{ \left(g_{k,l,j}(\mathbf{X}, \mathbf{Z}) - \eta_{n,\mathbf{x},h}(\mathbf{X}, \mathbf{Z}) \right)^2 \right\} = \mathbb{E} \left\{ \left(\eta_{n_k, \mathbf{x}_{k,l,j}, h'_{n_k,l}}(\mathbf{X}, \mathbf{Z}) - \eta_{n,\mathbf{x},h}(\mathbf{X}, \mathbf{Z}) \right)^2 \right\} \\ &\leq 2\mathbb{E} \left\{ \left((c_\psi(\mathbf{x}_{k,l,j}) - c_\psi(\mathbf{x}))\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{k,l,j}) - d_\psi(\mathbf{x}) \right)^2 K \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h'_{n_k,l}} \right)^2 \right\} \\ &\quad + 2\mathbb{E} \left\{ (c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}))^2 \left(K \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h'_{n_k,l}} \right) - K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right)^2 \right\} \\ &\leq 4 \left\{ \omega_{c_\psi}^2(\delta h'_{n_k,l}) \vee \omega_{d_\psi}^2(\delta h'_{n_k,l}) \right\} \beta \mathbb{E} \left\{ K \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h'_{n_k,l}} \right)^2 \right\} \\ &\quad + 4 \left\{ \|c_\psi\|_{I^\alpha}^2 \vee \|d_\psi\|_{I^\alpha}^2 \right\} \beta \mathbb{E} \left\{ \left(K \left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h'_{n_k,l}} \right) - K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right)^2 \right\} \\ &=: (I) + (II). \end{aligned} \quad (3.65)$$

Soit

$$\tilde{\Gamma}_{k,l,j} := \left\{ \mathbf{x}_{k,l,j} + [-h'_{n_k,l}/2, h'_{n_k,l}/2]^d \right\}. \quad (3.66)$$

Avec, comme précédemment, $\kappa = 1$ dans $(K.II)(i)$, il vient

$$K\left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h'_{n_k,l}}\right)^2 \leq \|K\|^2 \mathbb{1}\{\mathbf{X} \in \tilde{\Gamma}_{k,l,j}\}.$$

On a donc, *apcr*, pour tout $0 \leq l \leq R_k - 1$ et $1 \leq j \leq J_l$,

$$\begin{aligned} (I) &\leq 4 \left\{ \omega_{c_\psi}^2(\delta h'_{n_k,l}) \vee \omega_{d_\psi}^2(\delta h'_{n_k,l}) \right\} \beta \|K\|^2 \mathbb{P}(\mathbf{X} \in \tilde{\Gamma}_{k,l,j}) \\ &\leq 4 \left\{ \omega_{c_\psi}^2(\delta h'_{n_k,l}) \vee \omega_{d_\psi}^2(\delta h'_{n_k,l}) \right\} \beta \|K\|^2 \|f_{\mathbf{X}}\|_{I^\alpha} h'_{n_k,l}. \end{aligned} \quad (3.67)$$

Il nous suffit maintenant d'obtenir une majoration convenable de (II) , avec

$$(II) = 4 \left\{ \|c\|_{I^\alpha}^2 \vee \|d\|_{I^\alpha}^2 \right\} \beta \mathbb{E} \left\{ \left(K\left(\frac{\mathbf{x}_{k,l,j} - \mathbf{X}}{h'_{n_k,l}}\right) - K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right) \right)^2 \right\}.$$

Pour majorer ce terme, nous allons utiliser la condition $(K.I)$. Soient les fonctions B_1 et B_2 , respectivement définies sur \mathbb{R} et $[1, +\infty)$, et telles que

$$B_1(\delta) = \sup_{|\mathbf{u}| \leq \delta} \int_{\mathbb{R}^d} (K(\mathbf{x}) - K(\mathbf{x} + \mathbf{u}))^2 d\mathbf{x}, \quad (3.68)$$

$$B_2(\gamma) = \sup_{1/\gamma \leq \lambda \leq \gamma} \int_{\mathbb{R}^d} (K(\lambda \mathbf{x}) - K(\mathbf{x}))^2 d\mathbf{x}. \quad (3.69)$$

D'après la condition $(K.I)$,

$$\lim_{\delta \rightarrow 0} B_1(\delta) = 0 \quad \text{et} \quad \lim_{\lambda \rightarrow 1} B_2(\lambda) = 0. \quad (3.70)$$

Soit $\Delta := \int_{\mathbb{R}^d} (K((\mathbf{x} - \mathbf{t})/h) - K((\mathbf{x}_{k,l,j} - \mathbf{t})/h'_{n_k,l}))^2 d\mathbf{t}$. En posant $\mathbf{u} = (\mathbf{x}_{k,l,j} - \mathbf{t})/h'_{n_k,l}$, il vient

$$\begin{aligned} h'_{n_k,l} \Delta &= \int_{\mathbb{R}^d} \left[K\left(\frac{\mathbf{x} - \mathbf{x}_{k,l,j}}{h} + \mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K(\mathbf{u}) \right]^2 d\mathbf{u} \\ &= \int_{\mathbb{R}^d} \left[K\left(\frac{\mathbf{x} - \mathbf{x}_{k,l,j}}{h} + \mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K\left(\mathbf{u} \frac{h'_{n_k,l}}{h}\right) + K\left(\mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K(\mathbf{u}) \right]^2 d\mathbf{u} \\ &= \int_{\mathbb{R}^d} \left[K\left(\frac{\mathbf{x} - \mathbf{x}_{k,l,j}}{h} + \mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K\left(\mathbf{u} \frac{h'_{n_k,l}}{h}\right) \right]^2 d\mathbf{u} + \int_{\mathbb{R}^d} \left[K\left(\mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K(\mathbf{u}) \right]^2 d\mathbf{u} \\ &\quad + 2 \int_{\mathbb{R}^d} \left[K\left(\frac{\mathbf{x} - \mathbf{x}_{k,l,j}}{h} + \mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K\left(\mathbf{u} \frac{h'_{n_k,l}}{h}\right) \right] \left[K\left(\mathbf{u} \frac{h'_{n_k,l}}{h}\right) - K(\mathbf{u}) \right] d\mathbf{u} \\ &=: \Delta_1 + \Delta_2 + \Delta_3. \end{aligned}$$

Comme $h \in [h'_{n_k,l}, h'_{n_k,l+1}]$ et $\mathbf{x} \in \Gamma_{k,l,j}$, on a, d'après les définitions (3.39), (3.40), (3.68) et (3.69),

$$\begin{aligned} \Delta_1 &\leq 2B_1\left(\frac{2\delta}{\sqrt{d}}\right), \\ \Delta_2 &\leq B_2(\lambda), \\ \Delta_3 &\leq 2\sqrt{2B_1\left(\frac{2\delta}{\sqrt{d}}\right)B_2(\lambda)}, \end{aligned}$$

où nous avons utilisé l'inégalité de Cauchy-Schwarz pour majorer Δ_3 . On en conclut que $\Delta \leq \left(\sqrt{2B_1\left(\frac{2\delta}{\sqrt{d}}\right)} + \sqrt{B_2(\lambda)} \right)^2$, si bien que

$$(II) \leq 4 \left\{ \|c_\psi\|_{I^\alpha}^2 \vee \|d_\psi\|_{I^\alpha}^2 \right\} \beta \|f_{\mathbf{X}}\|_{I^\alpha} \left(\sqrt{2B_1\left(\frac{2\delta}{\sqrt{d}}\right)} + \sqrt{B_2(\lambda)} \right)^2 h_{n_k,l}^{td}.$$

D'après (3.65) et (3.67), on a

$$\begin{aligned} \text{Var}(g_{k,l,j}(\mathbf{X}, \mathbf{Z}) - \eta_{\mathbf{x},h}(\mathbf{X}, \mathbf{Z})) &\leq \mathbb{E} \left\{ \left(g_{k,l,j}(\mathbf{X}, \mathbf{Z}) - \eta_{\mathbf{x},h}(\mathbf{X}, \mathbf{Z}) \right)^2 \right\} \\ &\leq B \left\{ (\omega_c^2(\delta h'_{n_k,l}) \vee \omega_d^2(\delta h'_{n_k,l})) \right. \\ &\quad \left. + \left(\sqrt{2B_1\left(\frac{2\delta}{\sqrt{d}}\right)} + \sqrt{B_2(\lambda)} \right)^2 \right\} h_{n_k,l}^{td}, \end{aligned}$$

avec $B = 4 \|f_{\mathbf{X}}\|_{I^\alpha} \beta \left\{ \|K\|^2 + (\|c_\psi\|_{I^\alpha}^2 \vee \|d_\psi\|_{I^\alpha}^2) |K|_v^2 \right\}$. Ainsi, en introduisant, pour $k \geq 1$, $0 \leq l \leq R_k - 1$ et $1 \leq j \leq J_l$,

$$\sigma_{k,l,j}^2(\psi) = \sup \left\{ \text{Var}(g(\mathbf{X}, \mathbf{Z})) : g \in \overline{\mathcal{F}}'_{k,l,j} \right\},$$

et en sélectionnant un $\delta > 0$ suffisamment petit et un $\lambda > 1$ suffisamment proche de 1, il vient, d'après (3.70) et la continuité des fonctions c_ψ et d_ψ ,

$$\sigma_{k,l,j}^2(\psi) \leq \frac{\varepsilon_1^2 \sigma(\psi)^2}{[(1 + \sqrt{2/A_2}) D_1(\nu)]^2} h_{n_k,l}^{td}, \quad \text{apcr}, \quad (3.71)$$

où A_2 et $D_1(\nu)$ sont les constantes du fait 3.4.3. On rappelle que pour tout $g \in \overline{\mathcal{F}}'_{k,l,j}$ $\|g\| \leq M_1$. De plus, $\overline{\mathcal{F}}'_{k,l,j} \subset \mathcal{G}'$, où la classe de fonctions \mathcal{G}' est définie en (3.59) et vérifie la condition $[\mathcal{E}]$. De plus, $\overline{\mathcal{F}}'_{k,l,j}$ est séparable point par point (par la continuité des fonctions c_ψ , d_ψ , et le caractère séparable point par point de \mathcal{K}). On peut donc appliquer le fait 3.4.3 avec $\tau = \varepsilon_1 \sigma(\psi) / [(1 + \sqrt{2/A_2}) D_1(\nu)]$ et $\rho = \tau \sqrt{2/A_2}$. On obtient

$$\begin{aligned} \mathbb{P}_{2,1,k,l,j} &\leq \mathbb{P} \left[\max_{1 \leq n \leq n_k} \|n^{1/2} \Upsilon_n\|_{\overline{\mathcal{F}}'_{k,l,j}} \geq D_1(\nu)(\tau + \rho) \sqrt{2n_k h_{n_k,l}^{td} \log(1/h_{n_k,l}^{td})} \right] \\ &\leq 4 \exp \left[-A_2 \left(\frac{\rho}{\tau} \right)^2 \log(1/h_{n_k,l}^{td}) \right] \\ &= 4h_{n_k,l}^{2d}. \end{aligned} \quad (3.72)$$

En répétant les étapes menant à (3.51) et (3.52), il vient, au vu de (3.62) et (3.72),

$$\begin{aligned} \mathbb{P}_{2,1,k} &\leq 4C(\delta) \sum_{l=0}^{R_k-1} h_{n_k,l}^{td} \\ &\leq 4C(\delta) h_{n_k}^{td} \sum_{l=0}^{R_k-1} \lambda^{dl} \\ &\leq \frac{4C(\delta)}{\lambda^d - 1} h_{n_{k-1}}^{td}. \end{aligned} \quad (3.73)$$

D'après (H.III), on conclut que $\sum_{k \geq 1} \mathbb{P}_{2,1,k} < \infty$.

Bilan : En combinant les résultats issus de (3.61) et de (3.73), on conclut que $\sum_{k \geq 1} \mathbb{P}_{2,k} < \infty$. Le fait que $\sum_{k \geq 1} \mathbb{P}_k < \infty$ est alors une conséquence directe de (3.52), et le résultat (3.33) est ainsi établi.

Obtention de la borne inférieure

Dans cette partie, on ne travaille plus uniformément en $h \in [h'_n, h''_n]$. Les résultats sont obtenus pour $h = h'_n$. Il est évident que si le résultat de la proposition 3.2.2 ci-dessous est valable, alors il reste valable en prenant le supremum sur l'intervalle $[h'_n, h''_n]$.

On suppose toujours que ψ est une fonction définie sur \mathbb{R}^q , mesurable, à valeur réelles, uniformément bornée et telle que $\sigma(\psi) > 0$.

Proposition 3.2.2. *Sous les hypothèses du théorème 3.2.1, on a, presque sûrement,*

$$\liminf_{n \rightarrow \infty} \left\{ \sup_{\mathbf{x} \in I} \pm W_{n, h'_n}(\mathbf{x}, \psi) / \sqrt{2n h_n'^d \log(1/h_n'^d)} \right\} \geq \sigma(\psi). \quad (3.74)$$

Cette proposition est l'extension au cas multidimensionnel de la proposition 3 de Einmahl et Mason (2000). Nous suivrons donc la même méthodologie, et passerons assez vite sur les démonstrations des résultats techniques, renvoyant le lecteur à cet article pour plus de détails. La démonstration est présentée pour le cas '+' ; le cas '-' se traitant de façon similaire.

Résultats nécessaires Soit $\mathbf{U}, \mathbf{U}_1, \mathbf{U}_2, \dots$, des répliques i.i.d. de variables aléatoires à valeur dans \mathbb{R}^{d+q} . Pour $n \geq 1$, on considère la *fdr* empirique basée sur les n premiers vecteurs, définie par :

$$G_n(\mathbf{u}) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{U}_i \leq \mathbf{u}\}}, \quad \mathbf{u} \in \mathbb{R}^{d+q}.$$

On introduit le processus de Poisson, sur \mathbb{R}^{d+q} , $n\mathcal{N}_n$ défini par $n\mathbb{E}\mathcal{N}_n(A) = n\mathbb{P}(\mathbf{U} \in A)$ pour tout borélien A de \mathbb{R}^{d+q} . Par ailleurs, pour toute fonction g mesurable définie sur \mathbb{R}^{d+q} et à valeur dans \mathbb{R} , on note :

$$\begin{aligned} G_n(g) &= \int_{\mathbb{R}^{d+q}} g(\mathbf{u}) dG_n(\mathbf{u}), & \mathcal{N}_n(g) &= \int_{\mathbb{R}^{d+q}} g(\mathbf{u}) d\mathcal{N}_n(\mathbf{u}), \\ \mu(g) &= \mathbb{E}g(\mathbf{U}) & \text{et } \bar{\sigma}(g) &= \sqrt{\text{Var}(g(\mathbf{U}))}. \end{aligned}$$

On introduit, de plus, une suite $\{a_n : n \geq 1\}$ de constantes positives, tendant vers 0, et vérifiant

$$|\log a_n| / \log \log n \rightarrow \infty. \quad (3.75)$$

Soit $\{k_n, n \geq 1\}$ une suite de constantes positives tendant vers l'infini lorsque $n \rightarrow \infty$. On considère la suite $\mathcal{G}_n = \{g_i^{(n)} : i = 1, \dots, k_n\}$ d'ensembles de fonctions mesurables définies sur \mathbb{R}^{d+q} et à valeur dans \mathbb{R} telles que :

$$\mathbb{P}\{g_i^{(n)}(\mathbf{U}) \neq 0, g_j^{(n)}(\mathbf{U}) \neq 0\} = 0, \quad \forall 1 \leq i \neq j \leq k_n, \quad \text{et} \quad \sum_{i=1}^{k_n} \mathbb{P}\{g_i^{(n)}(\mathbf{U}) \neq 0\} \leq 1/2.$$

On supposera que :

(R.1) $\exists 0 < r < \infty$ tel que $a_n k_n \rightarrow r$ lorsque $n \rightarrow \infty$;

(R.2) $\exists -\infty < \mu_1, \mu_2 < \infty$ tels que, uniformément en $i = 1, \dots, k_n$, *apcr*,

$$a_n \mu_1 \leq \mu(g_i^{(n)}) \leq a_n \mu_2; \quad (3.76)$$

(R.3) $\exists 0 < \sigma_1 < \sigma_2 < \infty$, tels que, uniformément en $i = 1, \dots, k_n$, *apcr*,

$$\sigma_1 \sqrt{a_n} \leq \bar{\sigma}(g_i^{(n)}) \leq \sigma_2 \sqrt{a_n}; \quad (3.77)$$

(R.4) $\exists 0 < M_2 < \infty$ tel que, uniformément en $i = 1, \dots, k_n$, *apcr*, $\|g_i^{(n)}\| \leq M_2$.

Lemme 3.2.1. *Sous les hypothèses (R.1-2-3-4), pour tout $0 < \epsilon < 1$, il existe presque sûrement un N_ϵ tel que pour tout $n \geq N_\epsilon$,*

$$\max_{1 \leq i \leq k_n} \frac{\sqrt{n} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)}) \sqrt{2 \log(1/a_n)}} \geq 1 - \epsilon. \quad (3.78)$$

Démonstration. La démonstration est identique à celle présentée par Einmahl et Mason (2000). Il est montré, en utilisant un résultat emprunté à Deheuvels et Mason (1992) et basé sur l'approximation par les processus de Poisson, que, pour tout ξ compris entre $(1 - \epsilon/2)(1 - (\epsilon/2)^2)$ et 1, on a :

$$\mathbb{P} \left\{ \max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)}) \sqrt{2 \log(1/a_n)}} < 1 - \epsilon \right\} \leq \left(1 - k_n^{-\xi}\right)^{k_n}.$$

Comme (3.75) et (R.1) assurent conjointement que $\sum_{n=1}^{\infty} (1 - k_n^{-\xi})^{k_n} < \infty$, on aboutit au résultat du lemme. \square

Lemme 3.2.2. *Sous les hypothèses du théorème 3.2.1, on a, presque sûrement,*

$$\liminf_{n \rightarrow \infty} \left\{ \sup_{\mathbf{x} \in I} \frac{W_{n, h'_n}(\mathbf{x}, \psi)}{\sqrt{2n h_n'^d \log(1/h_n'^d)}} \right\} \geq \sigma(\psi). \quad (3.79)$$

Démonstration. Pour tout $\varepsilon > 0$, on sélectionne un sous-ensemble \mathcal{I} de I tel que :

$$\inf_{\mathbf{x} \in \mathcal{I}} \mathbb{E} \left((c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}))^2 \mid \mathbf{X} = \mathbf{x} \right) f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(t) dt > \sigma^2(\psi)(1 - \epsilon/2), \quad (3.80)$$

où $\sigma^2(\psi)$ est définie comme précédemment, et

$$\mathbb{P} \{ \mathbf{X} \in \mathcal{I} \} \leq 1/2. \quad (3.81)$$

Notons que l'existence d'un tel sous-ensemble \mathcal{I} de I est assurée par la continuité de la fonction

$$\mathbf{x} \rightarrow \mathbb{E} \left((c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}))^2 \mid \mathbf{X} = \mathbf{x} \right) f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(t) dt$$

(notamment sous $(F.I - II)$). On cherche maintenant à appliquer le lemme 3.2.1. On sélectionne alors une grille de \mathcal{I} de pas h'_n , i.e. une suite de cubes *disjoints* d'arrête h'_n :

$$\Gamma_{i,n} = \left\{ \mathbf{x}_{i,n} +] - h'_n/2, h'_n/2]^d \right\}, \quad (3.82)$$

$1 \leq i \leq k_n$, avec $k_n \leq C_{\mathcal{I}}/h_n'^d$, $C_{\mathcal{I}}$ ne dépendant que des dimensions de \mathcal{I} . Pour tout $\mathbf{x}_{i,n}$, soit la fonction, définie sur $\mathbb{R}^d \times \mathbb{R}^q$,

$$g_i^{(n)}(\mathbf{u}, \mathbf{v}) := \left\{ c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{v}) + d_\psi(\mathbf{x}_{i,n}) \right\} K \left(\frac{\mathbf{x}_{i,n} - \mathbf{u}}{h'_n} \right). \quad (3.83)$$

On a $\|g_i^{(n)}\| \leq \{\|c_\psi\|_{I^\alpha} \|\psi\| + \|d_\psi\|_{I^\alpha}\} \|K\| := M_2$. De plus, comme $K(\mathbf{s}) = 0$ pour $|\mathbf{s}| \geq 1/2$, $\mathbb{P}(g_i^{(n)}(\mathbf{X}, \mathbf{Z}) \neq 0, g_j^{(n)}(\mathbf{X}, \mathbf{Z}) \neq 0) = 0$ pour $1 \leq i, j \leq k_n$. Par ailleurs,

$$\begin{aligned} \mathbb{P}(g_i^{(n)}(\mathbf{X}, \mathbf{Z}) \neq 0) &= \mathbb{P}\left[\left(c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right)K\left(\frac{\mathbf{x}_{i,n} - \mathbf{X}}{h'_n}\right) \neq 0\right] \\ &\leq \mathbb{P}\left[K\left(\frac{\mathbf{x}_{i,n} - \mathbf{X}}{h'_n}\right) \neq 0\right] \\ &\leq \mathbb{P}(\mathbf{X} \in [\mathbf{x}_{i,n} - h'_n/2, \mathbf{x}_{i,n} + h'_n/2]), \\ \text{d'où } \sum_{i=1}^{k_n} \mathbb{P}(g_i^{(n)}(\mathbf{X}, \mathbf{Z}) \neq 0) &\leq \sum_{i=1}^{k_n} \mathbb{P}(\mathbf{X} \in \Gamma_{i,n}) \\ &= \mathbb{P}(\mathbf{X} \in \mathcal{I}) \leq 1/2. \end{aligned}$$

D'autre part,

$$\begin{aligned} \text{Var}[g_i^{(n)}(\mathbf{X}, \mathbf{Z})] &\leq \mathbb{E}\left[\left(g_i^{(n)}(\mathbf{X}, \mathbf{Z})\right)^2\right] \\ &= \mathbb{E}\left[\left\{c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right\}^2 K^2\left(\frac{\mathbf{x}_{i,n} - \mathbf{X}}{h'_n}\right)\right] \\ &= \int_{\mathbf{R}^d} \mathbb{E}\left[\left\{c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right\}^2 \mid \mathbf{X} = \mathbf{t}\right] K^2\left(\frac{\mathbf{x}_{i,n} - \mathbf{t}}{h'_n}\right) f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t} \\ &=: \mathcal{V}_{i,n}. \end{aligned}$$

Soit $\theta(\mathbf{t}) = \mathbb{E}\left[\left\{c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right\}^2 \mid \mathbf{X} = \mathbf{t}\right] f_{\mathbf{X}}(\mathbf{t})$. Comme θ est continue et bornée, on peut appliquer le lemme de Bochner avec " $K = K^2$ ". On obtient alors que $\mathcal{V}_{i,n} \sim h_n'^d \theta(\mathbf{x}_{i,n}) \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t}$, lorsque $n \rightarrow \infty$. Finalement, d'après la définition de $\sigma^2(\psi)$, on a, *apcr*,

$$\forall \varepsilon > 0, \text{Var}\left[g_i^{(n)}(\mathbf{X}, \mathbf{Z})\right] \leq \sigma^2(\psi)(1 + \varepsilon)h_n'^d. \quad (3.84)$$

De plus,

$$\begin{aligned} A &:= \mathbb{E}\left[\left\{c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right\}^2 K^2\left(\frac{\mathbf{x}_{i,n} - \mathbf{X}}{h'_n}\right)\right] \\ &= \int_{\mathbf{R}^d} \mathbb{E}\left[\left\{c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right\}^2 \mid \mathbf{X} = \mathbf{t}\right] f_{\mathbf{X}}(\mathbf{t}) K^2\left(\frac{\mathbf{x}_{i,n} - \mathbf{t}}{h'_n}\right) d\mathbf{t} \\ &\geq \sigma^2(\psi)(1 - \varepsilon/2)h_n'^d, \end{aligned} \quad (3.85)$$

où nous avons utilisé la condition (3.80). D'autre part, en appliquant une nouvelle fois le lemme de Bochner, il est immédiat que

$$B^2 := \left\{\mathbb{E}\left[\left\{c_\psi(\mathbf{x}_{i,n})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}_{i,n})\right\} K\left(\frac{\mathbf{x}_{i,n} - \mathbf{X}}{h'_n}\right)\right]\right\}^2 = O(h_n'^{2d}) = o(h_n'^d). \quad (3.86)$$

Et donc, on a, *apcr*,

$$\forall \varepsilon > 0, B^2 \leq \frac{\varepsilon}{2} h_n'^d \sigma^2(\psi). \quad (3.87)$$

En combinant ce résultat à (3.84) et (3.85), il vient, pour tout $\varepsilon > 0$,

$$\sigma^2(\psi)(1 - \varepsilon)h_n'^d \leq \bar{\sigma}^2(g_i^{(n)}) := \text{Var}(g_i^{(n)}(\mathbf{X}, \mathbf{Z})) \leq \sigma^2(\psi)(1 + \varepsilon)h_n'^d. \quad (3.88)$$

On conclut que (3.77) est vérifiée pour le choix particulier $a_n = h_n'^d$. De la même manière, on peut montrer que (3.76) est vérifiée pour ce même choix $a_n = h_n'^d$. Nous pouvons alors appliquer le lemme 3.2.1, qui implique que, presque sûrement,

$$\liminf_{n \rightarrow \infty} \left\{ \sup_{\mathbf{x} \in I} \frac{W_{n, h_n'}(\mathbf{x}, \psi)}{\sqrt{2nh_n'^d \log(1/h_n'^d)}} \right\} \geq \sigma(\psi)(1 - \varepsilon)^{3/2}. \quad (3.89)$$

Comme $\varepsilon > 0$ peut être choisi indifféremment petit, l'assertion du lemme 3.2.2 est finalement démontrée \square .

D'après ce dernier résultat, on a, avec probabilité un,

$$\forall \varepsilon > 0, \liminf_{n \rightarrow \infty} \left\{ \sup_{h \in [h_n', h_n'']} \frac{\sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} W_{n, h}(\mathbf{x}, \psi)}{\sqrt{2nh^d \log(1/h^d)}} \right\} \geq \sigma(\psi)(1 - \varepsilon), \quad (3.90)$$

ce qui achève la démonstration de la partie « borne inférieure ».

En combinant la proposition 3.2.2 au résultat (3.33), la démonstration du Théorème 3.2.1 est immédiate. Le résultat est donc établi pour une fonction ψ , définie sur \mathbb{R}^d , mesurable, à valeur réelle, uniformément bornée et telle que $\sigma(\psi) > 0$, fixée.

3.2.2 Démonstration du théorème 3.1.3 dans le cas général

Seule la partie « borne supérieure » est à traiter. On déduit de ce qui précède que, pour toute sous-classe de fonctions finie $\mathcal{G} \subset \mathcal{F}$, on a, avec probabilité un,

$$\limsup_{n \rightarrow \infty} \sup_{h \in [h_n', h_n'']} \frac{\sup_{\psi \in \mathcal{G}} \sup_{\mathbf{x} \in I} |W_{n, h}(\mathbf{x}, \psi)|}{\sqrt{2nh^d \log(1/h^d)}} \leq \sup_{\psi \in \mathcal{G}} \sigma(\psi). \quad (3.91)$$

Il reste à montrer comment étendre le résultat (3.91) à la classe \mathcal{F} entière. Cette extension est obtenue à partir de la série de lemmes suivants. Posons

$$C_{\mathcal{F}} := \sup\{\|c_{\psi}\|_{I^{\alpha}} : \psi \in \mathcal{F}\} \quad \text{et} \quad D_{\mathcal{F}} := \sup\{\|d_{\psi}\|_{I^{\alpha}} : \psi \in \mathcal{F}\}.$$

Remarquons que d'après l'hypothèse (F.III), la classe \mathcal{F} possède une fonction enveloppe Ψ uniformément bornée, où $\Psi(\mathbf{z}) \geq \sup_{\psi \in \mathcal{F}} \psi(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^d$.

Lemme 3.2.3. *Pour tout $\varepsilon > 0$, il existe une sous-classe finie $\mathcal{G}_{\varepsilon} \subset \mathcal{F}$ telle que, pour tout $\psi \in \mathcal{F}$, apcr,*

$$\min_{\phi \in \mathcal{G}_{\varepsilon}} \sup_{\substack{h \in [h_n', h_n''] \\ \mathbf{x} \in I}} \frac{1}{h^d} \mathbb{E} \left[\left\{ c_{\psi}(\mathbf{x})\psi(\mathbf{Z}) + d_{\psi}(\mathbf{x}) - c_{\phi}(\mathbf{x})\phi(\mathbf{Z}) - d_{\phi}(\mathbf{x}) \right\}^2 K^2 \left(\frac{\mathbf{X} - \mathbf{x}}{h} \right) \right] \leq \varepsilon. \quad (3.92)$$

Démonstration. Soit J un compact de \mathbb{R}^d . Pour $\psi, \phi \in \mathcal{F}$, $\mathbf{x} \in I$ et $h \in [h_n', h_n'']$, on a

$$\begin{aligned} & \frac{1}{h^d} \mathbb{E} \left[\left\{ c_{\psi}(\mathbf{x})\psi(\mathbf{Z}) + d_{\psi}(\mathbf{x}) - c_{\phi}(\mathbf{x})\phi(\mathbf{Z}) - d_{\phi}(\mathbf{x}) \right\}^2 K^2 \left(\frac{\mathbf{X} - \mathbf{x}}{h} \right) \right] \\ & \leq \frac{1}{h^d} \mathbb{E} \left[\left\{ c_{\psi}(\mathbf{x})\psi(\mathbf{Z}) + d_{\psi}(\mathbf{x}) - c_{\phi}(\mathbf{x})\phi(\mathbf{Z}) - d_{\phi}(\mathbf{x}) \right\}^2 \mathbb{1}_{\{|\mathbf{X} - \mathbf{x}| \leq h/2\}} \right] \|K^2\|_{\infty} \\ & \leq \frac{4}{h^d} \mathbb{E}[(C_{\mathcal{F}}\Psi(\mathbf{Z}) + D_{\mathcal{F}})^2 \mathbb{1}_{\{\mathbf{Z} \notin J, |\mathbf{X} - \mathbf{x}| \leq h/2\}}] \|K^2\|_{\infty} \\ & \quad + \eta(J) \int_J \{c_{\psi}(\mathbf{x})\psi(\mathbf{Z}) + d_{\psi}(\mathbf{x}) - c_{\phi}(\mathbf{x})\phi(\mathbf{Z}) - d_{\phi}(\mathbf{x})\}^2 d\mathbf{z}, \\ & =: (I) + (II). \end{aligned} \quad (3.93)$$

où $\eta(J) := \sup_{(\mathbf{x}, \mathbf{z}) \in I^\alpha \times J} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}) \|K^2\|_\infty$.

Pour traiter le terme (I), remarquons que d'après l'inégalité d'Hölder, on a, pour tout $s > 1$,

$$(I) \leq 4 \|K^2\|_\infty \tilde{\alpha}^{2/s} \|f_{\mathbf{X}}\|_{I^\alpha}^2 \sup_{\mathbf{x} \in I} \mathbb{P}(\mathbf{Z} \notin J | \mathbf{X} = \mathbf{x})^{1-2/s},$$

où $\tilde{\alpha} := \sup_{\mathbf{x} \in I} \mathbb{E}[(C_{\mathcal{F}}\Psi(\mathbf{Z}) + D_{\mathcal{F}})^s | \mathbf{X} = \mathbf{x}] < \infty$. Sous nos hypothèses de continuité, et d'après le lemme de Scheffé, la fonction $\mathbf{x} \rightarrow \mathbb{P}(\mathbf{Z} \in \cdot | \mathbf{X} = \mathbf{x})$ est continue de I vers l'espace de toutes les mesures de probabilité sur les boréliens de \mathbb{R} muni de la topologie de la convergence faible. Ainsi, l'ensemble $\{\mathbb{P}(\mathbf{Z} \in \cdot | \mathbf{X} = \mathbf{x}), \mathbf{x} \in I\}$ est compact dans cet espace, ce qui entraîne, au vu du théorème de Prohorov, qu'il est uniformément tendu. Autrement dit, pour tout $\varepsilon > 0$, il existe un compact $J_\varepsilon \subset \mathbb{R}^q$ tel que $\mathbb{P}(\mathbf{Z} \in J_\varepsilon | \mathbf{X} = \mathbf{x}) > 1 - \varepsilon/2$. Finalement, on en déduit que, uniformément en $\mathbf{x} \in I$, il existe un compact $J = J_\varepsilon$ tel que,

$$(I) \leq \frac{\varepsilon}{2}.$$

Pour traiter (II), nous allons utiliser le fait que \mathcal{F} est une classe de fonctions VC. Cela assure en effet qu'elle est totalement bornée par rapport à d_{Q_ε} , où Q_ε est la loi uniforme sur J^ε . On en déduit que pour tout $\delta > 0$, il existe une sous-classe finie $\mathcal{G}_1 \subset \mathcal{F}$ pour laquelle

$$\sup_{\psi \in \mathcal{F}} \min_{\phi \in \mathcal{G}_1} \int_{J^\varepsilon} (\psi(\mathbf{z}) - \phi(\mathbf{z}))^2 d\mathbf{z} < \delta.$$

Par ailleurs, la relative compacité des classes $\mathcal{F}_{\mathcal{C}}$ et $\mathcal{F}_{\mathcal{D}}$ entraîne l'existence de sous-classes finies $\mathcal{G}_2, \mathcal{G}_3 \subset \mathcal{F}$ telles que

$$\sup_{\psi \in \mathcal{F}} \min_{\phi \in \mathcal{G}_2} \|c_\psi - c_\phi\|_I \vee \sup_{\psi \in \mathcal{F}} \min_{\phi \in \mathcal{G}_3} \|d_\psi - d_\phi\|_I < \delta.$$

En combinant les faits

$$\sup_{\psi \in \mathcal{F}} \|c_\psi\|_{I^\alpha} < \infty \text{ et } \sup_{\psi \in \mathcal{F}} \int_{J^\varepsilon} \psi(\mathbf{z})^2 d\mathbf{z} < \infty,$$

et en choisissant un $\delta > 0$ suffisamment petit, il vient

$$\sup_{\psi \in \mathcal{F}} \min_{\phi_1, \phi_2, \phi_3} \sup_{\mathbf{x} \in I} \int_{J^\varepsilon} \{c_\psi(\mathbf{x})\psi(\mathbf{Z}) + d_\psi(\mathbf{x}) - c_{\phi_2}(\mathbf{x})\phi_1(\mathbf{Z}) - d_{\phi_3}(\mathbf{x})\}^2 d\mathbf{z} \leq \varepsilon/(4\eta(J_\varepsilon)),$$

où le minimum est pris sur $\mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3$. Pour tout triplet $(\phi_1, \phi_2, \phi_3) \in \mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3$ pour lequel il existe un $\phi \in \mathcal{F}$ tel que

$$\sup_{\mathbf{x} \in I} \int_{J^\varepsilon} \{c_\phi(\mathbf{x})\phi(\mathbf{Z}) + d_\phi(\mathbf{x}) - c_{\phi_2}(\mathbf{x})\phi_1(\mathbf{Z}) - d_{\phi_3}(\mathbf{x})\}^2 d\mathbf{z} \leq \varepsilon/(4\eta(J_\varepsilon)),$$

on sélectionne un de ces $\phi \in \mathcal{F}$ pour former la classe \mathcal{G}_ε . En appliquant l'inégalité triangulaire, il vient facilement

$$(II) \leq \frac{\varepsilon}{2},$$

ce qui conclut la démonstration du lemme. \square

Soient $\varepsilon > 0$ et n_0 tel que (3.92) soit vrai pour tout $n \geq n_0$. Pour tout $\psi, \phi \in \mathcal{F}$, on définit,

$$d^2(\psi, \phi) := \sup_{n \geq n_0} \sup_{h \in [h'_n, h''_n]} h^{-d} \sup_{\mathbf{x} \in I} \mathbb{E} \left[\left\{ c_\psi(\mathbf{x}) \psi(\mathbf{Z}) + d_\psi(\mathbf{x}) - c_\phi(\mathbf{x}) \phi(\mathbf{Z}) - d_\phi(\mathbf{x}) \right\}^2 K^2 \left(\frac{\mathbf{X} - \mathbf{x}}{h} \right) \right].$$

Lemme 3.2.4. *Sous les hypothèses du théorème 3.1.3, il existe une constante universelle A telle que, presque sûrement,*

$$\limsup_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \frac{\sup_{d^2(\psi, \phi) \leq \varepsilon} \sup_{\mathbf{x} \in I} |W_{n,h}(\mathbf{x}, \psi) - W_{n,h}(\mathbf{x}, \phi)|}{\sqrt{nh^d \log(1/h^d)}} \leq AD_1(\nu) \sqrt{\varepsilon}. \quad (3.94)$$

Démonstration. Considérons les classes de fonctions

$$\tilde{\mathcal{F}}(\varepsilon, h) := \{\eta_{\mathbf{x}, h, \psi} - \eta_{\mathbf{x}, h, \phi} : d^2(\psi, \phi) \leq \varepsilon, \mathbf{x} \in I\}, \text{ pour } h \in [h'_n, h''_n],$$

où les fonctions $\eta_{\mathbf{x}, h, \psi}$ sont définies en (3.35). Pour tout entier $k \geq 1$, on pose $n_k = 2^k$. En posant

$$Q_k := \max_{n_{k-1} \leq n \leq n_k} \sup_{h \in [h'_n, h''_n]} \frac{\sup_{d^2(\psi, \phi) \leq \varepsilon} \sup_{\mathbf{x} \in I} |W_{n,h}(\mathbf{x}, \psi) - W_{n,h}(\mathbf{x}, \phi)|}{\sqrt{2nh^d \log(1/h^d)}},$$

il vient

$$Q_k = \max_{n_{k-1} \leq n \leq n_k} \sup_{h \in [h'_n, h''_n]} \frac{\|n^{1/2} \Upsilon_n\|_{\tilde{\mathcal{F}}(\varepsilon, h)}}{\sqrt{2nh^d \log(1/h^d)}}.$$

Considérons, comme précédemment, la discrétisation suivante de l'intervalle $[h'_{n_k}, h''_{n_{k-1}}] \supset [h'_n, h''_n]$, pour $n_{k-1} \leq n \leq n_k$.

$$\begin{aligned} h'_{n_k, R_k} &:= h''_{n_{k-1}} \\ h'_{n_k, l} &:= \lambda^l h'_{n_k}, \quad l = 0 \dots R_k - 1, \end{aligned} \quad (3.95)$$

où R_k vérifie la condition

$$R_k = \left\lfloor \frac{\log(h''_{n_{k-1}}/h'_{n_k})}{\log(\lambda)} \right\rfloor + 1. \quad (3.96)$$

Posons alors, pour tout $l \geq 0$,

$$\tilde{\mathcal{F}}_{k,l}(\varepsilon) := \{\eta_{\mathbf{x}, h, \psi} - \eta_{\mathbf{x}, h, \phi} : d^2(\psi, \phi) \leq \varepsilon, \mathbf{x} \in I, h \in [h'_{n_k, l}, h'_{n_k, l+1}]\}. \quad (3.97)$$

Au vu la définition des classes $\tilde{\mathcal{F}}_{k,l}(\varepsilon)$, on a, *apcr*, pour toute fonction $v \in \tilde{\mathcal{F}}_{k,l}(\varepsilon)$,

$$h^{-d} \text{Var}(v(\mathbf{X}, \mathbf{Z})) \leq d^2(\psi, \phi). \quad (3.98)$$

Ainsi, d'après (3.97) et (3.98), en posant

$$\tilde{\sigma}_{k,l,\varepsilon}^2 := \sup\{\text{Var}(v(\mathbf{X}, \mathbf{Z})) : v \in \tilde{\mathcal{F}}_{k,l}(\varepsilon)\},$$

on a $\tilde{\sigma}_{k,l,\varepsilon}^2 \leq \lambda^{dl} \varepsilon h_{n_k}^d$.

Par ailleurs, pour toute fonction $v \in \tilde{\mathcal{F}}_{k,l}(\varepsilon)$, on a $\|v\| \leq M_1$ (où M_1 est défini en (3.46)). On peut, de plus, montrer que chaque classe $\tilde{\mathcal{F}}_{k,l}(\varepsilon)$ est séparable point par point et telle que $\tilde{\mathcal{F}}_{k,l}(\varepsilon) \subset \tilde{\mathcal{G}}$, où $\tilde{\mathcal{G}}$ est une classe de fonctions vérifiant la condition $[\mathcal{E}]$ (pour

ce faire, on utilise les mêmes arguments que ceux utilisés pour établir que \mathcal{G}' était VC , ainsi que la propriété VC de \mathcal{F} ; voir, par exemple, la démonstration du lemme 5 dans Einmahl et Mason (2000). Nous sommes donc sous les conditions d'application du fait 3.4.3, avec $\tau = \sqrt{\lambda^{dl}}\varepsilon$ et $\rho = \tau\sqrt{1/A_2}$. D'après celui-ci, on a, avec comme d'habitude $N_k = \{n_{k-1} + 1, \dots, n_k\}$, et en reprenant la méthodologie menant à (3.60),

$$\begin{aligned} \mathbb{P}_{\varepsilon,k,l} &:= \mathbb{P}\left[\max_{1 \leq n \leq n_k} \frac{n^{1/2} \|\Upsilon_n\|_{\tilde{\mathcal{F}}_{k,l}(\varepsilon)}}{\sqrt{2n_k h_{n_k,l}^{td} \log(1/h_{n_k,l}^{td})}} \geq \sqrt{\varepsilon \lambda^{dl}} (1 + \sqrt{1/A_2}) D_1(\nu)\right] \\ &\leq 4h_{n_k,l}^{td}. \end{aligned}$$

Ainsi,

$$\begin{aligned} \mathbb{P}_{\varepsilon,k} &:= \mathbb{P}\left[\max_{0 \leq l \leq R_k-1} \max_{1 \leq n \leq n_k} \frac{n^{1/2} \|\Upsilon_n\|_{\tilde{\mathcal{F}}_{k,l}(\varepsilon)}}{\sqrt{2n_k h_{n_k,l}^{td} \log(1/h_{n_k,l}^{td})}} \geq \sqrt{\varepsilon \lambda^{dl}} (1 + \sqrt{1/A_2}) D_1(\nu)\right] \\ &\leq 4 \sum_{l=0}^{R_k-1} h_{n_k,l}^{td} \\ &\leq \frac{4}{\lambda^d - 1} h_{n_{k-1}}^{td}, \end{aligned}$$

ce qui conduit au résultat du lemme 3.2.4, *via* le lemme de Borel-Cantelli. \square

En combinant les lemmes 3.2.3 et 3.2.4 au résultat (3.91), on obtient l'uniformité sur la classe \mathcal{F} , ce qui conclut la démonstration du théorème 3.1.3 dans le cas général.

3.3 Démonstration du corollaire 3.1.3

Dans cette partie, on montre comment les résultats obtenus sur le processus $W_{n,h}(\mathbf{x}, \psi)$ se transposent aux estimateurs non paramétriques fonctionnels de la régression (pour les estimateurs de la densité, la démonstration est directe et omise). La démonstration est quasiment identique à celles présentées par Einmahl et Mason (2000) et Deheuvels et Mason (2004). Nous la rappelons cependant compte tenu de son intérêt. Pour ce faire, on utilise les définitions (3.11), (3.12), (3.15) et (3.16) de $f_{\mathbf{X};n}(\mathbf{x}, h)$, $r_{\psi;n}(\mathbf{x}, h)$, $\tilde{f}_{\mathbf{X}}(\mathbf{x}, h)$ et $\tilde{r}_{\psi}(\mathbf{x}, h)$. En prenant $c_{\psi}(\mathbf{x}) = 1/f_{\mathbf{X}}(\mathbf{x})$ et $d_{\psi}(\mathbf{x}) = -m_{\psi}(\mathbf{x})/f_{\mathbf{X}}(\mathbf{x})$ dans la définition (3.20) de $W_{n,h}(\mathbf{x}, \psi)$, on a la relation suivante, pour tout $h > 0$,

$$W_{n,h}(\mathbf{x}, \psi) = nh^d \left\{ \frac{r_{\psi;n}(\mathbf{x}, h)}{f_{\mathbf{X}}(\mathbf{x})} - \frac{\tilde{r}_{\psi}(\mathbf{x}, h)}{f_{\mathbf{X}}(\mathbf{x})} - \frac{r_{\psi}(\mathbf{x}) \{f_{\mathbf{X};n}(\mathbf{x}, h) - \tilde{f}_{\mathbf{X}}(\mathbf{x}, h)\}}{f_{\mathbf{X}}(\mathbf{x})^2} \right\}. \quad (3.99)$$

Lemme 3.3.1. *Sous les hypothèses du théorème 3.1.3, on a, presque sûrement,*

$$\begin{aligned} \mathcal{E}_n &:= \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{1}{\sqrt{nh^d \log(1/h^d)}} \left| W_{n,h}(\mathbf{x}, \psi) \right. \\ &\quad \left. - nh^d \left(m_{\psi;n}(\mathbf{x}, h) - \frac{r_{\psi;n}(\mathbf{x}, h)}{f_{\mathbf{X};n}(\mathbf{x}, h)} \right) \right| = o(1). \end{aligned} \quad (3.100)$$

Démonstration. Appliquons le théorème 3.1.3 avec $c_{\psi}(\mathbf{x}) = 1$ et $d_{\psi}(\mathbf{x}) = 0$, puis avec $c_{\psi}(\mathbf{x}) = 0$ et $d_{\psi}(\mathbf{x}) = 1$ (remarquons que les classes $\mathcal{F}_{\mathcal{C}}$ et $\mathcal{F}_{\mathcal{D}}$ associées vérifient trivialement la condition de relative compacité par rapport à la topologie engendrée par la norme uniforme sur I^α). On obtient,

- $\sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \sqrt{nh^d [\log(1/h^d)]^{-1}} |r_{\psi;n}(\mathbf{x}, h) - \tilde{r}_\psi(\mathbf{x}, h)| = \mathcal{O}(1)$ p.s.
- et
- $\sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} \sqrt{nh^d [\log(1/h^d)]^{-1}} |f_{\mathbf{X};n}(\mathbf{x}, h) - \tilde{f}_{\mathbf{X}}(\mathbf{x}, h)| = \mathcal{O}(1)$ p.s. .

De plus, d'après le fait 3.4.2,

- $\sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} |r_\psi(\mathbf{x}) - \tilde{r}_\psi(\mathbf{x}, h)| = o(1)$ p.s.
- et
- $\sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} |f_{\mathbf{X}}(\mathbf{x}) - \tilde{f}_{\mathbf{X}}(\mathbf{x}, h)| = o(1)$ p.s. .

Le résultat du lemme 3.3.1 est alors immédiat, au vu de la condition (F.IV). \square

En combinant le lemme 3.3.1 et le théorème 3.1.3, le corollaire 3.1.3 est immédiat.

3.4 Appendice

3.4.1 Classes de fonctions séparables point par point

Nous rappelons ici la définition des classes de fonctions séparables point par point (cf. p.110 dans van der Vaart et Wellner (1996)).

Définition. Soit \mathcal{G} une classe de fonctions définies sur un ensemble \mathcal{X} . Soit \mathcal{G}_1 un sous-ensemble dénombrable de \mathcal{G} . Si, pour tout $g \in \mathcal{G}$, il existe une suite $\{g_m\}_{m \geq 1} \in \mathcal{G}_1$ telle que $g_m(x) \rightarrow g(x)$ lorsque $n \rightarrow \infty$, pour tout $x \in \mathcal{X}$, alors \mathcal{G} est séparable point par point.

3.4.2 Faits importants

Nous présentons dans cette section les résultats techniques dont nous avons fait usage dans ce chapitre. Les démonstrations sont omises.

Fait 3.4.1. (*Inégalité maximale de Bernstein*). Soit U_1, \dots, U_n des variables indépendantes de moyenne nulle et de variance $\sigma_j^2 = \text{Var}(U_j) < \infty$, $j = 1, \dots, n$. Supposons de plus qu'il existe une constante $0 < M < \infty$ telle que pour tout $1 \leq j \leq n$, $|U_j| \leq M$. Alors, pour tout $x \geq 0$, on a

$$\mathbb{P}\left(\max_{1 \leq k \leq n} \sum_{j=1}^k U_j \geq x\right) \leq \exp\left(-\frac{x^2}{2 \sum_{j=1}^n \sigma_j^2 + \frac{2}{3}Mt}\right). \quad (3.101)$$

Le résultat suivant est une version améliorée du lemme de Bochner. Sa démonstration figure dans Einmahl et Mason (2005).

Fait 3.4.2. Soit I un pavé compact de \mathbb{R}^d , $d \geq 1$. Supposons que \mathcal{H} est une famille uniformément bornée et équicontinue de fonctions φ définies sur I^α , l' α -voisinage de I , et à valeur réelle. Supposons que K est un noyau à support dans $[-1/2, 1/2]^d$ tel que

$$\int_{\mathbb{R}^d} K(\mathbf{u}) d\mathbf{u} = 1.$$

Alors on a, pour toute suite de constantes positives $b_n \rightarrow 0$,

$$\sup_{\varphi \in \mathcal{H}} \sup_{0 < h < b_n} \|\varphi * K_h - \varphi\|_I \rightarrow 0,$$

où

$$\varphi * K_h(\mathbf{x}) := h^{-1} \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) \varphi(\mathbf{t}) d\mathbf{t}.$$

Le fait suivant a été présenté par Varron (2006) (voir également Mason (2004)).

Fait 3.4.3. *Soit \mathcal{F} une classe de fonctions séparable point par point vérifiant*

$$\sup_{f \in \mathcal{F}} \text{Var}(f(\mathbf{Z})) \leq \tau^2 h,$$

avec $\tau, h > 0$. Supposons qu'il existe $M, C, \nu > 0$ tels que, pour tout $0 < \epsilon < 1$,

$$\mathcal{N}(\epsilon, \mathcal{F}) \leq C\epsilon^{-\nu}, \quad (3.102)$$

$$\sup_{f \in \mathcal{F}, \mathbf{z} \in \mathbb{R}^d} |f(\mathbf{z})| \leq M. \quad (3.103)$$

Sélectionnons $\rho > 0$ arbitrairement. Alors il existe une constante universelle $A_2 > 0$ et un paramètre $D_1(\nu) > 0$ dépendant seulement de ν tels que, si $h > 0$ vérifie

$$K_1 := \max \left\{ \frac{4M\sqrt{\nu+1}}{\tau}, \frac{M\rho}{\tau^2} \right\} \leq \sqrt{\frac{nh}{\log(1/h)}}, \quad (3.104)$$

$$K_2 := \min \left\{ \frac{1}{\tau^2 M}, \tau^2 \right\} \geq h, \quad (3.105)$$

alors on a, avec $T_n(g) = \sum_{j=1}^n \{g(\mathbf{Z}_j) - \mathbb{E}(g(\mathbf{Z}))\}$, $g \in \mathcal{F}$,

$$\mathbb{P} \left(\sup_{1 \leq m \leq n} \|T_m(\cdot)\|_{\mathcal{F}} \geq (\tau + \rho) D_1 \sqrt{nh \log(1/h)} \right) \leq 4 \exp \left(-A_2 \left(\frac{\rho}{\tau}\right)^2 \log(1/h) \right). \quad (3.106)$$

Ci-dessous, nous présentons les deux résultats qui, combinés, conduisent directement au fait 3.4.3 ci-dessus (voir un exemple de démonstration dans Mason (2004)).

Soit $\varepsilon_1, \dots, \varepsilon_n$ une suite de variables i.i.d. de Rademacher, indépendantes de $\mathbf{X}_1, \dots, \mathbf{X}_n$. L'inégalité suivante a été établie par Talagrand (1994).

Fait 3.4.4. *Soit \mathcal{G} une classe séparable point par point de fonctions définies sur \mathbb{R}^d et à valeurs réelles vérifiant, pour une constante $0 < M_0 < \infty$,*

$$\sup_{g \in \mathcal{G}} \|g\| \leq M_0.$$

Alors pour tout $t > 0$ il existe des constantes universelles $A_1, A_2 > 0$ telles que,

$$P \left\{ \max_{1 \leq m \leq n} \|m^{1/2} \alpha_m\|_{\mathcal{G}} \geq A_1 \left(t + \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(\mathbf{X}_i) \right\|_{\mathcal{G}} \right) \right\} \leq 2 \left[\exp \left(-\frac{A_2 t^2}{n \sigma_{\mathcal{G}}^2} \right) + \exp \left(-\frac{A_2 t}{M_0} \right) \right],$$

où $\sigma_{\mathcal{G}}^2 = \sup_{g \in \mathcal{G}} \text{Var}(g(\mathbf{X}))$.

Le fait suivant est dû à Einmahl et Mason (2000).

Fait 3.4.5. *Soit \mathcal{G} une classe séparable point par point de fonctions définies sur \mathbb{R}^d , à valeurs réelles et bornées. Soit G la fonction enveloppe de \mathcal{G} . Supposons qu'il existe des constantes $\beta, \nu, C_0 > 1$, $\sigma \leq 1/(8C_0)$ telles que les conditions (A.1), (A.2), (A.3) et (A.4) ci-dessous sont satisfaites :*

$$(A.1) \quad E[G^2(\mathbf{X})] \leq \beta^2;$$

$$(A.2) \quad N(\epsilon, \mathcal{G}) \leq C_0 \epsilon^{-\nu}, \quad 0 < \epsilon < 1;$$

$$(A.3) \quad \sigma_0^2 := \sup_{g \in \mathcal{G}} E[g^2(\mathbf{X})] \leq \sigma^2;$$

$$(A.4) \quad \sup_{g \in \mathcal{G}} \|g\| \leq \frac{1}{2\sqrt{\nu+1}} \sqrt{n\sigma^2 / \log(\beta \vee 1/\sigma)}.$$

Alors il existe une constante universelle $A_3 > 0$ telle que

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(\mathbf{X}_i) \right\|_{\mathcal{G}} \leq A_3 \sqrt{\nu n \sigma^2 \log(\beta \vee 1/\sigma)}. \quad (3.107)$$

Chapitre 4

Lois du logarithme uniformes pour un estimateur non paramétrique de la régression en données censurées

Avant-propos. L'objet de ce chapitre est de développer et de généraliser les résultats de Viallon (2006b).

Résumé. Dans ce chapitre, nous allons appliquer les résultats du **chapitre 3** précédent au cas des *données censurées*. En particulier, nous établissons, à l'aide de ces méthodes, le comportement asymptotique d'un estimateur de la régression, de type « *Inverse Probability of Censoring Weighted estimator* » [I.P.C.W.], introduit par Kohler *et al.* (2002).

Aparté. L'erratum joint à la version papier relative à ce manuscrit a été « pris en compte » dans cette version pdf. Les erreurs de notation ont ainsi été corrigées, et l'hypothèse d'existence de la densité de la loi de censure ajoutée.

4.1 Introduction

Dans un premier temps, nous rappelons les notations nécessaires, relatives au cas des données censurées, et que nous avons, en grande partie, introduites dans le **chapitre 1** d'introduction. Nous supposons que l'*échantillon de base* (non observé directement) est constitué de copies indépendantes (Y_i, C_i, \mathbf{X}_i) , $i = 1, 2, \dots$, du vecteur aléatoire (Y, C, \mathbf{X}) à valeur dans $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. Ici, Y désigne la variable d'intérêt, C la variable de censure, et \mathbf{X} un vecteur de variables concomitantes, dont le lien avec Y est à étudier. Nous travaillerons avec l'*échantillon observé*

$$\mathcal{D}_n = \{(Z_i, \delta_i, \mathbf{X}_i), i = 1, \dots, n\}, \quad (4.1)$$

pour $n \geq 1$, où $Z_i = Y_i \wedge C_i$ et $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$, $\mathbb{1}_E$ désignant, comme d'habitude, la fonction indicatrice de E . Soit alors, pour $-\infty < t < \infty$, $F(t) = \mathbb{P}(Y \leq t)$, $G(t) = \mathbb{P}(C \leq t)$ et $H(t) = \mathbb{P}(Z \leq t)$, les versions continues à droite des fonctions de répartition de Y , C et Z . Nous supposons dans ce chapitre que la variable C a une densité f_C par rapport à la mesure de Lebesgue sur \mathbb{R} . Par ailleurs, on notera, pour toute fonction de répartition L continue à droite, $T_L = \sup\{t : L(t) < 1\}$, le point supérieur de la distribution correspondante.

Comme dans le chapitre précédent, nous considérons un ensemble (une « classe ») \mathcal{F} de fonctions mesurables, définies cette fois sur \mathbb{R} et à valeur réelle. La classe \mathcal{F} sera supposée, « séparable point par point » (voir l'appendice du **chapitre 3**, et les développements p.110 de van der Vaart et Wellner (1996)), et de type « VC subgraph class » (voir l'annexe du présent mémoire, et le §2.6.2 de van der Vaart et Wellner (1996)). Etant donnée une fonction $\psi \in \mathcal{F}$, nous nous intéressons, dans ce chapitre, à l'espérance conditionnelle de $\psi(Y)$ sachant que $\mathbf{X} = \mathbf{x}$, soit

$$m_\psi(\mathbf{x}) = \mathbb{E}(\psi(Y) \mid \mathbf{X} = \mathbf{x}), \quad (4.2)$$

si tant est que cette fonction existe. Plus explicitement, nous nous intéressons à la fonction de régression de $\psi(Y)$, évaluée, conditionnellement à $\mathbf{X} = \mathbf{x}$, dans le cas où la variable Y est censurée à droite par la variable aléatoire (de censure) C . Comme mentionné par Beran (1981) et Dabrowska (1987), le problème de l'estimation de la fonction de répartition conditionnelle de Y , sachant $\mathbf{X} = \mathbf{x}$, tout comme celui de la fonction de régression m_ψ , sachant $\mathbf{X} = \mathbf{x}$, ne peut être réalisé sans hypothèses complémentaires sur (Z, δ, \mathbf{X}) . Il s'agit d'un problème d'identifiabilité. Pour permettre ces opérations, on travaille, généralement, sous l'hypothèse d'indépendance conditionnelle entre Y et C , sachant \mathbf{X} . Nous ferons donc, par la suite, cette hypothèse.

Dans ce chapitre, nous supposons, en sus de l'hypothèse précédente, que les variables aléatoires \mathbf{X} et C sont mutuellement indépendantes. La combinaison de ces deux hypothèses (à savoir, X et C sont conditionnellement indépendantes, sachant \mathbf{X} , et \mathbf{X} et C sont mutuellement indépendantes) se trouve être équivalente à la condition :

(\mathcal{I}) C et (Y, \mathbf{X}) sont mutuellement indépendantes.

L'hypothèse (\mathcal{I}) est conforme à l'intuition dans la plupart des cas rencontrés en pratique, notamment en épidémiologie, où la censure est généralement fixée par la date de la fin d'étude. Dans ce cas, la censure est, tout naturellement, indépendante des caractéristiques \mathbf{X} de l'individu, et l'hypothèse (\mathcal{I}) est alors satisfaite, dès lors que l'indépendance conditionnelle de Y et C sachant \mathbf{X} a bien lieu.

Remarque 4.1.1. (i) *La condition (\mathcal{I}) est plus forte que celle communément utilisée dans les travaux concernant la régression censurée. Par exemple, Beran (1981) a travaillé sous l'hypothèse d'indépendance conditionnelle de Y et C sachant \mathbf{X} pour construire (et établir des propriétés de) un estimateur de la fonction de répartition conditionnelle (voir aussi Dabrowska (1989) et Deheuveldt et Derzko (2006)). Cependant, l'hypothèse (\mathcal{I}) se trouve être essentielle lorsqu'on utilise les estimateurs de la régression censurée proposés par Carbonez et al. (1995). Elle conduit au résultat suivant. Introduisons une fonction Ψ_ψ , définie dans \mathbb{R}^2 et à valeur dans \mathbb{R} . On pose*

$$\Psi_\psi(y, c) = \frac{\mathbb{I}_{\{y \leq c\}} \psi(y \wedge c)}{1 - G(y \wedge c)}.$$

Alors, sous l'hypothèse (\mathcal{I}), on a

$$\begin{aligned} m_{\Psi_\psi}(\mathbf{x}) &= \mathbb{E}(\Psi_\psi(Y, C) \mid \mathbf{X} = \mathbf{x}) \\ &= \mathbb{E}\left\{ \frac{\mathbb{I}_{\{Y \leq C\}} \psi(Z)}{\bar{G}(Z)} \mid \mathbf{X} = \mathbf{x} \right\} \\ &= \mathbb{E}\left\{ \frac{\psi(Y)}{\bar{G}(Y)} \mathbb{E}(\mathbb{I}_{\{Y \leq C\}} \mid \mathbf{X}, Y) \mid \mathbf{X} = \mathbf{x} \right\} \\ &= m_\psi(\mathbf{x}). \end{aligned}$$

Cette propriété jouera un rôle essentiel dans la démonstration de nos résultats, énoncés dans la suite de ce chapitre.

(ii) L'hypothèse (\mathcal{I}) peut être légèrement affaiblie, en l'hypothèse (\mathcal{I}') ci-dessous.

- (\mathcal{I}') (a) Y et C sont indépendants ;
 (b) $\mathbb{P}(Y \leq C \mid Y, \mathbf{X}) = \mathbb{P}(Y \leq C \mid Y)$.

Ces conditions sont, notamment, utilisées par Stute (1999) et Delecroix et al. (2006).

Dans la littérature scientifique, plusieurs estimateurs non paramétriques ont été proposés pour évaluer m_ψ , dans un contexte de données censurées. Dans le cas où $\psi = \mathbb{I}_{(-\infty, t]}$, pour un $t \in \mathbb{R}$, Beran (1981) a proposé une variante de l'estimateur de Kaplan-Meier parfois appelé *estimateur de Kaplan-Meier local*, ou *conditionnel*. Certaines propriétés asymptotiques de cet estimateur ont été étudiées par Dabrowska (1987), Dabrowska (1989), Dabrowska (1995), Deheuvels et Derzko (2006) et Deheuvels et Mason (2006). D'autres auteurs, comme Fan et Gijbels (1994), ont eux proposé des estimateurs issus de la catégorie dite de « synthetic data ». En particulier, ils obtiennent une version conditionnelle d'un estimateur rentrant dans la catégorie connue sous le nom de « *Self Consistent Estimators* » (*SCE*). Cette dernière catégorie d'estimateurs a été introduite par Efron (1967). Notons qu'en l'absence de covariables, les estimateurs de Kaplan-Meier et les *SCE* coïncident (voir la discussion au chapitre introductif de ce mémoire). Cependant, en présence de covariables, les liens existant entre les estimateurs de Beran (1981) et de Fan et Gijbels (1994) n'ont, à notre connaissance, pas été étudiés. Dans ce chapitre, nous allons étudier un troisième estimateur, du type connu sous le nom de « *Inverse Probability of Censoring Weighted estimator* » [IPCW] (voir, sur ce point, van der Laan et Robins (2002)). Il s'agit d'une généralisation de l'estimateur de Carbonez et al. (1995), introduite par Kohler et al. (2002). De par sa forme, cet estimateur est très simple à décrire. Il a donné lieu, dans la littérature scientifique récente, à de nombreux travaux, tels ceux de Ould-Saïd et Cai (2005), Kohler et al. (2006), Brunel et Comte (2006a), Brunel et Comte (2006b), ainsi que Delecroix et al. (2006). De plus, son utilisation pratique, tout comme sa généralisation au cas du modèle additif, est simple et naturelle. On consultera à ce sujet Brunel et Comte (2006b), ainsi que le **chapitres 5** du présent mémoire.

Après avoir rappelé la définition de cet estimateur, nous établissons ci-dessous des *lois du logarithme* décrivant sa vitesse de convergence uniforme presque sûre. Ces lois sont très semblables à celles que nous avons établies au **chapitre 3** précédent, dans le cadre non censuré. Nous déduisons de ces *lois du logarithme* la vitesse de convergence uniforme presque sûre d'estimateurs non paramétriques de la fonction de répartition conditionnelle en données censurées. Ces résultats sont analogues à ceux obtenus par Deheuvels et Mason (2006) et Deheuvels et Derzko (2006) pour des estimateurs voisins de ceux de Beran (1981). Poursuivant les idées de Deheuvels et Mason (2004), nous construisons, à l'aide de ces résultats, des bandes de confiance asymptotiques simultanées presque sûres pour la fonction de régression définie en (4.2).

4.2 Définition de l'estimateur

À l'instar des estimateurs *IPCW*, l'estimateur de Carbonez et al. (1995) repose sur l'observation suivante. Soit $\ell : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$, une fonction mesurable. Supposons que l'on souhaite estimer $\mathbb{E}(\ell(Y, \mathbf{X}))$ à partir de l'échantillon \mathcal{D}_n (voir (4.1)). Sous l'hypothèse

d'indépendance (\mathcal{I}), un estimateur sans biais de cette quantité est donné par

$$\widehat{\mathbb{E}}(\ell(Y, \mathbf{X})) = n^{-1} \sum_{i=1}^n \frac{\delta_i \ell(Z_i, \mathbf{X}_i)}{1 - G(Z_i)}.$$

En effet, sous l'hypothèse (\mathcal{I}), en utilisant des arguments classiques de conditionnement, il vient

$$\begin{aligned} \mathbb{E}\left\{\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \ell(Z_i, \mathbf{X}_i)}{1 - G(Z_i)}\right\} &= \mathbb{E}\left\{\frac{\mathbb{1}_{\{Y \leq C\}} \ell(Z, \mathbf{X})}{1 - G(Z)}\right\} \\ &= \mathbb{E}\left\{\frac{\ell(Y, \mathbf{X})}{1 - G(Y)} \mathbb{E}\left[\mathbb{1}_{\{Y \leq C\}} \mid \mathbf{X}, Y\right]\right\} \\ &= \mathbb{E}\ell(Y, \mathbf{X}). \end{aligned} \quad (4.3)$$

En pratique, G est inconnue. On ne peut donc pas utiliser l'estimateur tel quel, et il est alors logique de remplacer G dans ces formules par un estimateur de cette même quantité. Pour cela, il est commode de rappeler les définitions des estimateurs de Kaplan-Meier de F et G données dans le **chapitre 2**. Pour tout $y \in \mathbb{R}$, on pose

$$\begin{aligned} F_n^*(y) &= 1 - \prod_{\substack{i: Z_i \leq y \\ 1 \leq i \leq n}} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)}\right)^{\delta_i}, \\ G_n^*(y) &= 1 - \prod_{\substack{i: Z_i \leq y \\ 1 \leq i \leq n}} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)}\right)^{1 - \delta_i}, \end{aligned} \quad (4.4)$$

où $N_n(x) = \sum_{i=1}^n \mathbb{1}_{\{Z_i \geq x\}}$ et avec les conventions $\prod_{\emptyset} = 1$ et $0^0 = 1$. Un estimateur possible de $\mathbb{E}\ell(Y, \mathbf{X})$ est alors donné par $\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \ell(Z_i, \mathbf{X}_i)}{1 - G_n^*(Z_i)}$.

Une autre justification de ce choix provient du résultat suivant,

$$\begin{aligned} M_n &:= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i Z_i}{1 - G_n^*(Z_i)} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i Y_i}{1 - G_n^*(Z_i)} \\ &\stackrel{(1)}{=} \int_0^{\infty} F_n^*(t) dt \\ &= \int_0^{T_F} F_n^*(t) dt, \end{aligned} \quad (4.5)$$

où (1) vient de l'observation de Susarla *et al.* (1984) selon laquelle $\delta_i [n(1 - G_n^*(Z_i))]^{-1}$ correspond au saut de l'estimateur de Kaplan-Meier F_n^* de F en Z_i (pour une justification de ce résultat, voir l'appendice de ce chapitre). Cette observation nous permet d'ailleurs de réécrire l'estimateur de Kaplan-Meier $F_n^*(y)$ de $F(y)$ sous la forme

$$F_n^*(y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbb{1}_{\{Z_i \leq y\}}}{1 - G_n^*(Z_i)}. \quad (4.6)$$

A partir de (4.3), (4.4), (4.5) et (4.6), Carbonez *et al.* (1995) ont construit un estimateur de la fonction de régression en données censurées de type *histogramme*. Nous ne revenons pas

sur cet estimateur mais plutôt sur l'un de ceux proposés par Kohler *et al.* (2002) basé sur des noyaux (dans leur article sont aussi proposés des estimateurs *spline* et des estimateurs de type *moindre carrés pénalisés*, qui pourraient être traités par une méthode similaire à celle exposée dans ce chapitre). Pour un choix de *fenêtre* $h > 0$ approprié, l'expression de l'estimateur considéré ici est :

$$\widehat{m}_{\psi,n,h}^*(\mathbf{x}) := \sum_{i=1}^n \omega_{n,h,i}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{1 - G_n^*(Z_i)}, \quad (4.7)$$

avec la fonction de poids $\omega_{n,h,i}(\mathbf{x})$ définie comme suit,

$$\omega_{n,h,i}(\mathbf{x}) := \begin{cases} \frac{K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h}\right)}, & \text{si } \sum_{j=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h}\right) \neq 0; \\ 1/n, & \text{sinon,} \end{cases} \quad (4.8)$$

K étant un *noyau*, i.e. une fonction mesurable à valeur réelle et d'intégrale un, défini sur \mathbb{R}^d .

Remarquons que le choix $\psi = \psi_t := \mathbb{I}_{(-\infty, t]}$, $t \in \mathbb{R}$, permet de traiter le cas de la fonction de répartition conditionnelle. Nous pouvons définir l'estimateur de cette quantité suivant, pour $\mathbf{x} \in \mathbb{R}^d$,

$$\widehat{F}_{n,h}^*(t|\mathbf{x}) := \sum_{i=1}^n \omega_{n,h,i}(\mathbf{x}) \frac{\delta_i \mathbb{I}_{\{Z_i \leq t\}}}{1 - G_n^*(Z_i)}. \quad (4.9)$$

Remarque 4.2.1. *En l'absence de covariables, l'estimateur de Kaplan-Meier est égal à une transformation simple de l'estimateur de Carbonez (et aussi de Kohler). Cependant, dans le cas qui nous intéresse plus particulièrement dans ce chapitre où certains facteurs influent sur la variable d'intérêt, l'estimateur de Kaplan-Meier local et l'estimateur de Carbonez sont différents (voir Carbonez et al. (1995) pour un exposé complet sur ces différences).*

En vue de l'obtention de nos résultats, nous travaillerons sous l'hypothèse (A) ci-dessous.

(A) $\mathcal{F} = \{\psi := \psi_1 \mathbb{I}_{(-\infty, \tau]}, \psi_1 \in \mathcal{F}_1\}$, où $\tau < T_H$ et \mathcal{F}_1 est une classe de fonctions mesurables, définies sur \mathbb{R} , à valeur dans \mathbb{R} , séparable point par point et de type *VC*.

Une alternative à cette hypothèse est de supposer l'existence d'un réel T tel que $\mathbb{P}(0 \leq Y \leq T) = 1$, $\mathbb{P}(Y = T) > 0$ et $\mathbb{P}(C > T) > 0$ (voir la condition (A.3) dans Kohler *et al.* (2006), et la définition de la quantité $Y_T = Y \wedge T$ dans Brunel et Comte (2006b)).

Dans le §4.3 qui suit, nous présentons nos hypothèses de travail et nos résultats principaux. Dans le §4.4, nous montrons comment ces résultats peuvent conduire à l'obtention de bandes de confiance simultanées presque sûres asymptotiques. Le §4.5 est, quant à lui, consacré aux démonstrations de nos résultats. Enfin, en annexe, nous présentons une justification de l'écriture additive de l'estimateur de Kaplan-Meier, ainsi que la démonstration d'un fait technique, qui sera utilisé pour l'obtention des bandes de confiance.

4.3 Résultats principaux

4.3.1 Notations et hypothèses supplémentaires

Commençons par préciser les hypothèses additionnelles qui seront utilisées par la suite. Ces hypothèses sont essentiellement identiques à celles faites dans le chapitre précédent,

et sont rappelées ici par commodité.

Soit I une boule de \mathbb{R}^d d'intérieur non vide. On supposera qu'il existe un $\alpha > 0$ tel que (\mathbf{X}, Y) ait une densité jointe $f_{\mathbf{X}, Y}$ par rapport à la mesure de Lebesgue sur $I^\alpha \times \mathbb{R}$, avec

$$I^\alpha := \left\{ \mathbf{x} \in \mathbb{R}^d : \inf_{\mathbf{y} \in I} \|\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^d} \leq \alpha \right\}.$$

Dans cette formule, $\|\cdot\|_{\mathbb{R}^d}$ désigne la norme euclidienne classique sur \mathbb{R}^d . En notant $f_{\mathbf{X}}$ la densité marginale de \mathbf{X} sur I^α , on supposera vérifiées les conditions de régularité suivantes.

(F.I) Pour tout $\mathbf{x} \in I^\alpha$, $\lim_{\mathbf{x}' \rightarrow \mathbf{x}; \mathbf{x}' \in I^\alpha} f_{\mathbf{X}, Y}(\mathbf{x}', y) = f_{\mathbf{X}, Y}(\mathbf{x}, y)$ pour presque tout $y \in \mathbb{R}$.

(F.II) $f_{\mathbf{X}}$ est continue et strictement positive sur I^α .

(F.III) \mathcal{F} est bornée (au sens que la classe \mathcal{F} possède une fonction enveloppe Ψ uniformément bornée, et mesurable, telle que $\Psi(y) \geq \sup_{\psi \in \mathcal{F}} \psi(y)$, $y \in \mathbb{R}$).

Introduisons, par ailleurs, la famille de fonctions

$$\mathcal{M} := \{m_\psi / f_{\mathbf{X}}, \psi \in \mathcal{F}\}. \quad (4.10)$$

Dans les théorèmes qui suivent, concernant les estimateurs de la fonction de régression, nous ferons usage de l'hypothèse suivante.

(F.IV) L'ensemble de fonctions \mathcal{M} est relativement compact, pour la topologie engendrée par la norme uniforme sur I^α .

Remarquons que l'hypothèse (F.IV) est notamment vérifiée lorsque $\mathcal{F}_1 = \{\mathbb{1}_{(-\infty, t]}, t \in \mathbb{R}\}$ (voir pp. 6-7 dans Einmahl et Mason (2000)). Cette propriété nous permettra de décrire, sans difficulté, par la suite, comme conséquence de résultats généraux, la convergence uniforme en $t \in \mathbb{R}$ de l'estimateur de la fonction de répartition conditionnelle $F(t|\cdot) = \mathbb{P}(Y \leq t|\cdot)$.

Intéressons-nous maintenant aux conditions qui seront imposées sur le noyau K . Soit $\mathcal{K} := \{K(\lambda(\cdot - \mathbf{z})), \mathbf{z} \in \mathbb{R}^d, \lambda > 0\}$.

(K.I) (i) $\lim_{|\mathbf{u}| \rightarrow 0} \int_{\mathbb{R}^d} (K(\mathbf{x}) - K(\mathbf{x} + \mathbf{u}))^2 d\mathbf{x} = 0$.

(ii) $\lim_{\lambda \rightarrow 1} \int_{\mathbb{R}^d} (K(\lambda \mathbf{x}) - K(\mathbf{x}))^2 d\mathbf{x} = 0$.

($\widetilde{K.II}$) (i) Il existe un $0 < \kappa < \infty$ tel que $K(\mathbf{s}) = 0$, pour $|\mathbf{s}| \geq \frac{\kappa}{2}$.

(ii) Il existe une constante C_K telle que $\sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| \leq C_K$.

(iii) K est positif.

(K.III) (i) $\exists C > 0, v > 0, \forall 0 < \varepsilon < 1, \mathcal{N}(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-v}$.

(ii) \mathcal{K} est séparable point par point.

Ici, et par la suite, on note $|\mathbf{s}| = \max_{1 \leq j \leq d} |s_j|$ pour tout $\mathbf{s} \in \mathbb{R}^d$. De plus,

$$\mathcal{N}(\varepsilon, \mathcal{K}) := \sup \left\{ \mathcal{N}(\varepsilon, \mathcal{K}, L_2(\mathbb{P})), \mathbb{P} \text{ mesure de probabilité} \right\}, \quad (4.11)$$

représente le *nombre de recouvrement uniforme* de \mathcal{K} , relatif à $\varepsilon > 0$, et à la classe de normes $\{L_2(\mathbb{P})\}$. Dans (4.11), \mathbb{P} varie dans l'ensemble des mesures de probabilité sur \mathbb{R}^d . Nous renvoyons à la définition A.1 de l'annexe du présent mémoire, pour une définition précise du *nombre de recouvrement* d'un espace métrique.

Les hypothèses (K.I)-(K.II)-(K.III) imposent que la quantité $\int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t}$ soit finie, mais permettent à celle-ci de prendre la valeur 0. Pour éviter les difficultés inhérentes à ce cas, nous supposons, par la suite, que K satisfait la condition additionnelle

$$(K.IV) \quad \int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1.$$

Dans le présent chapitre, nous supposons que la *fenêtre* h est telle que $h \in [h'_n, h''_n]$. Ici, $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ désignent deux suites de constantes réelles positives vérifiant les conditions (H.I-II-III) ci-dessous. Celles-ci sont exprimées en fonction d'une suite générique, $\{h_n\}_{n \geq 1}$, de réels positifs.

$$(H.I) \quad h_n \downarrow 0, 0 < h_n < 1, \text{ et } nh_n^d \uparrow \infty;$$

$$(H.II) \quad nh_n^d / \log n \rightarrow \infty \text{ lorsque } n \rightarrow \infty;$$

$$(H.III) \quad \log(1/h_n) / \log \log n \rightarrow \infty \text{ lorsque } n \rightarrow \infty.$$

Remarque 4.3.1. Sur les hypothèses concernant la loi de (Y, C, \mathbf{X})

(i) Une des hypothèses nécessaires à la consistance de l'estimateur de Kaplan-Meier local est que $\mathbb{P}[Y \geq t | \mathbf{X} = \mathbf{x}] > 0 \Rightarrow \mathbb{P}[C \geq t | \mathbf{X} = \mathbf{x}] > 0$. Dans le cas de l'estimateur de Kohler et al., l'hypothèse devient $\mathbb{P}[Y \geq t] > 0 \Rightarrow \mathbb{P}[C \geq t] > 0$. Ainsi, l'hypothèse dans le cas de l'estimateur de Kaplan-Meier local est que le processus de censure soit "locally fair", alors que pour l'estimateur de Kohler, il suffit que ce processus soit "globally fair" (ici, on a comme condition $G(\tau) > 0$, ce qui, au vu du fait que $\tau < T_H$, est toujours vérifié; c'est pourquoi l'hypothèse de "global fairness" n'apparaît pas clairement dans notre étude). En ce sens, l'estimateur étudié dans ce chapitre requiert des hypothèses moins fortes que celles imposées par la procédure de Beran (1981).

(ii) D'autre part, il est bon de rappeler que la condition (I) d'indépendance entre (\mathbf{X}, Y) et C , imposée ici, est plus forte que l'hypothèse classique d'indépendance conditionnelle entre Y et C sachant \mathbf{X} . Cependant, ces deux hypothèses coïncident si $\mathbb{P}[C > t | \mathbf{X} = \mathbf{x}]$ ne dépend pas de \mathbf{X} (i.e. si \mathbf{X} et C sont indépendants).

Dans la suite, comme dans Deheuvels et Mason (2004), nous ferons usage d'une fonction auxiliaire $\{\Theta(\mathbf{x}) : \mathbf{x} \in I\}$, supposée continue et strictement positive sur I . Nous supposons également disposer d'un estimateur de cette fonction, noté $\{\Theta_n(\mathbf{x}) : \mathbf{x} \in I\}$, pour $n = 1, 2, \dots$. On suppose que cet estimateur converge vers $\Theta(\cdot)$ sur I , au sens que, presque sûrement,

$$(\Theta.1) \quad \lim_{n \rightarrow \infty} \sup_{\mathbf{x} \in I} \left| \frac{\Theta_n(\mathbf{x})}{\Theta(\mathbf{x})} - 1 \right| = 0.$$

On introduit maintenant la quantité $\tilde{\sigma}_\psi^2(\mathbf{x})$ qui va remplacer la variance conditionnelle $\sigma_\psi^2(\mathbf{x})$ de $\psi(Y)$ sachant que $\mathbf{X} = \mathbf{x}$ ($\sigma_\psi^2(\mathbf{x}) = \text{Var}(\psi(Y) | \mathbf{X} = \mathbf{x})$, voir (3.8) dans le **chapitre 3**). Il est à noter que cette quantité est supérieure à $\sigma_\psi^2(\mathbf{x})$, cette « inflation de variance » étant due au phénomène de censure. Comme G est supposée continue, cette fonction de $\mathbf{x} \in I^\alpha$ est définie, sous les conditions (F.I-II-III) et (A), par

$$\tilde{\sigma}_\psi^2(\mathbf{x}) = \mathbb{E} \left(\frac{\psi^2(Y)}{1 - G(Y)} \mid \mathbf{X} = \mathbf{x} \right) - m_\psi^2(\mathbf{x}). \quad (4.12)$$

Soit une constante positive $h > 0$. On rappelle la définition de $\widehat{\mathbb{E}} m_{\psi;n}$, introduite dans le chapitre précédent.

$$\widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h) := \begin{cases} \frac{\mathbb{E}\left\{\psi(Y)K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\}}{\mathbb{E}\left\{K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\}}, & \text{pour } \mathbb{E} f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \mathbb{E}(\psi(Y)), & \text{pour } \mathbb{E} f_{\mathbf{X};n}(\mathbf{x}; h) = 0, \end{cases} \quad (4.13)$$

avec, comme en (3.11),

$$f_{\mathbf{X};n}(\mathbf{x}; h) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h}\right). \quad (4.14)$$

Enfin, on introduit la version de $\widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h)$ correspondant au cas où $\mathcal{F} = \{\mathbb{I}_{(-\infty, t]} : t < T_H\}$,

$$F_h(t|\mathbf{x}) := \begin{cases} \frac{\mathbb{E}\left\{\mathbb{I}_{\{Y \leq t\}}K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\}}{\mathbb{E}\left\{K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\}}, & \text{pour } \mathbb{E} f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \mathbb{E}(\mathbb{I}_{\{Y \leq t\}}) = F(t), & \text{pour } \mathbb{E} f_{\mathbf{X};n}(\mathbf{x}; h) = 0. \end{cases} \quad (4.15)$$

4.3.2 Résultats

Nous pouvons maintenant présenter nos résultats. Les démonstrations détaillées de ces derniers sont établies au §4.5.

Théorème 4.3.1. *Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III) avec $0 < h'_n \leq h''_n < 1$. Sous les hypothèses (A), (I), (F.I-II-III-IV), (K.I), ($\widetilde{K.II}$), (K.III-IV) et ($\Theta.1$), on a, presque sûrement,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \Theta_n(\mathbf{x}) \left\{ \widehat{m}_{\psi, n, h}^*(\mathbf{x}) - \widehat{\mathbb{E}} m_{\psi; n}(\mathbf{x}; h) \right\}}{\sqrt{2 \log(1/h^d)}} \\ &= \left\{ \int_{\mathbb{R}^d} K^2(t) dt \sup_{\mathbf{x} \in I} \frac{\Theta^2(\mathbf{x}) \sup_{\psi \in \mathcal{F}} \widetilde{\sigma}_{\psi}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right\}^{1/2}. \end{aligned} \quad (4.16)$$

Corollaire 4.3.1. *Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III) avec $0 < h'_n \leq h''_n < 1$. Sous les hypothèses (I), (F.I-II), (K.I), ($\widetilde{K.II}$), (K.III-IV) et ($\Theta.1$), on a, presque sûrement, pour tout $\tau_0 < T_H$,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \Theta_n(\mathbf{x}) \sup_{t \leq \tau_0} \left\{ \widehat{F}_{n, h}^*(t|\mathbf{x}) - F_h(t|\mathbf{x}) \right\}}{\sqrt{2 \log(1/h^d)}} \\ &= \left\{ \int_{\mathbb{R}^d} K^2(t) dt \sup_{\mathbf{x} \in I} \frac{\Theta^2(\mathbf{x}) \sup_{t \leq \tau_0} \widetilde{\sigma}_{\mathbb{I}_{[0, t]}}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right\}^{1/2}. \end{aligned} \quad (4.17)$$

4.4 Applications : construction de bandes de confiance

4.4.1 Résultat nécessaire

Dans ce paragraphe, nous présentons une façon de construire des bandes de confiance asymptotiques simultanées presque sûres pour la fonction de régression définie en (4.2). Pour ce faire, nous ferons usage d'une version du théorème 4.3.1, énoncée dans le théorème 4.4.1 ci-dessous.

Soit une fenêtre H_n , fonction mesurable de \mathbf{x} et $\mathbf{X}_1, \dots, \mathbf{X}_n$, et, de ce fait, aléatoire. Afin d'évaluer le comportement asymptotique de la fenêtre $H_n(\mathbf{x})$, il est nécessaire de supposer qu'elle reste proche, au sens de l'hypothèse (B.1) ci-dessous, d'une suite de constantes positives déterministes $\{h_n : n \geq 1\}$, vérifiant les conditions (H.I-II-III). Soient deux constantes $0 < c_1 \leq c_2 < \infty$.

$$(B.1) \quad c_1 h_n \leq \inf_{\mathbf{x} \in I} H_n(\mathbf{x}) \leq \sup_{\mathbf{x} \in I} H_n(\mathbf{x}) \leq c_2 h_n \quad \text{presque sûrement lorsque } n \rightarrow \infty.$$

Remarquons que l'hypothèse (B.1) ci-dessus est la version « presque sûre » de l'hypothèse (B.1) de Deheuvels et Mason (2004).

En vue de l'énoncé du théorème 4.4.1, on introduit \mathcal{V}_I , la mesure de Lebesgue de la boule I (i.e., le volume de I), et

$$\log_{\theta, K}(u) := \log \left(\theta \vee u \left\{ \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\} \right),$$

où $\theta > 1$ est une constante fixée, dont on donnera un exemple de choix judicieux dans la suite (cf. §4.4.4). Nous pouvons maintenant présenter, dans le théorème 4.4.1 ci-dessous, la version du théorème 4.3.1 dont nous ferons usage pour obtenir des bandes de confiance simultanées pour la régression en présence de censure.

Théorème 4.4.1. *Soit $\{h_n\}_{n \geq 1}$ une suite de constantes positives vérifiant les hypothèses (H.I-II-III). Sous les hypothèses (A), (T), (F.I-II-III-IV), (K.I), ($\widetilde{K.II}$), (K.III-IV), ($\Theta.1$) et (B.1), on a, presque sûrement,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\sqrt{n H_n^d(\mathbf{x})} \pm \Theta_n(\mathbf{x}) \left\{ \widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - \widehat{\mathbf{E}} m_{\psi; n}(\mathbf{x}; H_n(\mathbf{x})) \right\}}{\sqrt{2 \log_{\theta, K}(\mathcal{V}_I / H_n^d(\mathbf{x}))}} \\ &= \left\{ \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t} \sup_{\mathbf{x} \in I} \frac{\Theta^2(\mathbf{x}) \sup_{\psi \in \mathcal{F}} \widetilde{\sigma}_{\psi}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right\}^{1/2}. \end{aligned} \quad (4.18)$$

Le théorème 4.4.1 se déduit du théorème 4.3.1, de la même façon que la première partie du théorème 1.2 dans Deheuvels et Mason (2004) se déduit de leur corollaire 3.2. Les détails de la démonstration sont donc omis.

Remarque 4.4.1. *Un résultat analogue est valable pour la fonction de répartition conditionnelle. Comme précédemment, les hypothèses (F.III-IV) sont alors automatiquement vérifiées.*

4.4.2 Hypothèses nécessaires supplémentaires

Nous présentons maintenant des conditions suffisantes pour que les termes déterministes de type biais soient négligeables devant les termes aléatoires de type variance traités par nos théorèmes. Les résultats de ce paragraphe sont connus (voir Deheuvels et Mason (2004)), nous les rappelons néanmoins pour faciliter la lecture du présent chapitre.

Nous supposons que les fonctions $f_{\mathbf{X}}$ et $f_{\mathbf{X},Y}$ vérifient les conditions supplémentaires de régularité (F.V)(i-ii).

(F.V) (i) $f_{\mathbf{X}}$ est trois fois continûment différentiable sur I^α .

(ii) $f_{\mathbf{X},Y}$ est trois fois continûment différentiable sur $I^\alpha \times \mathbb{R}$.

Par ailleurs, on impose la condition suivante au noyau K .

(K.V) (i) $\int_{\mathbb{R}^d} u_1^{j_1} \dots u_d^{j_d} K(\mathbf{u}) d\mathbf{u} = 0$, $j_1, \dots, j_d \geq 0$, $j_1 + \dots + j_d = 0, \dots, \ell - 1$.

(ii) $\int_{\mathbb{R}^d} |u_1^{j_1} \dots u_d^{j_d}| K(\mathbf{u}) d\mathbf{u} < \infty$, $j_1, \dots, j_d \geq 0$, $j_1 + \dots + j_d = \ell$.

Sous ces conditions additionnelles, on montre, au prix de calculs analytiques détaillés dans l'appendice de ce chapitre (voir (4.51)), que, pour un choix de fenêtre H_n déterministe, si $h_n = An^{-\delta_1}$ avec $0 < A < \infty$ et $1/(4+d) \leq \delta_1 < 1$, alors le terme déterministe, de type biais, devient négligeable devant le terme aléatoire, de type variance, au sens où les expressions (4.21) et (4.22) ci-dessous sont vraies (voir Deheuvels et Mason (2004) et Nadaraya (1989)). Nous rappelons donc que, sous des hypothèses générales de régularité des fonctions $f_{\mathbf{X}}$ et $f_{\mathbf{X},Y}$, de tels choix de fenêtre h_n , et notamment les choix donnés par $h_n = Cn^{-1/(4+d)}$, $0 < C < \infty$, sont optimaux pour la norme \mathbb{L}_2 (i.e., ils minimisent les critères *MISE* ou *MSE*), mais sont sous-optimaux pour la norme \mathbb{L}_∞ . Au regard de cette norme, le compromis (*trade-off*) biais-variance est obtenu, sous ces mêmes hypothèses de régularité, pour des fenêtres du type $C_2(\log n)^{1/2}n^{-1/(4+d)}$, $0 < C_2 < \infty$.

4.4.3 Construction des bandes de confiance

A partir de maintenant, et jusqu'à la fin du présent paragraphe, nous supposons que la fenêtre H_n est déterministe. Le cas, plus général, où H_n est aléatoire n'est pas traité dans ce mémoire. Par ailleurs, on considère une fonction $\psi \in \mathcal{F}$ fixée.

Introduisons l'estimateur suivant de la quantité $\tilde{\sigma}_\psi^2(\mathbf{x})$.

$$\tilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h) = \begin{cases} \sum_{i=1}^n \frac{\delta_i(\psi(Z_i))^2}{(1 - G_n^*(Z_i))^2} \omega_{n,h,i}(\mathbf{x}) - \hat{m}_{\psi,n,h}^{*2}(\mathbf{x}), & \text{pour } f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i(\psi(Z_i))^2}{(1 - G_n^*(Z_i))^2} - \hat{m}_{\psi,n,h}^{*2}(\mathbf{x}), & \text{pour } f_{\mathbf{X};n}(\mathbf{x}; h) = 0. \end{cases} \quad (4.19)$$

Pour construire les bandes de confiance qui suivent, nous nous appuyerons, notamment, sur le résultat du lemme suivant, dont la démonstration est reportée à la fin du présent chapitre.

Lemme 4.4.1. *Supposons que $h_n = An^{-\delta_1}$, avec $0 < A < \infty$ et $1/(4+d) \leq \delta_1 < 1$. Soient $h'_n = c_1 h_n$ et $h''_n = c_2 h_n$, avec $0 < c_1 \leq c_2 < \infty$. Sous les hypothèses (F.I-III-III-V), (K.I), (\bar{K} .II), (K.III-IV-V), on a, presque sûrement lorsque $n \rightarrow \infty$,*

$$\sup_{\mathbf{x} \in I} \sup_{h \in [h'_n, h''_n]} \left| \frac{\tilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h)}{\tilde{\sigma}_\psi^2(\mathbf{x})} - 1 \right| \rightarrow 0. \quad (4.20)$$

Le théorème 4.4.1 et le lemme 4.4.1 permettent d'exhiber des fonctions (aléatoires) $L_n(\mathbf{x}) = L_n(\mathbf{X}_1, \dots, \mathbf{X}_n; \mathbf{x}) > 0$ (voir (4.29) plus bas), telles que

$$\sup_{\mathbf{x} \in I} \pm \left(\frac{1}{L_n(\mathbf{x})} \right) \{ \widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - \widehat{\mathbb{E}} m_{\psi; n}(\mathbf{x}; H_n(\mathbf{x})) \} \xrightarrow{p.s.} 1. \quad (4.21)$$

Par ailleurs, sous les hypothèses du lemme 4.4.1, on a (voir (4.51) dans l'appendice, à la fin de ce chapitre)

$$\sup_{\mathbf{x} \in I} \pm \left(\frac{1}{L_n(\mathbf{x})} \right) \{ \widehat{\mathbb{E}} m_{\psi; n}(\mathbf{x}; H_n(\mathbf{x})) - m_{\psi}(\mathbf{x}) \} \xrightarrow{p.s.} 0. \quad (4.22)$$

En supposant que les expressions (4.21) et (4.22) soient valides, on obtient aisément des expressions de *bandes de confiance asymptotiques simultanées*, au sens où, pour tout ε fixé vérifiant $0 < \varepsilon < 1$, il existe un $n_0 = n_0(\varepsilon)$ tel que pour tout $n \geq n_0$,

$$\begin{aligned} \mathbb{P} \left(m_{\psi}(\mathbf{x}) \in [\widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - (1 + \varepsilon)L_n(\mathbf{x}), \right. \\ \left. \widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) + (1 + \varepsilon)L_n(\mathbf{x})], \forall \mathbf{x} \in I \right) &= 1, \\ \mathbb{P} \left(m_{\psi}(\mathbf{x}) \in [\widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - (1 - \varepsilon)L_n(\mathbf{x}), \right. \\ \left. \widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) + (1 - \varepsilon)L_n(\mathbf{x})], \forall \mathbf{x} \in I \right) &= 0. \end{aligned} \quad (4.23)$$

Lorsque (4.23) sera vérifiée pour tout $0 < \varepsilon < 1$, on dira que l'intervalle

$$[A_n(\mathbf{x}), B_n(\mathbf{x})] = [\widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - L_n(\mathbf{x}), \widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) + L_n(\mathbf{x})] \quad (4.24)$$

est une *bande de confiance simultanée asymptotique optimale* (au niveau de confiance asymptotique de 100%) pour $m_{\psi}(\mathbf{x})$ sur $\mathbf{x} \in I$. Remarquons que cette méthode ne fournit pas de régions de confiance au sens usuel du terme, puisqu'elles ne sont pas rattachées ici à un *niveau de confiance* $1 - \alpha$ spécifié. Il s'agit plutôt d'intervalles de confiance *presque sûrs* puisque renfermant la vraie valeur avec probabilité un lorsque $n \rightarrow \infty$.

Intéressons-nous maintenant aux choix des fonctions Θ_n et L_n . Sous les hypothèses du lemme 4.4.1, et avec une fonction déterministe H_n vérifiant la condition (B.1), on a, d'après le lemme 4.4.1 et le théorème 3.1.2 du chapitre précédent,

$$\sup_{\mathbf{x} \in I} \left| \frac{f_{\mathbf{X}; n}(\mathbf{x}, H_n(\mathbf{x}))}{f_{\mathbf{X}}(\mathbf{x})} - 1 \right| \xrightarrow{p.s.} 0, \quad \sup_{\mathbf{x} \in I} \left| \frac{\widetilde{\sigma}_{\psi; n}^{*2}(\mathbf{x}, H_n(\mathbf{x}))}{\widetilde{\sigma}_{\psi}^2(\mathbf{x})} - 1 \right| \xrightarrow{p.s.} 0. \quad (4.25)$$

Ainsi, les choix de

$$\Theta_n(\mathbf{x}) = \left\{ \frac{f_{\mathbf{X}; n}(\mathbf{x}, H_n(\mathbf{x}))}{\widetilde{\sigma}_{\psi; n}^{*2}(\mathbf{x}, H_n(\mathbf{x}))} \right\}^{1/2} \quad \text{et} \quad \Theta(\mathbf{x}) = \left\{ \frac{f_{\mathbf{X}}(\mathbf{x})}{\widetilde{\sigma}_{\psi}^2(\mathbf{x})} \right\}^2, \quad (4.26)$$

conduisent, sous les hypothèses du théorème 4.4.1, à

$$\sup_{\mathbf{x} \in I} \left\{ \frac{nH_n^d(\mathbf{x})}{2 \log_{\theta, K}(\mathcal{V}_I/H_n^d(\mathbf{x}))} \right\} \pm \{ \widehat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - \widehat{\mathbb{E}} m_{\psi; n}(\mathbf{x}; H_n(\mathbf{x})) \} \xrightarrow{p.s.} \left[\int_{\mathbf{R}^d} K^2(\mathbf{t}) dt \right]^{1/2}. \quad (4.27)$$

De plus, sous les hypothèses du lemme 4.4.1, avec toujours une fonction déterministe H_n vérifiant la condition (B.1)

$$\sup_{\mathbf{x} \in I} \left\{ \frac{nH_n^d(\mathbf{x})}{2 \log_{\theta, K}(\mathcal{V}_I/H_n^d(\mathbf{x}))} \right\} |\widehat{\mathbb{E}} m_{\psi; n}(\mathbf{x}; H_n(\mathbf{x})) - m_{\psi}(\mathbf{x})| \rightarrow 0. \quad (4.28)$$

Nous renvoyons à l'annexe du présent chapitre pour une démonstration de ce résultat (cf. (4.51)).

Finalement, au vu de (4.27) et (4.28), et en posant

$$L_n(\mathbf{x}) = \left\{ \frac{2 \log_{\theta, K}(\mathcal{V}_I / H_n^d(\mathbf{x}))}{n H_n^d(\mathbf{x})} \times \frac{\tilde{\sigma}_{\psi; n}^{*2}(\mathbf{x}, H_n(\mathbf{x}))}{f_{\mathbf{X}; n}(\mathbf{x}, H_n(\mathbf{x}))} \right\}^{1/2} \times \left[\int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right]^{1/2}, \quad (4.29)$$

on conclut que l'intervalle

$$[\hat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) - L_n(\mathbf{x}), \hat{m}_{\psi, n, H_n(\mathbf{x})}^*(\mathbf{x}) + L_n(\mathbf{x})] \quad (4.30)$$

fournit une *bande de confiance simultanée asymptotique optimale*, au sens précisé plus haut.

4.4.4 Motivation de l'introduction de \mathcal{V}_I et $\log_{\theta, K}$ - Choix de θ

Comme nous l'avons déjà évoqué, l'intérêt de ces deux quantités apparaît dans le cas où l'on considère des échantillons de taille n finie. Elles jouent ainsi un rôle important dans la construction des bandes de confiance présentées dans le paragraphe précédent, alors que d'un point de vue purement théorique (e.g. pour l'établissement formel des lois du logarithme telles que celles présentées dans le théorème 4.3.1), on peut se placer à l'asymptotique sans plus de précaution, et se passer de ce type de « subtilité ». Cette partie est largement inspirée des remarques discutées dans l'article de Deheuvels et Mason (2004).

Motivation de l'introduction de \mathcal{V}_I et $\log_{\theta, K}$

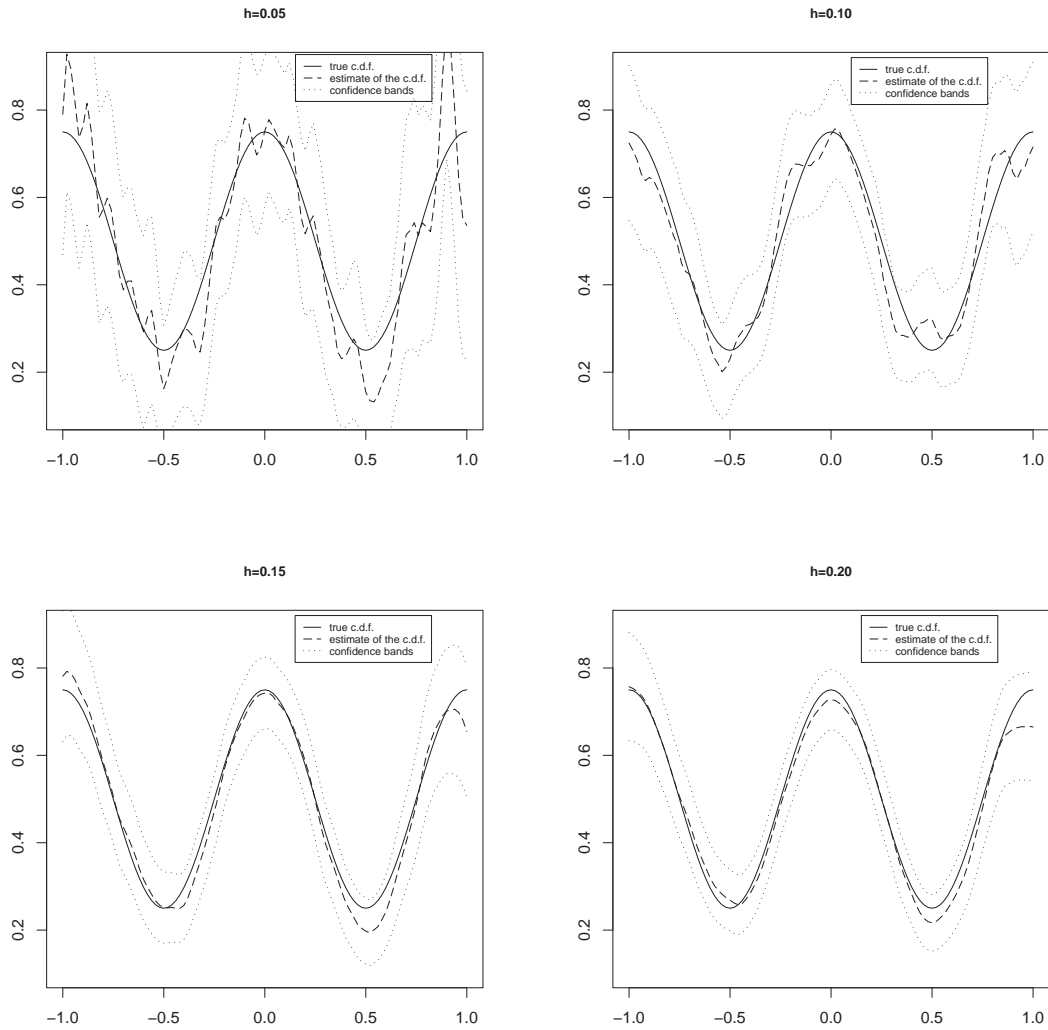
Nous commençons par montrer pourquoi, pour des tailles d'échantillon finie, l'expression $\log_{\theta, K}(\mathcal{V}_I / H_n^d(\mathbf{x}))$ est à préférer à $\log(1 / H_n^d(\mathbf{x}))$ en (4.18). Premièrement, avec $\{h_n\}_{n \geq 1}$ une suite de constantes positives, le remplacement de $K(\mathbf{t})$ et h_n par, formellement, $\tilde{K}(\mathbf{t}) = \lambda K(\lambda \mathbf{t})$ et $\tilde{h}_n = \lambda h_n$, dans les expressions de $f_{\mathbf{X}; n}(\mathbf{x}, h_n)$ et $m_{\psi; n}(\mathbf{x}, h_n)$, ne modifie pas ces estimateurs. Or, on remarque que pour $\lambda \neq 1$, $\log_{\theta, \tilde{K}}(\mathcal{V}_I / \tilde{h}_n^d) = \log_{\theta, K}(\mathcal{V}_I / h_n^d)$, alors que $\log(1 / \tilde{h}_n^d) \neq \log(1 / h_n^d)$. Il paraît donc légitime de préférer $\log_{\theta, K}(\mathcal{V}_I / h_n^d)$, cette expression étant invariante par *changement d'échelle au niveau du noyau*.

D'autre part, un changement d'échelle dans les données, remplaçant l'échantillon $\mathbf{X}_1, \dots, \mathbf{X}_n$ par $\rho \mathbf{X}_1, \dots, \rho \mathbf{X}_n$, $\rho > 0$, transforme logiquement \mathcal{V}_I et h_n^d en $\rho^d \mathcal{V}_I$ et $\rho^d h_n^d$ respectivement. Ceci tend une nouvelle fois à nous faire préférer l'expression (\mathcal{V}_I / h_n^d) à l'expression $1 / h_n^d$ puisque la première est invariante par tout *changement d'échelle au niveau des données*.

Choix de θ

On se place dans le cas où $\mathbb{P}(C \geq T_F) = 1$, i.e., dans le cas non censuré. On rappelle alors qu'un estimateur de la fonction de régression de $\psi(Y)$ sachant que $\mathbf{X} = \mathbf{x}$ est donné dans le chapitre précédent en (3.13). Pour comprendre l'intérêt de l'introduction de la constante θ et pour en choisir une valeur pertinente, on rappelle que sous les hypothèses (F.I-II-III) du chapitre précédent, les conditions $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ impliquent la normalité asymptotique de $m_{I_d; n}(\mathbf{x}, h_n)$ pour tout $\mathbf{x} \in I$, avec I_d qui désigne la fonction identité. Plus précisément, on a (cf. §3.1 dans Härdle (1990))

$$(nh_n^d)^{1/2} \{m_{I_d; n}(\mathbf{x}, h_n) - \hat{\mathbb{E}}m_{I_d; n}(\mathbf{x}, h_n)\} \xrightarrow{d} N\left(0, \frac{\sigma_{I_d}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbf{R}^d} K^2(\mathbf{t}) d\mathbf{t}\right). \quad (4.31)$$

FIG. 4.1 – Estimation de la c. d. f. ($\mathbb{P}(\delta = 1) \simeq 0.55$).

A partir de ce résultat, on peut, en supposant que h_n a été choisi de telle sorte que $|\widehat{\mathbb{E}}m_{I_d;n}(\mathbf{x}, h_n) - m_{I_d}(\mathbf{x})| = o((nh_n^d)^{-1/2})$, construire un intervalle de confiance ponctuel de niveau de confiance asymptotique 95% pour $m_{I_d}(\mathbf{x})$. Concrètement, pour n suffisamment grand,

$$\mathbb{P}\left(m_{I_d}(\mathbf{x}) \in \left[m_{I_d;n}(\mathbf{x}, h_n) \pm (1.96) \left\{ \frac{\sigma_{I_d}^2(\mathbf{x})}{nh_n^d f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2} \right]\right) \simeq 95\%, \quad (4.32)$$

où " \simeq " signifie "approximativement égal". Au vu de (4.29), on remarque que le choix $\theta = 7$ assure que $\{2 \log_{\theta, K}(u)\}^{1/2}$ est toujours supérieure à $\{2 \log 7\}^{1/2} \simeq 1.97. > 1.96$. Ainsi, le choix $\theta = 7$ assure que l'intervalle de confiance « ponctuel » au niveau 95%, (4.32), est inclus dans notre bande de confiance simultanée (4.30). Pour plus de détails, nous renvoyons une nouvelle fois à Deheuvels et Mason (2004).

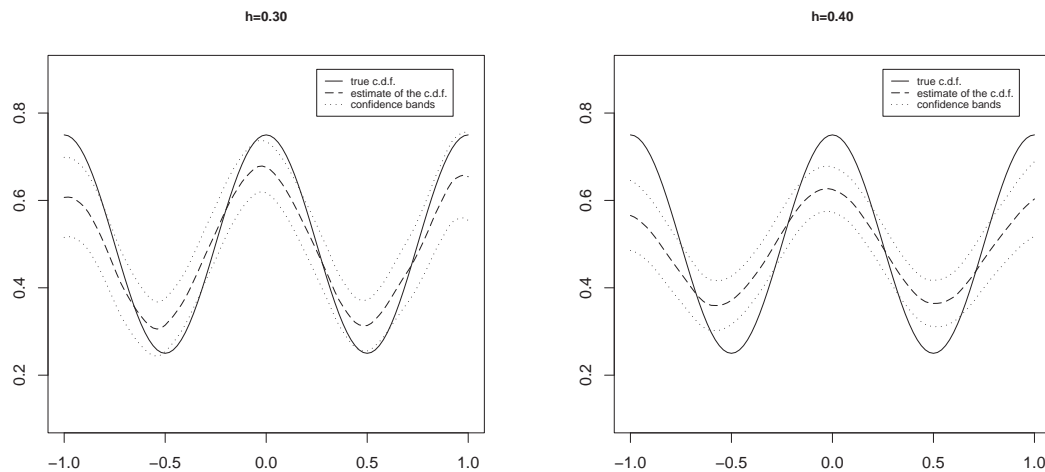


FIG. 4.2 – Estimation de la c. d. f. ($\mathbb{P}(\delta = 1) \simeq 0.55$).

4.4.5 Illustration

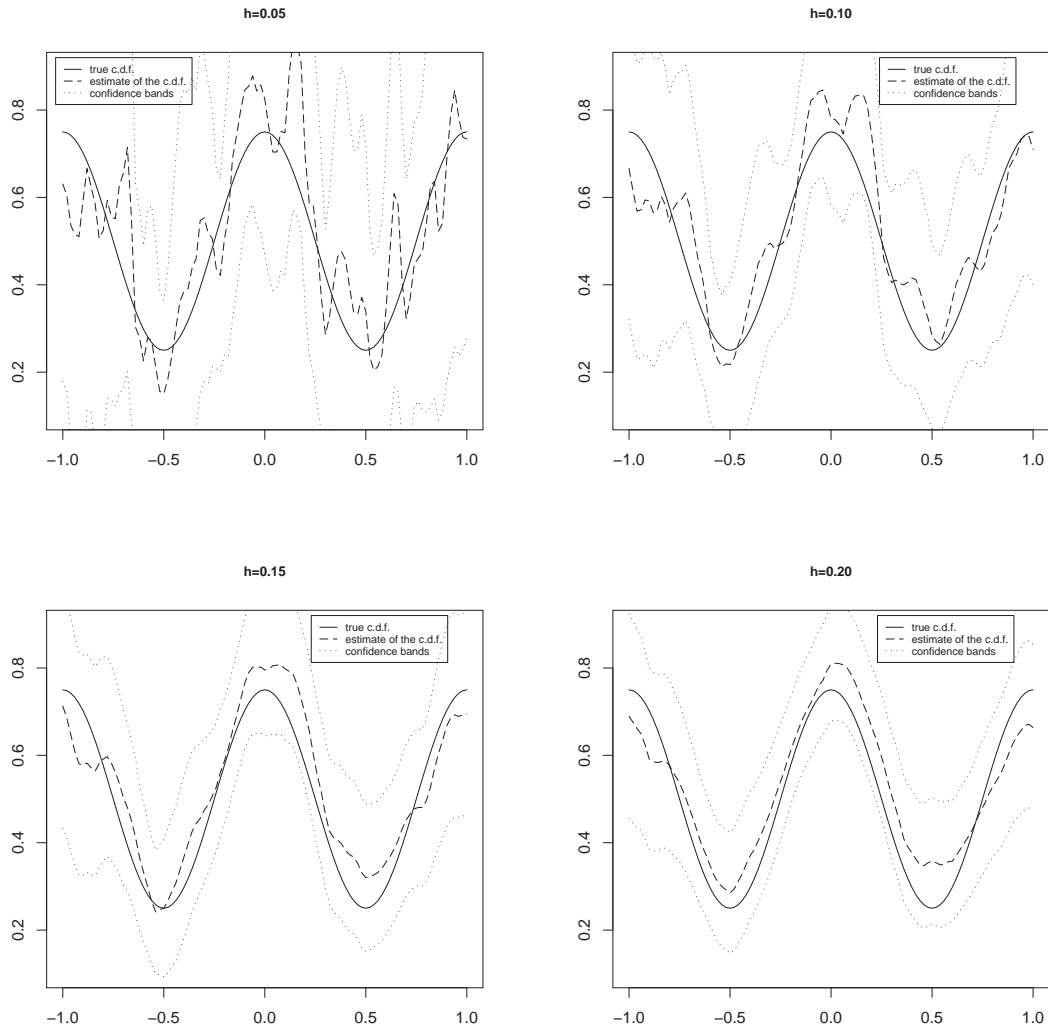
Dans cette partie, nous présentons, pour illustrer ce qui précède, quelques simulations de bandes de confiance simultanées, dont la forme est donnée en 4.29)–(4.30). Nous nous plaçons dans le cas où $\mathbf{X} = X \in \mathbb{R}$ (i.e. $d = 1$) est tel que $X \sim \mathcal{N}(0, 1)$. Nous considérons le choix $\psi = \mathbb{1}_{\{Y \leq \tau\}}$, et nous fixons $\tau = 0.9$. Soit alors le modèle

$$\mathbb{E}(\mathbb{1}_{\{Y \leq \tau\}} \mid X = x) = 0.25 + 0.5 * \cos^2(x).$$

La variable Y est simulée de la façon suivante. A chaque entier $1 \leq i \leq n$, nous faisons correspondre une probabilité $p_i = 0.25 + 0.5 * \cos^2(x_i)$, où x_i est la valeur observée de la variable X_i . Notons que $0 < p_i < 1$, pour tout entier $1 \leq i \leq n$. La quantité Y_i est ensuite tirée comme réalisation d’une variable aléatoire de loi $\mathcal{U}(\tau - p_i, 1 + \tau - p_i)$, uniforme sur $(\tau - p_i, 1 + \tau - p_i)$, (de telle sorte que $\mathbb{P}(Y_i \leq \tau \mid X_i = x_i) = p_i$, et donc que $\mathbb{1}_{\{Y_i \leq \tau\}} \mid X_i = x_i$ suive une loi de Bernoulli, de paramètre p_i).

Afin de mieux rendre compte de l’influence de la censure sur la procédure d’estimation, nous présentons des résultats obtenus pour différentes lois de censure. Le nombre d’observations total est, quant à lui, fixé à $n = 2000$. Pour chacune des simulations, nous fournissons la vraie valeur de la ”conditional distribution function” (c.d.f.) à estimer (en trait continu), les estimations obtenues (en tirets), ainsi que les bandes de confiance simultanées correspondantes (en pointillés), et ce, sur l’intervalle $[-1, 1]$. Les fenêtres utilisées $H_n(x) = h$ sont déterministes et ne dépendent pas de la position x . De plus, le noyau sélectionné est le noyau d’Epanechnikov, $K(u) = 0.75 * (1 - u^2) * \mathbb{1}_{\{|u| \leq 1\}}$, $u \in \mathbb{R}$.

Dans un premier temps, nous considérons le cas où la variable de censure est telle que $C \sim \mathcal{U}(0, 2)$. On observe alors que $\mathbb{P}(\delta = 1) \simeq 0.55$. Les graphiques de la Figure 4.1 représentent les estimations obtenues pour des fenêtres $h = 0.05$, $h = 0.10$, $h = 0.15$ et $h = 0.20$. Pour tous ces choix de fenêtre, les bandes de confiance simultanées associées renferment la vraie valeur de la c.d.f. en tout point de l’intervalle $[-1, 1]$, ce qui corrobore nos résultats théoriques. Dans la Figure 4.2, nous présentons les résultats des simulations pour $h = 0.3$ et $h = 0.4$. Les bandes de confiance ne contiennent alors plus la vraie valeur de la fonction en tout point. La raison est que le terme de type biais devient trop important

FIG. 4.3 – Estimation de la c. d. f. ($\mathbb{P}(\delta = 1) \simeq 0.20$).

pour des valeurs aussi élevées de h : les estimateurs, comme les bandes de confiance, ne sont plus satisfaisants. Ceci met en exergue l'importance du choix de h en pratique - à distance finie en fait. Nos résultats sont certes valables uniformément en $h \in [h'_n, h''_n]$, mais seulement pour des séquences $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ vérifiant $h'_n = An^{-\delta_1}$ et $h''_n = Bn^{-\delta_2}$, avec $0 < A, B < \infty$ et $1/(4+d) \leq \delta_2 \leq \delta_1 < 1$ tels que $0 < h'_n \leq h''_n < 1$. On ne peut donc évidemment pas choisir h « au hasard ».

Nous avons également effectué des simulations dans le cas où $C \sim \mathcal{U}(0, 1)$, ce qui correspond, *a posteriori* à $\mathbb{P}(\delta = 1) \simeq 0.20$. Les graphiques de la Figure 4.3 représentent les estimations obtenues pour des fenêtres $h = 0.05$, $h = 0.10$, $h = 0.15$ et $h = 0.20$. Là encore, les bandes de confiance simultanées renferment la vraie valeur de la fonction en tout point de $[-1, 1]$. Pour des valeurs de h plus grandes, nous sommes bien sûr confrontés au même type de problème que précédemment (simulations non fournies).

En conclusion, ces simulations, aussi limitées qu'elles soient, mettent en évidence que nos bandes de confiance simultanées sont correctes, et peuvent être utilisées en pratique. Elles

soulignent aussi le fait que le choix du paramètre de lissage reste crucial. Des procédures de sélection adaptatives existent (voir par exemple le package *lokern* du logiciel R), et pourraient être également utilisées dans le cas de la régression censurée.

4.5 Démonstrations

Nous présentons maintenant les démonstrations de nos principaux résultats.

4.5.1 Démonstration du théorème 4.3.1

Introduisons la quantité suivante,

$$\widehat{m}_{\psi,n,h}^{\star(1)}(\mathbf{x}) := \sum_{i=1}^n \omega_{n,h,i}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{1 - G(Z_i)}, \quad (4.33)$$

qui correspond à la version de $\widehat{m}_{\psi,n,h}^{\star}$ obtenue lorsque la fonction G est connue.

Le principe de la démonstration du théorème 4.3.1 va être d'obtenir son analogue pour $\widehat{m}_{\psi,n,h}^{\star(1)}$ (proposition 4.5.1 ci-dessous), puis de montrer que la différence entre $\widehat{m}_{\psi,n,h}^{\star(1)}$ et $\widehat{m}_{\psi,n,h}^{\star}$ est négligeable sous les conditions du théorème 4.3.1 (lemme 4.5.1 ci-dessous).

Un résultat utile

Proposition 4.5.1. *Soit $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les hypothèses (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Sous les hypothèses (\mathcal{A}), (\mathcal{I}), (F.I-II-III-IV), (K.I), ($\widetilde{K.II}$), (K.III-IV) et ($\Theta.1$), on a, presque sûrement,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} \frac{\sqrt{nh^d} \pm \Theta_n(\mathbf{x}) \left\{ \widehat{m}_{\psi,n,h}^{\star(1)} - \widehat{\mathbb{E}} m_{\psi;n}(\mathbf{x}; h) \right\}}{\sqrt{2 \log(1/h^d)}} \\ &= \sup_{\psi \in \mathcal{F}} \left\{ \int_{\mathbb{R}^d} K^2(t) dt \sup_{\mathbf{x} \in I} \frac{\Theta^2(\mathbf{x}) \widetilde{\sigma}_{\psi}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right\}^{1/2}. \end{aligned} \quad (4.34)$$

Démonstration. Ce résultat est une conséquence directe du théorème 3.1.1 du chapitre précédent. En effet, en rappelant la définition de la fonction Ψ_{ψ} , de \mathbb{R}^2 dans \mathbb{R} , telle que pour tout $(y, c) \in \mathbb{R}^2$,

$$\Psi_{\psi}(y, c) = \frac{\mathbb{1}_{\{y \leq c\}} \psi(y \wedge c)}{1 - G(y \wedge c)},$$

il est évident que Ψ_{ψ} est uniformément bornée, en $(y, c) \in \mathbb{R}^2$ et $\psi \in \mathcal{F}$, puisque \mathcal{F} est uniformément bornée, $\psi(t) = 0$ pour tout $t > \tau$ et $G(\tau) < 1$. Cette propriété, combinée au fait que la classe \mathcal{F}_1 est supposée de type VC assure que la classe de fonctions $\mathcal{G} := \{\Psi_{\psi}, \psi \in \mathcal{F}\}$ vérifie la condition $[\mathcal{E}]$ du **chapitre 3**. De même, la classe \mathcal{G} est clairement séparable point par point. Par ailleurs, le fait que la classe $\mathcal{M}' = \{\mathbb{E}(\Psi_{\psi}(Y, C) | \mathbf{X} = \cdot), \psi \in \mathcal{F}\}$ soit presque sûrement relativement compacte par rapport à la topologie uniforme sur I^{α} est quant à lui assuré par (F.IV) et le fait que $\mathbb{E}(\Psi_{\psi}(Y, C) | \mathbf{X} = \mathbf{x}) = m_{\psi}(\mathbf{x})$. On peut donc appliquer le théorème 3.1.1 avec le changement formel $\mathbf{Z} \rightsquigarrow (Y, C)^T$ et $\psi \rightsquigarrow \Psi_{\psi}$.

Comme, sous l'hypothèse (\mathcal{I}) , on a

$$\begin{aligned} \mathbb{E}\left\{\Psi_{\psi}(Y, C)K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right)\right\} &= \mathbb{E}\left\{\frac{\mathbb{1}_{\{Y \leq C\}}\psi(Z)}{1 - G(Z)}K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right)\right\} \\ &= \mathbb{E}\left\{\frac{\psi(Y)K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right)}{1 - G(Y)}\mathbb{E}[\mathbb{1}_{\{Y \leq C\}}|\mathbf{X}, Y]\right\} \\ &= \mathbb{E}\left\{\psi(Y)K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right)\right\}, \end{aligned} \quad (4.35)$$

le fait que ce soit le terme $\widehat{\mathbb{E}}m_{\psi;n}(\mathbf{x}; h)$ qui intervienne dans l'énoncé du résultat de la proposition 4.5.1 est établi. Enfin, pour $j = 1, 2$, sous (\mathcal{I}) , on a

$$\begin{aligned} \mathbb{E}\left\{\left(\frac{\delta\psi(Z)}{1 - G(Z)}\right)^j|\mathbf{X}\right\} &= \mathbb{E}\left\{\left(\frac{\psi(Y)}{1 - G(Y)}\right)^j\mathbb{E}[\mathbb{1}_{\{Y \leq C\}}|\mathbf{X}, Y]|\mathbf{X}\right\} \\ &= \mathbb{E}\left\{\frac{\psi^j(Y)}{(1 - G(Y))^{j-1}}|\mathbf{X}\right\}, \end{aligned}$$

ce qui justifie l'expression de $\widetilde{\sigma}_{\psi}^2$. \square

Un lemme d'approximation

Pour achever la démonstration du théorème 4.3.1, nous allons utiliser le résultat du lemme d'approximation suivant.

Lemme 4.5.1. *Sous les hypothèses du théorème 4.3.1, on a, presque sûrement lorsque $n \rightarrow \infty$,*

$$\sup_{h \in [h'_n, h''_n]} \sup_{\psi \in \mathcal{F}} \sup_{\mathbf{x} \in I} |\widehat{m}_{\psi, n, h}^{*(1)}(\mathbf{x}) - \widehat{m}_{\psi, n, h}^*(\mathbf{x})| = O\left(\sqrt{\frac{\log_2 n}{n}}\right). \quad (4.36)$$

Démonstration. On a

$$\begin{aligned} &\sup_{\psi \in \mathcal{F}} \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} |\widehat{m}_{\psi, n, h}^{*(1)}(\mathbf{x}) - \widehat{m}_{\psi, n, h}^*(\mathbf{x})| \\ &= \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} \left| \sum_{i=1}^n \omega_{n, h, i}(\mathbf{x}) \delta_i \psi(Z_i) \left(\frac{1}{1 - G(Z_i)} - \frac{1}{1 - G_n^*(Z_i)} \right) \right| \\ &\leq \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} \sum_{i=1}^n |\omega_{n, h, i}(\mathbf{x})| \sup_{t \leq \tau} \frac{\psi(t)}{G_n^*(t)G(t)} \times \sup_{t \leq \tau} |G_n^*(t) - G(t)|. \end{aligned}$$

On conclut en rappelant que $\sup_{\psi \in \mathcal{F}} \|\psi\| < \infty$, que le noyau K est supposé positif, que $\tau < T_H = T_F \leq T_G$, et en utilisant le fait que, d'après la loi du logarithme itéré portant sur G_n^* (voir Földes et Rejtő (1981) par exemple),

$$\sup_{t \leq \tau} |G_n^*(t) - G(t)| = O\left(\sqrt{\frac{\log_2 n}{n}}\right),$$

presque sûrement lorsque $n \rightarrow \infty$. \square

Bilan

En combinant les résultats de la proposition 4.5.1 et du lemme 4.5.1, le résultat du théorème 4.3.1 est immédiat en remarquant que, sous les conditions (H.I-III-III), on a, uniformément en $h \in [h'_n, h''_n]$, pour n suffisamment grand,

$$\frac{\log(1/h)}{nh} \gg \frac{\log_2 n}{n}.$$

4.5.2 Démonstration du lemme 4.4.1

La démonstration du lemme 4.4.1 s'obtient par les mêmes arguments techniques que la démonstration du théorème 4.3.1. Notons cependant que, au vu du théorème 4.3.1, et sous les hypothèses du lemme 4.4.1 (qui assurent que les termes déterministes de type biais sont négligeables), on a, presque sûrement lorsque $n \rightarrow \infty$,

$$\sup_{\psi \in \mathcal{F}} \sup_{h \in [h'_n, h''_n]} \sup_{\mathbf{x} \in I} |\widehat{m}_{\psi,n,h}^*(\mathbf{x}) - m_\psi(\mathbf{x})| \rightarrow 0.$$

On peut alors se contenter de l'étude des quantités suivantes. On introduit, pour toute fonction mesurable $\psi \in \mathcal{F}$, et, pour tout $\mathbf{x} \in \mathbb{R}$,

$$\widetilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h) := \sum_{i=1}^n \frac{\delta_i(\psi(Z_i))^2}{(1 - G(Z_i))^2} \omega_{n,h,i}(\mathbf{x}), \quad (4.37)$$

et

$$\widetilde{\sigma}_\psi^2(\mathbf{x}) := \widetilde{\sigma}_\psi^2(\mathbf{x}) + m_\psi^2(\mathbf{x}). \quad (4.38)$$

Un résultat utile

Commençons par établir le résultat suivant.

Proposition 4.5.2. *Sous les hypothèses du théorème 4.3.1, on a, presque sûrement,*

$$\sup_{\mathbf{x} \in I} \sup_{h \in [h'_n, h''_n]} \left| \frac{\widetilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h)}{\widetilde{\sigma}_\psi^2(\mathbf{x})} - 1 \right| \rightarrow 0. \quad (4.39)$$

Démonstration. En suivant, pas à pas, la démonstration de la proposition 4.5.1, avec le choix de $\Psi_\psi(y, c)$ donné par

$$\Psi_\psi(y, c) = \frac{\mathbb{I}_{\{y \leq c\}} \psi^2(y \wedge c)}{(1 - G(y \wedge c))^2},$$

on aboutit directement au résultat suivant. On a, presque sûrement lorsque $n \rightarrow \infty$,

$$\sup_{\mathbf{x} \in I} \sup_{h \in [h'_n, h''_n]} \left| \frac{\widetilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h)}{\widehat{\mathbb{E}} \widetilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h)} - 1 \right| \rightarrow 0, \quad (4.40)$$

avec, pour tout $\mathbf{x} \in I$ et tout $h \in [h'_n, h''_n]$,

$$\widehat{\mathbb{E}} \widetilde{\sigma}_{\psi;n}^{*2}(\mathbf{x}; h) = \frac{\mathbb{E} \left\{ \frac{\psi^2(Y)}{1 - G(Y)} K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right) \right\}}{\mathbb{E} \left\{ K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right) \right\}}. \quad (4.41)$$

La démonstration de la proposition 4.5.2 se déduit notamment du fait que, sous les hypothèses du lemme 4.4.1, les termes déterministes, de type biais, sont négligeables par rapport aux termes aléatoires (voir (4.51) en appendice de ce chapitre). \square

Pour aboutir à la conclusion du lemme 4.4.1, il suffit de procéder comme dans la démonstration du lemme 4.5.1, en faisant usage de la loi du logarithme itéré de Földes et Rejtő (1981).

4.6 Appendice

4.6.1 Justification de l'écriture additive de l'estimateur de Kaplan-Meier

Nous montrons, dans ce qui suit, comment justifier l'écriture additive de l'estimateur de Kaplan-Meier en l'absence de covariables (écriture qui, rappelons-le, est à la base de l'estimateur de Carbonez). Notons tout d'abord que, sous l'hypothèse d'indépendance entre Y et C , on a la décomposition suivante pour $H(y)$, $y \in \mathbb{R}$,

$$H(y) = 1 - (1 - F(y))(1 - G(y)). \quad (4.42)$$

Pour toute fonction L , on pose $L_-(x) = \lim_{\varepsilon \downarrow 0} L(x - \varepsilon)$, pour tout $x \in \mathbb{R}$, dès lors que cette limite existe. Introduisons les quantités suivantes.

$$H^{(1)}(y) := \mathbb{P}(Z \leq y, \delta = 1) = \int_0^y (1 - G_-(t)) dF(t) = H_+^{(1)}(y), \quad (4.43)$$

et

$$H^{(0)}(y) := \mathbb{P}(Z \leq y, \delta = 0) = \int_0^y (1 - F(t)) dG(t) = H_+^{(0)}(y), \quad (4.44)$$

qui vérifient l'identité $H(y) = H^{(1)}(y) + H^{(0)}(y)$.

D'après (4.44), on a $dH^{(0)}(t) = (1 - F_-(t))dG(t)$, ou encore,

$$\begin{aligned} \frac{dH_0(t)}{(1 - F_-(t))(1 - G_-(t))} &= \frac{dG(t)}{1 - G_-(t)} \\ \text{et donc} \quad \frac{dH^{(0)}(t)}{1 - H_-(t)} &= \frac{dG(t)}{1 - G_-(t)}. \end{aligned} \quad (4.45)$$

De la même manière, au vu de (4.43), il vient

$$\frac{dH^{(1)}(t)}{1 - H_-(t)} = \frac{dF(t)}{1 - F_-(t)}. \quad (4.46)$$

Au vu des relations (2.27) et (2.28) présentées au **chapitre 2**, il est immédiat que les estimateurs F_n^* et G_n^* , de F et G respectivement, vérifient les contreparties empiriques de ces équations différentielles, soit

$$\begin{aligned} dF_n^*(t) &= (1 - F_{n-}^*(t)) \frac{dH_n^{(1)}(t)}{1 - H_{n-}(t)}, \\ dG_n^*(t) &= (1 - F_{n-}^*(t)) \frac{dH_n^{(0)}(t)}{1 - H_{n-}(t)}. \end{aligned} \quad (4.47)$$

avec

$$\begin{aligned} H_n^{(1)}(t) &= \frac{1}{n} \sum_{i=1}^n \delta_i \mathbb{I}_{\{Z_i \leq t\}}, \\ H_n^{(0)}(t) &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \mathbb{I}_{\{Z_i \leq t\}}, \\ H_n(t) &= H_n^{(1)}(t) + H_n^{(0)}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{Z_i \leq t\}}. \end{aligned} \quad (4.48)$$

Remarque 4.6.1. Dans Deheuvels et Derzko (2006) et Deheuvels et Mason (2006), le même type de propriété est établi pour les versions conditionnelles des estimateurs de Kaplan-Meier. L'estimateur de Kaplan-Meier « classique », i.e. non conditionnel, étant, en quelque sorte, un cas particulier de l'estimateur de Kaplan-Meier conditionnel, (4.47) se déduit donc également des résultats de ces auteurs.

D'après les équations différentielles empiriques (4.47), les sauts de F_n^* ont lieu aux points de discontinuité de $H_n^{(1)}(t)$, i.e., sur les observations Z_i correspondant à des $\delta_i = 1$. D'autre part, les sauts de G_n^* ont lieu aux points Z_i correspondant à des $\delta_i = 0$. On peut donc écrire :

$$\begin{aligned} F_n^*(t) &= \frac{1}{n} \sum_{i=1}^n \delta_i \mathbb{I}_{\{Z_i \leq t\}} \pi_{F,n,i} \\ G_n^*(t) &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \mathbb{I}_{\{Z_i \leq t\}} \pi_{G,n,i}, \end{aligned} \quad (4.49)$$

où les termes $\pi_{F,n,i}$ et $\pi_{G,n,i}$ s'interprètent comme des *poids*. L'observation selon laquelle $\pi_{F,n,i} = 1/[n(1 - G_n^*(Z_i))]$, est due à Susarla *et al.* (1984) (une démonstration complète de ce résultat figure dans Satten et Datta (2001)).

4.6.2 Etude du biais des estimateurs à noyau de la régression

Soient $\{h'_n\}_{n \geq 1}$ et $\{h''_n\}_{n \geq 1}$ deux suites de constantes positives vérifiant les conditions (H.I-II-III), avec $0 < h'_n \leq h''_n < \infty$. Considérons une fonction $\Phi(\cdot)$, mesurable et telle que $m_\Phi(\mathbf{x}) = \mathbb{E}(\Phi(Y) \mid \mathbf{X} = \mathbf{x})$ existe, pour tout $\mathbf{x} \in I^\alpha$. Les termes, de type biais, invoqués dans ce chapitre sont de la forme

$$\widehat{\mathbb{E}}m_{\Phi;n}(\mathbf{x}, h) - m_\Phi(\mathbf{x}), \quad (4.50)$$

où, pour tout $\mathbf{x} \in I$ et tout $h \in [h'_n, h''_n]$, les expressions de $m_{\Phi;n}(\mathbf{x}, h)$ et $\widehat{\mathbb{E}}m_{\Phi;n}(\mathbf{x}, h)$ sont respectivement données en (3.6) du **chapitre 3** et (4.13).

Soit, pour $\mathbf{t} \in \mathbb{R}^d$, $r(\mathbf{t}) = \int_{\mathbb{R}} \Phi(y) f_{\mathbf{X},Y}(\mathbf{t}, y) dy$. Sous les hypothèses du lemme 4.4.1, on a, pour tout $h \in [h'_n, h''_n]$ et tout $\mathbf{x} \in I$,

$$\begin{aligned} \mathbb{E} \left[\Phi(Y) K \left(\frac{\mathbf{x} - \mathbf{X}}{h} \right) \right] &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} \Phi(y) K \left(\frac{\mathbf{x} - \mathbf{t}}{h} \right) f_{\mathbf{X},Y}(\mathbf{t}, y) dt dy \\ &= \frac{h^{2+d}}{2} \sum_{i,j} \frac{\partial^2 r(\mathbf{x})}{\partial x_i \partial x_j} \int_{\mathbb{R}^d} u_i u_j K(\mathbf{u}) d\mathbf{u} + r(\mathbf{x}) + \mathcal{O}(h^{d+3}). \end{aligned}$$

De plus, sous ces mêmes hypothèses, on a, pour tout $h \in [h'_n, h''_n]$ et tout $\mathbf{x} \in I$,

$$m_\Phi(\mathbf{x})\mathbb{E}\left[K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right)\right] - r(\mathbf{x}) = \frac{h^{2+d}}{2} m_\Phi(\mathbf{x}) \sum_{i,j} \frac{\partial^2 f_{\mathbf{X}}(\mathbf{x})}{\partial x_i \partial x_j} \int_{\mathbf{R}^d} u_i u_j K(\mathbf{u}) d\mathbf{u} + \mathcal{O}(h^{d+3}).$$

Comme, enfin,

$$\mathbb{E}\left[K\left(\frac{\mathbf{x} - \mathbf{X}}{h}\right)\right] = h^d f_{\mathbf{X}}(\mathbf{x}) + \mathcal{O}(h^{d+2}),$$

on en déduit facilement que, sous les hypothèses du lemme 4.4.1, il existe une constante $0 < C_\Phi < \infty$ telle que, uniformément en $h \in [h'_n, h''_n]$ et en $\mathbf{x} \in I$, on ait, à partir d'un certain rang n ,

$$h^{-2} \{\widehat{\mathbb{E}}m_{\Phi;n}(\mathbf{x}, h) - m_\Phi(\mathbf{x})\} = C_\Phi + o(1). \quad (4.51)$$

Ce résultat est suffisant pour montrer que, sous les hypothèses du lemme 4.4.1, les *termes déterministes*, de type biais (selon la définition ci-dessus), sont négligeables par rapport aux *termes aléatoires*, de type variance, ces derniers ayant été décrits par les théorèmes précédents.

Le résultat (4.51) est notamment suffisant pour établir (4.28).

Chapitre 5

Convergence uniforme d'un estimateur de la régression additive en données censurées.

Avant-propos. Ce chapitre développe le contenu d'une note soumise aux *Comptes Rendus de l'Académie des Sciences* (Debbbarh et Viallon (2006c)).

Résumé. Dans ce chapitre, nous nous proposons d'établir la vitesse de convergence presque sûre optimale de l'estimateur d'une fonction de régression additive en données censurées. Pour construire nos estimateurs, nous utilisons la méthode d'intégration marginale associée à un estimateur de type *IPCW*. Ce dernier a été introduit au **chapitre 4** du présent mémoire.

5.1 Introduction

Les données censurées interviennent dans de nombreux domaines d'application de la statistique, notamment en épidémiologie, où les variables d'intérêt rencontrées (par exemple, l'instant futur où un patient en observation développe une pathologie) sont, le plus souvent, liées à de multiples facteurs concomitants pouvant être observés à l'instant initial. Considérons, par exemple, la construction de scores de risque pour le cancer du sein. Il s'agit, dans ce cas, d'évaluer la probabilité qu'une femme donnée a de développer cette pathologie dans un délai spécifié, conditionnellement à un certain nombre de facteurs de risque observés au départ (typiquement entre 5 et 15). L'estimation des distributions de survie conditionnelles relève alors d'un besoin logique pour pouvoir contrôler l'évolution de la maladie. Dans ce même contexte, l'utilisation d'estimateurs non paramétriques se heurte au problème pratique, bien connu, du *fléau de la dimension*. En effet, les vitesses de convergence des estimateurs, habituellement utilisés pour évaluer la distribution de survie conditionnelle, dépendent fortement, en général, de la dimension des covariables (dans notre exemple, celle-ci est comprise entre 5 et 15). On consultera à ce sujet Dabrowska (1989), Deheuvels et Derzko (2006), et le Corollaire 4.3.1 du **chapitre 4** de cette thèse.

Afin d'apporter des réponses les plus satisfaisantes possibles au problème du *fléau de la dimension*, nous nous placerons, dans ce qui suit, sous les hypothèses d'un *modèle additif* (on consultera à ce sujet Stone (1985), Hastie et Tibshirani (1990), et les références bibliographiques de ces textes). Notre résultat principal est le suivant. Nous considérons l'estimateur de la fonction de régression multivariée défini au (5.16). Ce dernier estimateur est construit en combinant les estimateurs de la fonction de régression censurée multivariée

introduits par Carbonez *et al.* (1995) et Kohler *et al.* (2002) (voir le **chapitre 4**), avec la méthode d'intégration marginale (voir, à ce sujet, Newey (1994), Tjøstheim et Auestad (1994) et Linton et Nielsen (1995)). Dans ce cadre général, nous décrivons la vitesse de convergence uniforme presque sûre de notre estimateur. Cette vitesse de convergence est identique à celle obtenue par Camlong (1999) dans le cas non censuré.

La suite du présent chapitre est ordonnée de la manière suivante. Après avoir introduit les notations nécessaires, nous présentons, dans le §5.2, un estimateur de la fonction de régression, adapté aux données censurées, dans le cadre du modèle additif. Dans le §5.3, nous exposons les hypothèses qui sont, ensuite, utilisées pour énoncer nos principaux résultats. Le §5.4, qui conclut le chapitre, est consacré aux détails de leur démonstration.

5.2 Notations

Considérons le triplet aléatoire (Y, C, \mathbf{X}) , à valeurs dans $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. Ici, $d \geq 2$ est un entier fixé, Y est la variable d'intérêt (typiquement, une durée de vie), C désigne une variable de censure, et $\mathbf{X} = (X_1, \dots, X_d)$ désigne une variable concomitante, de conditionnement. Nous supposons, dans toute la suite, l'existence de densités, $f_{\mathbf{X}, Y}$ et f , relativement à la mesure de Lebesgue, respectivement sur $\mathbb{R}^d \times \mathbb{R}$ et \mathbb{R}^d , pour le couple (\mathbf{X}, Y) et la variable \mathbf{X} . Nous travaillerons, par la suite, sur un échantillon aléatoire de taille $n \geq 1$, $(Y_i, C_i, \mathbf{X}_i)_{1 \leq i \leq n}$, de triplets indépendants et de même loi que (Y, C, \mathbf{X}) . Désignons par $\mathbb{1}_E$ la fonction indicatrice de E . Dans le cas de *censures à droite*, on observe les variables $Z_i = \min\{Y_i, C_i\}$, $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$ et \mathbf{X}_i , de sorte que, seul l'échantillon aléatoire $(Z_i, \delta_i, \mathbf{X}_i)_{1 \leq i \leq n}$ est à notre disposition.

Dans toute la suite de ce chapitre, nous poserons $F(t) = P(Y \leq t)$, $G(t) = P(C \leq t)$, $\bar{F}(t) = 1 - F(t) = P(Y > t)$, $\bar{G}(t) = 1 - G(t) = P(C > t)$, et nous placerons dans un cadre similaire à celui du **chapitre 4**. Nous introduisons les points extrêmes des distributions de Y et C , définis, respectivement, par $T_F = \sup\{t : \bar{F}(t) > 0\}$ et $T_G = \sup\{t : \bar{G}(t) > 0\}$. Comme dans le **chapitre 4**, nous introduisons une fonction ψ , réelle de variable réelle, mesurable et vérifiant l'hypothèse **(A)** ci-dessous. On se donne une constante $\tau < T_F \wedge T_G$.

(A) $\psi(y) = 0$ si $y \in (\tau, +\infty)$.

Nous nous intéressons à la *fonction de régression multivariée* $m_\psi(\mathbf{x})$ de $\psi(Y)$, évaluée en $\mathbf{X} = \mathbf{x} \in \mathbb{R}^d$, dans le cas où cette fonction est additive. Nous supposons que cette régression existe, et est définie, pour tout $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, par

$$m_\psi(\mathbf{x}) = \mathbb{E}(\psi(Y) \mid \mathbf{X} = \mathbf{x}). \quad (5.1)$$

Nous travaillerons sous le *modèle additif*, qui suppose que m_ψ est de la forme

$$m_\psi(\mathbf{x}) = \mu + \sum_{\ell=1}^d m_\ell(x_\ell). \quad (5.2)$$

Remarque 5.2.1. *Dans le modèle (5.2), les fonctions m_ℓ sont des fonctions réelles de variable réelle, définies à une constante additive près. Cette propriété impose naturellement que les procédures d'analyse statistique soient développées sous la condition d'identifiabilité*

$$\mathbb{E} m_\ell(X_\ell) = 0 \quad \text{pour } \ell = 1, \dots, d. \quad (5.3)$$

Cette dernière condition implique que $\mu = \mathbb{E}(\psi(Z))$.

En vue de la construction d'estimateurs de $m_\psi(\mathbf{x})$, nous introduisons des suites de constantes réelles positives $\{h_n\}_{n \geq 1}$ et $\{h_{j,n}\}_{n \geq 1}$, $j = 1, 2$. Nous introduisons également un noyau $K(\cdot)$, fonction réelle mesurable, définie sur \mathbb{R}^d , et telle que

$$\int_{\mathbb{R}^d} K(\mathbf{x}) d\mathbf{x} = 1.$$

L'estimateur à noyau, \hat{f}_n de f , correspondant à $K(\cdot)$ et $\{h_n\}_{n \geq 1}$, est alors défini, pour tout $\mathbf{x} \in \mathbb{R}^d$ et $n \geq 1$, par

$$\hat{f}_n(\mathbf{x}) = \frac{1}{nh_n^d} \sum_{j=1}^n K\left(\frac{\mathbf{X}_j - \mathbf{x}}{h_n}\right).$$

Dans la suite de l'exposé, \overline{G}_n^* désigne l'estimateur de Kaplan-Meier (Kaplan et Meier (1958)) de \overline{G} , défini, pour tout $y \in \mathbb{R}$, par,

$$\overline{G}_n^*(y) = \prod_{1 \leq i \leq n} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)} \right)^{\beta_i}, \quad (5.4)$$

où $\beta_i = \mathbb{1}_{\{Z_i \leq y\}}(1 - \delta_i)$, $N_n(x) = \sum_{i=1}^n \mathbb{1}_{\{Z_i \geq x\}}$ et avec les conventions $\prod_{\emptyset} = 1$ et $0^0 = 1$.

Introduisons, de plus, des noyaux auxiliaires, K_1 , K_2 et K_3 , définis, respectivement, sur \mathbb{R} , \mathbb{R}^{d-1} et \mathbb{R}^d , et, chacun, d'intégrale égale à 1 sur son domaine de définition (des hypothèses additionnelles sur K_1 , K_2 et K_3 seront faites au §5.3). Posons, pour tout $\mathbf{x} = (x_1, \dots, x_d)$, et pour tout $\ell = 1, \dots, d$,

$$\mathbf{x}_{-\ell} = (x_1, \dots, x_{\ell-1}, x_{\ell+1}, \dots, x_d) \in \mathbb{R}^{d-1}.$$

Pour estimer la régression multivariée $m_\psi(\mathbf{x})$, fonction de $\mathbf{x} \in \mathbb{R}^d$ définie par (5.1), nous ferons usage des estimateurs $\tilde{m}_{\psi,n}^*(\mathbf{x})$ et, pour $\ell = 1, \dots, d$, $\tilde{m}_{\psi,n,\ell}^*(\mathbf{x})$, définis respectivement en (5.7) et (5.5) ci-dessous. Ces estimateurs sont dans l'esprit des travaux de Carbonez *et al.* (1995), Kohler *et al.* (2002) et Jones *et al.* (1994). Tout d'abord, on pose, pour $\ell = 1, \dots, d$,

$$\tilde{m}_{\psi,n,\ell}^*(\mathbf{x}) := \sum_{i=1}^n W_{n,i}^\ell(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{\overline{G}_n^*(Z_i)}, \quad (5.5)$$

où

$$W_{n,i}^\ell(\mathbf{x}) = \frac{K_1\left(\frac{x_\ell - X_{i,\ell}}{h_{1,n}}\right) K_2\left(\frac{\mathbf{x}_{-\ell} - \mathbf{X}_{i,-\ell}}{h_{2,n}}\right)}{nh_{1,n} h_{2,n}^{d-1} \hat{f}_n(\mathbf{X}_i)}. \quad (5.6)$$

On considérera également la quantité $\tilde{m}_{\psi,n}^*$, définie par

$$\tilde{m}_{\psi,n}^*(\mathbf{x}) := \sum_{i=1}^n W_{n,i}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{\overline{G}_n^*(Z_i)}, \quad (5.7)$$

où, pour $i = 1, \dots, n$,

$$W_{n,i}(\mathbf{x}) = \frac{K_3\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}}\right)}{nh_{1,n}^d \hat{f}_n(\mathbf{X}_i)}. \quad (5.8)$$

Remarque 5.2.2. Introduisons la fonction à valeurs réelles, $\Psi_\psi(y, c)$, définie pour $(y, c) \in \mathbb{R}^2$ par

$$\Psi(y, c) = \frac{\mathbb{I}_{\{y \leq c\}} \psi(y \wedge c)}{\bar{G}(y \wedge c)}. \quad (5.9)$$

Lorsque la distribution de censure, \bar{G} , est connue, les estimateurs $\tilde{m}_{\psi,n}^*$ et $\tilde{m}_{\psi,n,\ell}^*$, $\ell = 1, \dots, d$, de la fonction de régression $m_\psi(\mathbf{x})$, définis, respectivement, par les relations (5.5) et (5.7), peuvent être exprimés sous la forme d'estimateurs de la régression multivariée généralisée, comme suit. On a

$$\tilde{m}_{\psi,n}^*(\mathbf{x}) = \sum_{i=1}^n W_{n,i}(\mathbf{x}) \Psi(Y_i, C_i), \quad (5.10)$$

et, pour $\ell = 1, \dots, d$,

$$\tilde{m}_{\psi,n,\ell}^*(\mathbf{x}) = \sum_{i=1}^n W_{n,i}^\ell(\mathbf{x}) \Psi(Y_i, C_i), \quad \ell = 1, \dots, d. \quad (5.11)$$

Pour estimer les composantes additives de m_ψ , nous utilisons la méthode dite d'*intégration marginale*. Nous rappelons ici brièvement son principe, et renvoyons au §0.3.1 du chapitre introductif pour plus de détails. Etant données q_1, \dots, q_d , d densités réelles bornées, nous posons

$$q(\mathbf{x}) = \prod_{\ell=1}^d q_\ell(x_\ell) \quad \text{et} \quad q_{-\ell}(\mathbf{x}_{-\ell}) = \prod_{j \neq \ell} q_j(x_j).$$

Définissons alors

$$\eta_\ell(x_\ell) = \int_{\mathbb{R}^{d-1}} m_\psi(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} m_\psi(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \quad (5.12)$$

Les fonctions η_ℓ , $\ell = 1, \dots, d$, vérifient les relations

$$\eta_\ell(x_\ell) = m_\ell(x_\ell) - \int_{\mathbf{R}} m_\ell(z_\ell) q_\ell(z_\ell) dz_\ell, \quad (5.13)$$

et

$$m_\psi(\mathbf{x}) = \sum_{\ell=1}^d \eta_\ell(x_\ell) + \int_{\mathbb{R}^d} m_\psi(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}. \quad (5.14)$$

Au vu de (5.13) et (5.14), pour chaque $\ell = 1, \dots, d$, les fonctions η_ℓ et m_ℓ sont égales à une constante additive près. Les fonctions η_ℓ , $\ell = 1, \dots, d$, constituent donc des composantes additives de m_ψ , vérifiant une hypothèse d'identifiabilité alternative à celle qui nous a permis de spécifier, pour $\ell = 1, \dots, d$, les m_ℓ à partir de m_ψ .

Remarque 5.2.3. Pour $\ell = 1, \dots, d$, l'égalité entre les fonctions η_ℓ et m_ℓ est obtenue pour le choix de $q_\ell = f_\ell$, où f_ℓ représente la densité associée à la variable X_ℓ (voir Sperlich et al. (2002)). Cependant, f_ℓ n'est généralement pas connue en pratique, et, sauf exception, $\eta_\ell \neq m_\ell$.

Considérant les expressions (5.7) et (5.12), pour $\ell = 1, \dots, d$, un estimateur naturel de la ℓ -ième composante additive η_ℓ de m_ψ est donné par

$$\hat{\eta}_\ell^*(x_\ell) = \int_{\mathbb{R}^{d-1}} \tilde{m}_{\psi,n,\ell}^*(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \tilde{m}_{\psi,n,\ell}^*(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad (5.15)$$

à partir duquel nous déduisons l'estimateur $\widehat{m}_{\psi,add}^*(\mathbf{x})$ de la fonction de régression additive $m_{\psi}(\mathbf{x})$, sous l'hypothèse (5.2), défini par

$$\widehat{m}_{\psi,add}^*(\mathbf{x}) = \sum_{\ell=1}^d \widehat{\eta}_{\ell}^*(x_{\ell}) + \int_{\mathbb{R}^d} \widetilde{m}_{\psi,n}^*(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \quad (5.16)$$

5.3 Présentation des hypothèses et résultats

Introduisons tout d'abord les hypothèses sur le triplet (Y, C, \mathbf{X}) qui seront utilisées pour l'obtention de nos résultats. Rappelons la définition (5.1)–(5.2) de m_{ψ} sous le modèle additif. En sus des hypothèses de base faites dans le §5.2, nous supposons que, pour un entier $k \geq 1$ convenable, et pour τ tel qu'en (A),

(I) C et (\mathbf{X}, Y) sont indépendants ;

(C.2) $\overline{G} = 1 - G$ est continue sur \mathbb{R} ;

(C.3) il existe une constante M telle que, $\sup_{t \leq \tau} |\psi(t)| \leq M < \infty$;

(C.4) m_{ψ} est k -fois continûment différentiable sur \mathbb{R}^d ,

$$\text{et } \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{\partial^k}{\partial x_{\ell}^k} m_{\psi}(\mathbf{x}) \right| < \infty; \ell = 1, \dots, d.$$

Soient $\mathcal{C}_1, \dots, \mathcal{C}_d$, d parties compactes de \mathbb{R} , et soit $\mathcal{C} := \mathcal{C}_1 \times \dots \times \mathcal{C}_d$ le compact produit de \mathbb{R}^d correspondant. Pour toute partie \mathcal{E} de \mathbb{R}^q , pour $q \geq 1$ quelconque, et pour tout $\alpha > 0$, définissons l' α -voisinage euclidien \mathcal{E}^{α} de \mathcal{E} , par

$$\mathcal{E}^{\alpha} = \left\{ \mathbf{x} : \inf_{\mathbf{y} \in \mathcal{E}} \|\mathbf{x} - \mathbf{y}\|_{\mathbb{R}^q} \leq \alpha \right\},$$

où $\|\cdot\|_{\mathbb{R}^q}$ désigne la norme euclidienne usuelle sur \mathbb{R}^q .

Nous supposons que les conditions suivantes sont satisfaites par les densités f et f_{ℓ} de \mathbf{X} et X_{ℓ} , respectivement, pour $\ell = 1, \dots, d$. On suppose, d'une part, que ces fonctions sont continues sur leurs domaines de définition. D'autre part, nous supposons l'existence d'une constante $\alpha > 0$ telle que les hypothèses (F.A-B) ci-dessous soient vérifiées.

(F.A) $\forall x_{\ell} \in \mathcal{C}_{\ell}^{\alpha}, f_{\ell}(x_{\ell}) > 0, \ell = 1, \dots, d$, et $\forall \mathbf{x} \in \mathcal{C}^{\alpha}, f(\mathbf{x}) > 0$.

(F.B) f est k' -fois continûment différentiable sur \mathcal{C}^{α} pour une constante entière $k' > dk$.

Nous supposons également que les fonctions de densité $q_{\ell}, \ell = 1, \dots, d$, sont à support inclus dans \mathcal{C}_{ℓ} .

Les noyaux K, K_1, K_2 et K_3 , utilisés dans le §5.2, et définis, respectivement, sur $\mathbb{R}^d, \mathbb{R}, \mathbb{R}^{d-1}$ et \mathbb{R}^d , seront supposés à supports compacts, et continus d'intégrales égales à 1 sur leurs domaines de définition respectifs. Nous travaillerons, de plus, sous les conditions suivantes.

(K.C) K_1 est lipschitzien.

(K.D) Les noyaux K_1 et K_3 sont d'ordre k , et le noyau K est d'ordre k' .

Nous travaillons enfin avec des constantes de lissage $h_n > 0$ et $h_{j,n} > 0, j = 1, 2$, pour $n = 1, 2, \dots$, vérifiant les conditions (H.A-B) ci-dessous.

(H.A) $h_n = a_1 \left(\frac{\log n}{n} \right)^{1/(2k'+d)}$ pour un $0 < a_1 < \infty$.

(H.B) $h_{1,n} = a_2 \left(\frac{\log n}{n} \right)^{1/(2k+1)}$ pour un $0 < a_2 < \infty$, et $h_{2,n} = o(1)$.

Nous sommes, maintenant, en mesure d'énoncer le résultat principal du présent chapitre, dans le théorème ci-dessous.

Théorème 5.3.1. *On suppose que les conditions (A), (I), (C.2-4), (F.A-B), (K.C-D) et (H.A-B) sont satisfaites. Alors, nous avons, presque sûrement lorsque $n \rightarrow \infty$,*

$$\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}^*(\mathbf{x}) - m_{\psi}(\mathbf{x})| = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{k/(2k+1)}\right). \quad (5.17)$$

5.4 Démonstration du théorème 5.3.1

5.4.1 Introduction

La démonstration de ce théorème repose sur une série de lemmes. Dans le premier de ceux-ci, le lemme 5.4.1, nous établissons une version simplifiée du théorème 5.3.1, correspondant au cas où la densité f des covariables, et la fonction de survie \overline{G} de la censure, sont connues. Les lemmes qui suivent ce premier résultat permettent d'évaluer l'écart entre l'estimateur obtenu dans ce cas particulier (défini en (5.19)) et l'estimateur correspondant au cas général, défini en (5.16).

5.4.2 Cas où la densité f des covariables et la fonction \overline{G} sont connues

Dans le cas où f et \overline{G} sont connues, les estimateurs des composantes additives $\eta_{\ell}(x_{\ell})$, $\ell = 1, \dots, d$, de la régression, et de la fonction de régression $m_{\psi}(\mathbf{x})$ elle-même, compte tenu de (5.13)–(5.14), sont, respectivement, définis par

$$\widehat{\eta}_{\ell}(x_{\ell}) = \int_{\mathbf{R}^{d-1}} \widetilde{m}_{\psi, n, \ell}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbf{R}^d} \widetilde{m}_{\psi, n, \ell}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad (5.18)$$

et

$$\widehat{m}_{\psi, add}(\mathbf{x}) = \sum_{\ell=1}^d \widehat{\eta}_{\ell}(x_{\ell}) + \int_{\mathbf{R}^d} \widetilde{m}_{\psi, n}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad (5.19)$$

en faisant usage des notations

$$\widetilde{m}_{\psi, n, \ell}(\mathbf{x}) = \sum_{i=1}^n \widetilde{W}_{n, i}^{\ell}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{\overline{G}(Z_i)}, \quad (5.20)$$

où, pour $\ell = 1, \dots, d$,

$$\widetilde{W}_{n, i}^{\ell}(\mathbf{x}) = \frac{K_1\left(\frac{x_{\ell} - X_{i, \ell}}{h_{1, n}}\right) K_2\left(\frac{\mathbf{x}_{-\ell} - \mathbf{X}_{i, -\ell}}{h_{2, n}}\right)}{nh_{1, n} h_{2, n}^{d-1} f(\mathbf{X}_i)},$$

et

$$\widetilde{m}_{\psi, n}(\mathbf{x}) = \sum_{i=1}^n \widetilde{W}_{n, i}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{\overline{G}(Z_i)} \quad \text{où} \quad \widetilde{W}_{n, i}(\mathbf{x}) = \frac{K_3\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1, n}}\right)}{nh_{1, n}^d f(\mathbf{X}_i)}. \quad (5.21)$$

Introduisons la fonction auxiliaire $\Psi(y, c)$ de $(y, c) \in \mathbb{R}^2$, définie par

$$\Psi(y, c) = \frac{\mathbb{I}_{\{y \leq c\}} \psi(y \wedge c)}{1 - G(y \wedge c)}.$$

Notons que, sous la conditions (\mathcal{I}) , cette fonction vérifie, pour $\mathbf{x} \in \mathbb{R}^d$, l'identité

$$\begin{aligned} \mathbb{E}\{\Psi(Y, C) | \mathbf{X} = \mathbf{x}\} &= \mathbb{E}\left\{\left(\frac{\psi(Y)}{1 - G(Y)}\right) \mathbb{E}[\mathbb{1}_{\{Y \leq C\}} | \mathbf{X}, Y] \mid \mathbf{X} = \mathbf{x}\right\} \\ &= m_\psi(\mathbf{x}). \end{aligned} \quad (5.22)$$

Nous pouvons énoncer maintenant le premier de nos lemmes composant la démonstration du théorème 5.3.1.

Lemme 5.4.1. *Sous les hypothèses du théorème 5.3.1, on a, presque sûrement lorsque $n \rightarrow \infty$,*

$$\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}(\mathbf{x}) - m_\psi(\mathbf{x})| = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{k/(2k+1)}\right).$$

Démonstration. Soit $\varepsilon_n > 0$, $n = 1, 2, \dots$, une suite de constantes que nous précisons plus loin (voir (5.44)). Comme \mathcal{C} est une partie compacte de \mathbb{R}^d , pour tout $n \geq 1$, on peut recouvrir \mathcal{C} par un nombre fini $r = r(n)$ de boules B_1, \dots, B_r , de centres respectifs $\mathbf{t}_1, \dots, \mathbf{t}_r$, de sorte que

$$\mathcal{C} \subset \bigcup_{p=1}^{r(n)} B_p. \quad (5.23)$$

Remarquons qu'en introduisant la norme

$$|\mathbf{x}| = \sum_{\ell=1}^d |x_\ell|,$$

pour tout $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, il vient

$$|\mathbf{x} - t(\mathbf{x})| \leq \frac{V_{\mathcal{C}}}{r(n)}, \quad (5.24)$$

où la quantité $V_{\mathcal{C}}$ ne dépend que du volume de \mathcal{C} . Le nombre $r = r(n)$ de boules nécessaires en fonction de ε_n sera précisé par la suite en (5.41).

Pour tout $\mathbf{x} \in \mathcal{C}$, il existe un indice $1 \leq p \leq r(n)$ tel que la boule B_p de centre $t_p = t(\mathbf{x})$ contienne \mathbf{x} . L'inégalité du triangle permet alors d'écrire

$$\begin{aligned} &\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}(\mathbf{x}) - m_\psi(\mathbf{x})| \\ &\leq \sup_{\mathbf{x} \in \mathcal{C}} |\mathbb{E}\widehat{m}_{\psi, add}(\mathbf{x}) - m_\psi(\mathbf{x})| + \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}(\mathbf{x}) - \widehat{m}_{\psi, add}(t(\mathbf{x}))| \\ &\quad + \sup_{\mathbf{x} \in \mathcal{C}} |\mathbb{E}\widehat{m}_{\psi, add}(\mathbf{x}) - \mathbb{E}\widehat{m}_{\psi, add}(t(\mathbf{x}))| + \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}(t(\mathbf{x})) - \mathbb{E}\widehat{m}_{\psi, add}(t(\mathbf{x}))|. \end{aligned}$$

Compte tenu de cette inégalité, la démonstration du lemme 5.4.1 peut se ramener à l'évaluation successive des termes apparaissant dans son membre de droite. Il est ainsi suffisant d'établir les quatre étapes suivantes, pour un choix convenable de la suite $\{\varepsilon_n\}_{n \in \mathbb{N}}$

qui sera précisé ultérieurement en (5.44).

$$\sup_{\mathbf{x} \in \mathcal{C}} |\mathbb{E} \widehat{m}_{\psi, add}(\mathbf{x}) - m_{\psi}(\mathbf{x})| = \mathcal{O}(h_{1,n}^k), \quad (5.25)$$

$$\mathbb{P}(\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}(\mathbf{x}) - \widehat{m}_{\psi, add}(t(\mathbf{x}))| \geq \varepsilon_n/3) = 0, \quad (5.26)$$

$$\mathbb{P}(\sup_{\mathbf{x} \in \mathcal{C}} |\mathbb{E} \widehat{m}_{\psi, add}(\mathbf{x}) - \mathbb{E} \widehat{m}_{\psi, add}(t(\mathbf{x}))| \geq \varepsilon_n/3) = 0, \quad (5.27)$$

$$\sum_n \mathbb{P}(\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi, add}(t(\mathbf{x})) - \mathbb{E} \widehat{m}_{\psi, add}(t(\mathbf{x}))| \geq \varepsilon_n/3) < \infty. \quad (5.28)$$

Étape 1 : Démonstration de (5.25). D'après les définitions des quantités η_{ℓ} , $\widehat{\eta}_{\ell}$ et $\widehat{m}_{\psi, add}$, données, respectivement, en (5.12), (5.18) et (5.19), on obtient la décomposition

$$\begin{aligned} \text{Biais}(\widehat{m}_{\psi, add}(\mathbf{x})) &:= \mathbb{E} \widehat{m}_{\psi, add}(\mathbf{x}) - m_{\psi}(\mathbf{x}) \\ &= \sum_{\ell=1}^d (\mathbb{E} \widehat{\eta}_{\ell}(x_{\ell}) - \eta_{\ell}(x_{\ell})) + \mathbb{E} \int_{\mathbf{R}^d} \widetilde{m}_{\psi, n}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x} - \int_{\mathbf{R}^d} m_{\psi}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Par le théorème de Fubini, on peut réécrire ceci sous la forme

$$\text{Biais}(\widehat{m}_{\psi, add}(\mathbf{x})) = \sum_{\ell=1}^d \text{Biais}(\widehat{\eta}_{\ell}(x_{\ell})) + \int_{\mathbf{R}^d} \text{Biais}(\widetilde{m}_{\psi, n}(\mathbf{x})) q(\mathbf{x}) d\mathbf{x}, \quad (5.29)$$

où, pour $\ell = 1, \dots, d$,

$$\text{Biais}(\widehat{\eta}_{\ell}(x_{\ell})) := \mathbb{E}(\widehat{\eta}_{\ell}(x_{\ell})) - \eta_{\ell}(x_{\ell}), \quad (5.30)$$

et

$$\text{Biais}(\widetilde{m}_{\psi, n}(\mathbf{x})) := \mathbb{E} \widetilde{m}_{\psi, n}(\mathbf{x}) - m_{\psi}(\mathbf{x}). \quad (5.31)$$

Les termes de biais définis ci-dessus sont maintenant évalués, successivement, comme suit.

Étude du terme $\text{Biais}(\widehat{\eta}_{\ell}(x_{\ell}))$. Nous nous limiterons à l'étude du cas où $\ell = 1$, les autres valeurs de l'indice ℓ pouvant être traitées de façon similaire. Soit $\hat{\alpha}_1(x_1)$, défini par

$$\begin{aligned} \hat{\alpha}_1(x_1) &:= \int_{\mathbf{R}^{d-1}} \widetilde{m}_{\psi, n, 1}(\mathbf{x}) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1} \\ &= \frac{1}{nh_{1,n}} \sum_{i=1}^n \frac{\widetilde{\Psi}_n(Y_i, C_i)}{f_1(X_{i,1})} K_1\left(\frac{x_1 - X_{i,1}}{h_{1,n}}\right), \end{aligned}$$

où

$$\widetilde{\Psi}_n(Y_i, C_i) = \Psi(Y_i, C_i) \int_{\mathbf{R}^{d-1}} \frac{1}{h_{2,n}^{d-1}} K_2\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{i,-1}}{h_{2,n}}\right) \frac{q_{-1}(\mathbf{x}_{-1})}{f(\mathbf{X}_{i,-1} | X_{i,1})} d\mathbf{x}_{-1}.$$

Introduisons les quantités suivantes. On pose

$$\mathcal{G}(\mathbf{u}_{-1}) := \int_{\mathbf{R}^{d-1}} \frac{1}{h_{2,n}^{d-1}} K_2\left(\frac{\mathbf{x}_{-1} - \mathbf{u}_{-1}}{h_{2,n}}\right) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1},$$

$$g(x_1) := \mathbb{E}\left(\widetilde{\Psi}_n(Y_i, C_i) \mid X_{i,1} = x_1\right),$$

$$C_n := \mu + \int_{\mathbf{R}^{d-1}} \sum_{j=2}^d m_j(u_j) \mathcal{G}(\mathbf{u}_{-1}) d\mathbf{u}_{-1},$$

$$\hat{C}_n := \int_{\mathbf{R}^d} \widetilde{m}_{\psi, n, 1}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x},$$

$$C := \int_{\mathbf{R}} m_1(x_1) q_1(x_1) dx_1.$$

Le terme Biais($\widehat{\eta}_1(x_1)$) peut alors être décomposé comme suit.

$$\begin{aligned} \text{Biais}(\widehat{\eta}_1(x_1)) &= \mathbf{E}(\widehat{\eta}_1(x_1)) - \eta_1(x_1) \\ &= \{\mathbf{E}(\widehat{\alpha}_1(x_1)) - g(x_1)\} + \mathbf{E}(\widehat{C}_n - C_n - C) \\ &=: (I) + (II). \end{aligned} \tag{5.32}$$

Commençons par traiter le terme (I) de (5.32). On a

$$\begin{aligned} g(x_1) &= \mathbf{E}\left(\widetilde{\Psi}_n(Y_i, C_i) \middle| X_{i,1} = x_1\right) \\ &= \mathbf{E}\left\{\Psi(Y_i, C_i) \int_{\mathbf{R}^{d-1}} \frac{1}{h_{2,n}^{d-1}} K_2\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{i,-1}}{h_{2,n}}\right) \frac{q_{-1}(\mathbf{x}_{-1})}{f(\mathbf{X}_{i,-1}|X_{i,1})} d\mathbf{x}_{-1} \middle| X_{i,1} = x_1\right\}. \end{aligned}$$

En utilisant des arguments classiques de conditionnement, et en faisant usage de l'hypothèse \mathcal{I} d'indépendance entre (Y, \mathbf{X}) et C (cf. (5.22)), on obtient que

$$\begin{aligned} g(x_1) &= \mathbf{E}\left\{\mathbf{E}(\Psi(Y_i, C_i)|\mathbf{X}_i) \frac{\mathcal{G}(\mathbf{X}_{i,-1})}{f(\mathbf{X}_{i,-1}|X_{i,1})} \middle| X_{i,1} = x_1\right\} \\ &= \mathbf{E}\left\{\mathbf{E}(\psi(Y_i)|\mathbf{X}_i) \frac{\mathcal{G}(\mathbf{X}_{i,-1})}{f(\mathbf{X}_{i,-1}|X_{i,1})} \middle| X_{i,1} = x_1\right\} \\ &= \int_{\mathbf{R}^{d-1}} m_\psi(x_1, \mathbf{u}_{-1}) \mathcal{G}(\mathbf{u}_{-1}) d\mathbf{u}_{-1}. \end{aligned}$$

Notons maintenant, d'une part, que

$$\begin{aligned} \mathbf{E}(\widehat{\alpha}_1(x_1)) - g(x_1) &= \int_{\mathbf{R}} \frac{1}{h_{1,n}} g(u_1) K_1\left(\frac{x_1 - u_1}{h_{1,n}}\right) du_1 - g(x_1) \\ &= \int_{\mathbf{R}} [g(x_1 - v_1 h_{1,n}) - g(x_1)] K_1(v_1) dv_1. \end{aligned}$$

D'autre part, sous l'hypothèse (C.4), il existe un réel $0 < \theta < 1$ tel que

$$\begin{aligned} &g(x_1 - v_1 h_{1,n}) - g(x_1) \\ &= \int_{\mathbf{R}^{d-1}} [m_\psi(x_1 - v_1 h_{1,n}, \mathbf{u}_{-1}) - m_\psi(x_1, \mathbf{u}_{-1})] \mathcal{G}(\mathbf{u}_{-1}) d\mathbf{u}_{-1} \\ &= \int_{\mathbf{R}^{d-1}} \left[\sum_{i=1}^{k-1} \frac{(-h_{1,n} v_1)^i}{i!} \frac{\partial^i m_\psi}{\partial x_1^i}(x_1, \mathbf{u}_{-1}) \right. \\ &\quad \left. + \frac{(-h_{1,n} v_1)^k}{k!} \frac{\partial^k m_\psi}{\partial x_1^k}(x_1 - \theta h_{1,n} v_1, \mathbf{u}_{-1}) \right] \mathcal{G}(\mathbf{u}_{-1}) d\mathbf{u}_{-1}. \end{aligned}$$

On déduit de ces deux derniers résultats que, sous l'hypothèse (K.D),

$$\begin{aligned} &\int_{\mathbf{R}} [g(x_1 - v_1 h_{1,n}) - g(x_1)] K_1(v_1) dv_1 \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}^{d-1}} \left[\sum_{i=1}^{k-1} \frac{(-h_{1,n} v_1)^i}{i!} \frac{\partial^i m_\psi}{\partial x_1^i}(x_1, \mathbf{u}_{-1}) \right. \\ &\quad \left. + \frac{(-h_{1,n} v_1)^k}{k!} \frac{\partial^k m_\psi}{\partial x_1^k}(x_1 - \theta h_{1,n} v_1, \mathbf{u}_{-1}) \right] \times \mathcal{G}(\mathbf{u}_{-1}) d\mathbf{u}_{-1} K_1(v_1) dv_1 \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}^{d-1}} \left[\frac{(-h_{1,n} v_1)^k}{k!} \frac{\partial^k m_\psi}{\partial x_1^k}(x_1 - \theta h_{1,n} v_1, \mathbf{u}_{-1}) \right] \mathcal{G}(\mathbf{u}_{-1}) d\mathbf{u}_{-1} K_1(v_1) dv_1. \end{aligned}$$

Sous les conditions (C.4), (K.D), en rappelant que les densités q_ℓ , $\ell = 1, \dots, d$, sont supposées bornées, on a donc

$$\sup_{x_1 \in \mathcal{C}_1} |\mathbb{E}(\hat{\alpha}_1(x_1)) - g(x_1)| = \mathcal{O}(h_{1,n}^k). \quad (5.33)$$

Pour traiter le terme (II) du membre de droite de (5.32), en reprenant, à nouveau, le raisonnement aboutissant à (5.22) (voir également (4.35) dans le **chapitre 4**), remarquons que

$$\begin{aligned} \mathbb{E}(\hat{C}_n) &= \int_{\mathbb{R}^d} \mathbb{E}(\tilde{m}_{\psi,n,1}(\mathbf{x}))q(\mathbf{x})d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \mathbb{E} \left\{ \mathbb{E} \left[\Psi(Y, C) \frac{K_1\left(\frac{x_1 - X_1}{h_{1,n}}\right) K_2\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1}}{h_{2,n}}\right)}{nh_{1,n}h_{2,n}^{d-1}f(\mathbf{X})} \middle| \mathbf{X} \right] \right\} q(\mathbf{x})d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{1}{h_{1,n}h_{2,n}^{d-1}} m_\psi(\mathbf{z}) K_1\left(\frac{x_1 - z_1}{h_{1,n}}\right) K_2\left(\frac{\mathbf{x}_{-1} - \mathbf{z}_{-1}}{h_{2,n}}\right) q(\mathbf{x})d\mathbf{x}d\mathbf{z}. \end{aligned}$$

La fonction de régression m_ψ étant, par hypothèse, supposée additive, et les fonctions q_ℓ , $\ell = 1, \dots, d$, étant des densités, on en déduit que

$$\begin{aligned} &\mathbb{E}(\hat{C}_n - C_n) \\ &= \sum_{j=1}^d \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{1}{h_{1,n}h_{2,n}^{d-1}} m_j(z_j) K_1\left(\frac{x_1 - z_1}{h_{1,n}}\right) K_2\left(\frac{\mathbf{x}_{-1} - \mathbf{z}_{-1}}{h_{2,n}}\right) q(\mathbf{x})d\mathbf{x}d\mathbf{z} \\ &\quad - \int_{\mathbb{R}^{d-1}} \sum_{j=2}^d m_j(z_j) \mathcal{G}(\mathbf{z}_{-1})d\mathbf{z}_{-1}, \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h_{1,n}} m_1(z_1) K_1\left(\frac{x_1 - z_1}{h_{1,n}}\right) q_1(x_1)dx_1dz_1. \end{aligned}$$

Notons que l'hypothèse (C.4), combinée au fait que, pour tout $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$m_\psi(\mathbf{x}) = \mu + \sum_{\ell=1}^d m_\ell(x_\ell),$$

assure que la fonction m_1 est k fois continûment dérivable, de k -ème dérivée uniformément bornée sur \mathbb{R} . Ainsi, en utilisant le changement de variable classique $v_1 = (x_1 - z_1)/h_{1,n}$, on obtient, sous les conditions (C.4) et (K.D),

$$\begin{aligned} &\mathbb{E}(\hat{C}_n - C_n) - C \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} q_1(x_1) m_1(x_1 + h_{1,n}v_1) K_1(v_1) dv_1 dx_1 - C \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} q_1(x_1) [m_1(x_1 + h_{1,n}v_1) - m_1(x_1)] K_1(v_1) dv_1 dx_1 \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} q_1(x_1) \left[\frac{h_{1,n}^k v_1^k}{k!} m_1^{(k)}(x_1 + \theta v_1 h_{1,n}) \right] K_1(v_1) dv_1 dx_1 \\ &= \mathcal{O}(h_{1,n}^k). \end{aligned} \quad (5.34)$$

En associant (5.32) avec (5.33) et (5.34), nous aboutissons à

$$\sup_{x_1 \in \mathcal{C}_1} |\mathbb{E}(\hat{\eta}_1(x_1)) - \eta_1(x_1)| = \mathcal{O}(h_{1,n}^k). \quad (5.35)$$

Étude du terme Biais($\tilde{m}_{\psi,n}(\mathbf{x})$). On a

$$\begin{aligned} \text{Biais}(\tilde{m}_{\psi,n}(\mathbf{x})) &= \mathbb{E}\tilde{m}_{\psi,n}(\mathbf{x}) - m_{\psi}(\mathbf{x}) \\ &= \int_{\mathbf{R}^d} [m(\mathbf{x} + h_{1,n}\mathbf{v}) - m(\mathbf{x})] K_3(\mathbf{v}) d\mathbf{v}. \end{aligned} \quad (5.36)$$

Sous les hypothèses (C.4) et (K.D), il existe un réel $0 < \theta_2 < 1$ tel que

$$\begin{aligned} \text{Biais}(\tilde{m}_{\psi,n}(\mathbf{x})) &= \int_{\mathbf{R}^d} \sum_{i_1+\dots+i_d=k} \frac{h_{1,n}^k}{k!} \frac{\partial^{i_1+\dots+i_d} m_{\psi}}{\partial x_1^{i_1} \dots \partial x_d^{i_d}}(\mathbf{x} + h_{1,n}\theta_2\mathbf{v}) v_1^{i_1} \dots v_d^{i_d} K_3(\mathbf{v}) d\mathbf{v} \\ &= \mathcal{O}(h_{1,n}^k), \end{aligned} \quad (5.37)$$

uniformément en $\mathbf{x} \in \mathcal{C}$, puisque, sous (C.4), m_{ψ} est k fois continûment différentiable, et \mathcal{C} est compact.

Au vu de (5.29), (5.35) et (5.37), la conclusion (5.25) devient maintenant évidente.

Étape 2 : Démonstration de (5.26). Posons

$$\widehat{m}_{\psi,add}(t(\mathbf{x})) - \mathbb{E}\widehat{m}_{\psi,add}(t(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n \xi_i(t(\mathbf{x})), \quad (5.38)$$

où, pour $i = 1, \dots, n$,

$$\xi_i(\mathbf{u}) = Z_i(\mathbf{u}) - \mathbb{E}Z_i(\mathbf{u}), \quad (5.39)$$

et, pour $i = 1, \dots, n$,

$$\begin{aligned} Z_i &= Z_i(\mathbf{u}) \\ &= \frac{\Psi(Y_i, C_i)}{h_{1,n} h_{2,n}^{d-1} f(\mathbf{X}_i)} \sum_{\ell=1}^d \left\{ \int_{\mathbf{R}^{d-1}} K_1\left(\frac{u_{\ell} - X_{i,\ell}}{h_{1,n}}\right) K_2\left(\frac{\mathbf{x}_{-\ell} - \mathbf{X}_{i,-\ell}}{h_{2,n}}\right) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} \right. \\ &\quad \left. - \int_{\mathbf{R}^d} K_1\left(\frac{x_{\ell} - X_{i,\ell}}{h_{1,n}}\right) K_2\left(\frac{\mathbf{x}_{-\ell} - \mathbf{X}_{i,-\ell}}{h_{2,n}}\right) q(\mathbf{x}) d\mathbf{x} \right\} \\ &\quad + \frac{1}{h_{1,n}^d} \frac{\Psi(Y_i, C_i)}{f(\mathbf{X}_i)} \int_{\mathbf{R}^{d-1}} K_3\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}}\right) q(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (5.40)$$

Comme toutes les fonctions intervenant dans l'expression de Z_i ci-dessus sont supposées bornées, et comme, sous la condition (F.A), f est bornée inférieurement sur \mathcal{C} , la condition (K.C) (K_1 est lipschitzien) implique qu'il existe une constante finie M_{K_1} telle que, uniformément en $\mathbf{x} \in \mathcal{C}$,

$$\begin{aligned} |\widehat{m}_{\psi,add}(\mathbf{x}) - \widehat{m}_{\psi,add}(t(\mathbf{x}))| &\leq \frac{1}{n} \sum_{i=1}^n |Z_i(\mathbf{x}) - Z_i(t(\mathbf{x}))| \\ &\leq \sum_{\ell=1}^d \frac{M_{K_1}}{h_{1,n}^2} |x_{\ell} - t(\mathbf{x})_{\ell}| \\ &\leq \frac{M_{K_1} V_{\mathcal{C}}}{r(n)}, \end{aligned}$$

où nous avons fait usage de la relation (5.24). En choisissant

$$r(n) = \frac{3M_{K_1}V_{\mathcal{C}}}{h_{1,n}^2\varepsilon_n}, \quad (5.41)$$

le résultat (5.26) est établi, de sorte que

$$\mathbb{P}\left\{\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}(\mathbf{x}) - \widehat{m}_{\psi,add}(t(\mathbf{x}))| \geq \varepsilon_n/3\right\} = 0.$$

Etape 3 : Démonstration de (5.27). Les techniques et les calculs étant les mêmes que ceux utilisés pour établir (5.26), nous omettons les détails correspondants.

Etape 4 : Démonstration de (5.28). Pour (5.28), on utilise l'inégalité exponentielle de Bernstein pour montrer que

$$\mathbb{P}\left\{\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}(\mathbf{t}(\mathbf{x})) - \mathbb{E}\widehat{m}_{\psi,add}(\mathbf{t}(\mathbf{x}))| \geq \varepsilon_n/3\right\} \leq 2r(n) \exp\{-nh_{1,n}\varepsilon_n^2\}.$$

Au vu de (5.23) et (5.39), on a

$$\begin{aligned} & \mathbb{P}\left\{\sup_{\mathbf{x} \in \mathcal{C}} \left|\frac{1}{n} \sum_{i=1}^n \xi_i(\mathbf{t}(\mathbf{x}))\right| \geq \varepsilon_n/3\right\} \\ & \leq r(n) \sup_{p=1, \dots, r(n)} \mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i(\mathbf{t}_p)\right| \geq \varepsilon_n/3\right\}. \end{aligned} \quad (5.42)$$

Notre but étant maintenant d'appliquer l'inégalité de Bernstein pour les v.a. indépendantes ξ_i , $i = 1, \dots, n$, il nous faut obtenir des bornes convenables pour $|\xi_i|$ et pour $\mathbb{E}|\xi_i|^2$. Pour cela, on utilise la forme particulière des Z_i , conformément à la définition (5.40). Par un changement de variable et en utilisant le fait que les noyaux sont bornés, on obtient facilement les bornes suivantes

$$|Z_i| \leq \frac{M_1}{h_{1,n}} \text{ et } \mathbb{E}Z_i^2 \leq \frac{M_1}{h_{1,n}}, \quad \text{pour un certain } 0 < M_1 < +\infty.$$

En effet, sous nos hypothèses de travail, les fonctions $\Psi(Y_i, C_i)$ et K_1 sont bornées, et f est bornée inférieurement sur \mathcal{C} . Il existe, par conséquent, une constante M_2 telle que

$$\begin{aligned} |Z_i| & \leq \frac{M_2}{h_{1,n}h_{2,n}^{d-1}} \sum_{l=1}^d \left| \int_{\mathbf{R}^{d-1}} K_2\left(\frac{\mathbf{x}_{-l} - \mathbf{X}_{i,-l}}{h_{2,n}}\right) q_{-l}(\mathbf{x}_{-l}) d\mathbf{x}_{-l} \right| \\ & \quad + \frac{M_2}{h_{1,n}h_{2,n}^{d-1}} \left| \int_{\mathbf{R}^d} K_2\left(\frac{\mathbf{x}_{-l} - \mathbf{X}_{i,-l}}{h_{2,n}}\right) q(\mathbf{x}) d\mathbf{x} \right| + \frac{M_2}{h_{1,n}^d} \left| \int_{\mathbf{R}^d} K_3\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}}\right) q(\mathbf{x}) d\mathbf{x} \right|. \end{aligned}$$

En utilisant des changements de variables simples, il est évident qu'il existe une constante M'_2 telle que

$$\frac{M_2}{h_{2,n}^{d-1}} \sum_{l=1}^d \left| \int_{\mathbf{R}^{d-1}} K_2\left(\frac{\mathbf{x}_{-l} - \mathbf{X}_{i,-l}}{h_{2,n}}\right) q_{-l}(\mathbf{x}_{-l}) d\mathbf{x}_{-l} \right| \leq M'_2,$$

et

$$\frac{M_2}{h_{1,n}^d} \left| \int_{\mathbf{R}^d} K_3\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}}\right) q(\mathbf{x}) d\mathbf{x} \right| \leq M'_2.$$

Il existe donc une constante M_3 , telle que, pour $i = 1, \dots, n$,

$$|Z_i| \leq \frac{M_3}{h_{1,n}}.$$

Par un raisonnement analogue, on peut montrer l'existence d'une constante M'_3 telle que $\mathbb{E}|Z_i| \leq M'_3$. De plus, pour $i = 1, \dots, n$,

$$\mathbb{E}|Z_i|^2 = \mathbb{E}(|Z_i| \times |Z_i|) \leq \frac{M_3}{h_{1,n}} \mathbb{E}|Z_i| \leq \frac{M_3 M'_3}{h_{1,n}} =: \frac{M_4}{h_{1,n}}.$$

Il suffit alors de choisir $M_1 = \max(M_2, M_4)$.

On en déduit que, pour $i = 1, \dots, n$,

$$|\xi_i| \leq \frac{M_1}{h_{1,n}},$$

et on peut alors appliquer l'inégalité de Bernstein. Cette dernière implique que, uniformément en $p = 1, \dots, r(n)$,

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i(\mathbf{t}_p)\right| \geq \frac{\varepsilon_n}{3}\right\} \leq 2 \exp\left\{-\frac{nh_{1,n}\varepsilon_n^2}{9M_1}\right\}. \quad (5.43)$$

En choisissant

$$\varepsilon_n^2 = 9cM_1 \left(\frac{\log n}{n}\right)^{\frac{2k}{2k+1}} = \mathcal{O}(h_{1,n}^{2k}), \quad (5.44)$$

avec $c > 1$, on obtient que, uniformément en $p = 1, \dots, r(n)$,

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i(\mathbf{t}_p)\right| \geq \varepsilon_n\right\} \leq n^{-c}. \quad (5.45)$$

Or, en rappelant que a_2 est la constante intervenant dans l'hypothèse (H.B) et que M_{K_1} et V_C sont celles intervenant dans (5.41) (voir également (5.24)), il vient

$$\begin{aligned} r(n) \times n^{-c} &= \frac{3M_{K_1}V_C}{h_{1,n}^2 3\sqrt{cM_1} \left(\frac{\log n}{n}\right)^{k/(2k+1)}} \times n^{-c} \\ &= \frac{M_{K_1}V_C}{a_2^2 \sqrt{cM_1} (\log n)^{(k+2)/(2k+1)}} \times n^{-c + \frac{2+k}{2k+1}} \\ &= \mathcal{O}\left(n^{-c + \frac{2+k}{2k+1}}\right). \end{aligned} \quad (5.46)$$

En choisissant, par exemple, $c > \frac{2+k}{2k+1} + 1$, et en combinant (5.42), (5.43), (5.45) et (5.46), on conclut la démonstration de (5.28).

Finalement, le résultat du lemme 5.4.1 est immédiat au vu de (5.25), (5.26), (5.27) et (5.28) et du choix (5.44) de ε_n . \square

5.4.3 Cas où la densité des covariables est inconnue et la fonction \bar{G} connue

Dans ce cas, la démonstration repose essentiellement sur la conclusion du lemme 5.4.1 et sur la décomposition (5.47) ci-dessous.

$$\frac{1}{\hat{f}_n} = \frac{1}{f} - \frac{\hat{f}_n - f}{\hat{f}_n f}. \quad (5.47)$$

Dans le cas où f est inconnue, alors que la fonction de survie \bar{G} est connue, les estimateurs des composantes additives et de la fonction de régression additive sont définis *via* les relations suivantes.

$$\hat{\eta}_\ell(x_\ell) = \int_{\mathbf{R}^{d-1}} \tilde{m}_{\psi,n,\ell}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbf{R}^d} \tilde{m}_{\psi,n,\ell}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \quad (5.48)$$

et

$$\hat{m}_{\psi,add}(\mathbf{x}) = \sum_{\ell=1}^d \hat{\eta}_\ell(x_\ell) + \int_{\mathbf{R}^d} \tilde{m}_{\psi,n}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad (5.49)$$

où, en rappelant les définitions (5.8) et (5.6) de $W_{n,i}$ et $W_{n,i}^\ell$,

$$\tilde{m}_{\psi,n}(\mathbf{x}) := \sum_{i=1}^n W_{n,i}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{\bar{G}(Z_i)} \quad \text{et} \quad \tilde{m}_{\psi,n,\ell}(\mathbf{x}) := \sum_{i=1}^n W_{n,i}^\ell(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{\bar{G}(Z_i)}.$$

Nous allons chercher à établir un résultat d'approximation permettant d'évaluer, uniformément en $\mathbf{x} \in \mathcal{C}$, l'écart entre $\hat{m}_{\psi,add}$ et $\tilde{m}_{\psi,add}$.

Lemme 5.4.2. *Sous les hypothèses du théorème 5.3.1, on a, avec probabilité 1,*

$$\sup_{\mathbf{x} \in \mathcal{C}} |\hat{m}_{\psi,add}(\mathbf{x}) - \tilde{m}_{\psi,add}(\mathbf{x})| = \mathcal{O} \left(\sqrt{\frac{\log n}{nh_n^d}} \right). \quad (5.50)$$

Démonstration. Soit $b > 0$, tel que $\inf_{\mathbf{x} \in \mathcal{C}^\alpha} \{\min(f(\mathbf{x}), \hat{f}_n(\mathbf{x}))\} \geq b$ (l'existence d'un tel réel est assuré, à partir d'un certain rang [apcr], par l'hypothèse (F.A)). Par les définitions (5.7) et (5.21), et la décomposition (5.47), on constate que apcr,

$$\begin{aligned} & \sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n |W_{n,i}(\mathbf{x}) - \tilde{W}_{n,i}(\mathbf{x})| \\ & \leq \sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n \left| \frac{K_3 \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}} \right)}{nh_{1,n}^d f(\mathbf{X}_i) \hat{f}_n(\mathbf{X}_i)} \right| \times |\hat{f}_n(\mathbf{x}) - f(\mathbf{x})| \\ & \leq \frac{1}{b^2 + o(1)} \frac{1}{h_{1,n}^d} E \left| K_3 \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}} \right) \right| \sum_{i=1}^n \frac{\left| K_3 \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}} \right) \right|}{E \left| K_3 \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_{1,n}} \right) \right|} \sup_{\mathbf{x} \in \mathcal{C}^\alpha} |\hat{f}_n(\mathbf{x}) - f(\mathbf{x})| \\ & = \mathcal{O} \left(\sqrt{\frac{\log n}{nh_n^d}} \right) \quad \text{p.s.}, \end{aligned} \quad (5.51)$$

où l'on a utilisé le caractère borné de K_3 (impliqué par l'hypothèse que K_3 est continu et à support compact), le fait que q soit à support inclus dans \mathcal{C} , ainsi que le résultat suivant sur l'estimateur \hat{f}_n (voir, par exemple, Ango-Nze et Rios (2000)). Sous les hypothèses (F.A-B), (H.A) et (K.B), on a

$$\sup_{\mathbf{x} \in \mathcal{C}^\alpha} |f(\mathbf{x}) - \hat{f}_n(\mathbf{x})| = \mathcal{O} \left(\sqrt{\frac{\log n}{nh_n^d}} \right) \text{ p.s.}$$

En reprenant les mêmes techniques, on obtient également, pour $\ell = 1, \dots, d$,

$$\sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n |W_{n,i}^\ell(\mathbf{x}) - \widetilde{W}_{n,i}^\ell(\mathbf{x})| = \mathcal{O} \left(\sqrt{\frac{\log n}{nh_n^d}} \right) \text{ p.s.}$$

De plus, sous les conditions (A), (C.2) et (C.3), on a $\max_{1 \leq i \leq n} \psi(Z_i)/\overline{G}(Z_i) < \infty$.

Le support de chaque densité $q_\ell, \ell = 1, \dots, d$ étant inclus dans \mathcal{C}_ℓ , et compte tenu des définitions (5.18), (5.19), (5.48) et (5.49), il s'ensuit que

$$\begin{aligned} & \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}(\mathbf{x}) - \widehat{m}_{\psi,add}(\mathbf{x})| \\ & \leq 2d \max_{1 \leq \ell \leq d} \sup_{\mathbf{x} \in \mathcal{C}} |\widetilde{m}_{\psi,n,\ell}(\mathbf{x}) - \widetilde{m}_{\psi,n,\ell}(\mathbf{x})| + \sup_{\mathbf{x} \in \mathcal{C}} |\widetilde{m}_{\psi,n}(\mathbf{x}) - \widetilde{m}_{\psi,n}(\mathbf{x})| \\ & \leq 2d \max_{1 \leq \ell \leq d} \sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n |W_{n,i}^\ell(\mathbf{x}) - \widetilde{W}_{n,i}^\ell(\mathbf{x})| + \sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n |W_{n,i}(\mathbf{x}) - \widetilde{W}_{n,i}(\mathbf{x})| \\ & = \mathcal{O} \left(\sqrt{\frac{\log n}{nh_n^d}} \right) \text{ p.s.,} \end{aligned}$$

ce qui achève la démonstration du lemme 5.4.2. \square

5.4.4 Cas où la densité des covariables et la fonction \overline{G} sont inconnues

Remarquons que,

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}^*(\mathbf{x}) - m_\psi(\mathbf{x})| & \leq \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}^*(\mathbf{x}) - \widehat{m}_{\psi,add}(\mathbf{x})| \\ & \quad + \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}(\mathbf{x}) - m_\psi(\mathbf{x})|. \end{aligned}$$

Au vu des résultats précédents, il nous suffit, maintenant, d'établir le lemme suivant.

Lemme 5.4.3. *Sous les hypothèses du Théorème 5.3.1, on a*

$$\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}^*(\mathbf{x}) - \widehat{m}_{\psi,add}(\mathbf{x})| = \mathcal{O} \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ p.s.} \quad (5.52)$$

Démonstration. Le support de chaque densité $q_\ell, \ell = 1, \dots, d$ étant inclus dans \mathcal{C}_ℓ , on a

$$\begin{aligned} & \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_{\psi,add}^*(\mathbf{x}) - \widehat{m}_{\psi,add}(\mathbf{x})| \\ & \leq \sum_{\ell=1}^d \sup_{x_\ell \in \mathcal{C}_\ell} |\widehat{\eta}_\ell^*(x_\ell) - \widehat{\eta}_\ell(x_\ell)| + \int_{\mathbf{R}^d} \sup_{\mathbf{x} \in \mathcal{C}} |\widetilde{m}_{\psi,n}^*(\mathbf{x}) - \widetilde{m}_{\psi,n}(\mathbf{x})| q(\mathbf{x}) d\mathbf{x} \\ & \leq 2d \max_{1 \leq \ell \leq d} \sup_{\mathbf{x} \in \mathcal{C}} |\widetilde{m}_{\psi,n,\ell}^*(\mathbf{x}) - \widetilde{m}_{\psi,n,\ell}(\mathbf{x})| + \sup_{\mathbf{x} \in \mathcal{C}} |\widetilde{m}_{\psi,n}^*(\mathbf{x}) - \widetilde{m}_{\psi,n}(\mathbf{x})|. \end{aligned}$$

D'autre part, sous l'hypothèse **(A)**, il ressort de ce qui précède que,

$$\begin{aligned}
& \sup_{\mathbf{x} \in \mathcal{C}} |\tilde{m}_{\psi,n}^*(\mathbf{x}) - \tilde{m}_{\psi,n}(\mathbf{x})| \\
& \leq \sup_{t \leq \tau} |\psi(t)| \sup_{t \leq \tau} \left| \frac{1}{\bar{G}(t)} - \frac{1}{\bar{G}_n^*(t)} \right| \sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n |W_{n,i}(\mathbf{x})| \\
& \leq \sup_{t \leq \tau} |\psi(t)| \sup_{t \leq \tau} |\bar{G}(t) - \bar{G}_n^*(t)| \sup_{t \leq \tau} \frac{1}{\bar{G}_n^*(t)\bar{G}(t)} \sup_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^n |W_{n,i}(\mathbf{x})|.
\end{aligned} \tag{5.53}$$

Or, comme $\bar{G}(\tau) > 0$ et $\bar{F}(\tau) > 0$, la loi du logarithme itéré de Földes et Rejtő (1981) implique que

$$\sup_{t \leq \tau} |\bar{G}(t) - \bar{G}_n^*(t)| = \mathcal{O}\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ p.s. .} \tag{5.54}$$

De plus, en suivant le même raisonnement que celui menant à (5.51), on peut montrer que la quantité $\sum_{i=1}^n |W_{n,i}(\mathbf{x})|$ est presque sûrement uniformément bornée. En combinant ce dernier résultat à (5.54), on obtient,

$$\sup_{\mathbf{x} \in \mathcal{C}} |\tilde{m}_{\psi,n}^*(\mathbf{x}) - \tilde{m}_{\psi,n}(\mathbf{x})| = \mathcal{O}\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ p.s. .} \tag{5.55}$$

En utilisant une argumentation analogue, on peut montrer que

$$\sup_{\mathbf{x} \in \mathcal{C}} |\tilde{m}_{\psi,n,\ell}^*(\mathbf{x}) - \tilde{m}_{\psi,n,\ell}(\mathbf{x})| = \mathcal{O}\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ p.s., } \forall \ell = 1, \dots, d. \tag{5.56}$$

On déduit de (5.55) et (5.56), finalement, le résultat énoncé dans le lemme 5.4.3 . \square

En combinant les lemmes 5.4.1, 5.4.2 et 5.4.3, et au vu de l'hypothèse **(H.A)**, on achève, directement, la démonstration du Théorème 5.3.1.

Annexe A

Les classes à nombre de recouvrement polynomial

Résumé. Dans cette annexe, après avoir défini les notions de *nombre de recouvrement*, nous nous intéresserons aux classes présentant un nombre de recouvrement polynomial. Ces classes de fonctions interviennent notamment dans l'étude du processus empirique indexé par des classes de fonction, en vue de la généralisation des théorèmes de Glivenko-Cantelli et de Donsker. Nous nous attarderons notamment sur les classes de Vapnik-Červonenkis (VC) d'ensembles et les classes VC de fonctions (classes VC d'hypographe). Ces dernières constituent un sous-ensemble des classes de fonctions à nombre de recouvrement polynomial qu'il est intéressant d'étudier, car il est aisé de les caractériser, de construire des VC classes à partir d'autres VC classes, etc.

Ce chapitre est largement inspiré des ouvrages de Pollard (1984), van der Vaart et Wellner (1996) et Dudley (1999), auxquels nous renvoyons pour plus de détails sur le sujet (et notamment pour les preuves des résultats non démontrés dans cette annexe).

A.1 Nombres de recouvrement d'un ensemble

Soit (S, d) un espace semi-métrique, et ε un réel strictement positif. On définit le *nombre de recouvrement* $N(\varepsilon, S, d)$ de (S, d) comme le nombre minimum de boules de rayon ε , relativement à la semi-métrique d , nécessaire pour recouvrir S . Plus formellement, on a

$$N(\varepsilon, S, d) = \min\{m : \exists y_1, \dots, y_m \in S : \forall x \in S, \exists i : d(x, y_i) \leq \varepsilon\}. \quad (\text{A.1})$$

D'autre part, on définit le *packing number* (ou *bracketing number*) $D(\varepsilon, S, d)$ de (S, d) comme le nombre maximum de points de S ε -séparés (où une collection de points est dite ε -séparée si $d(x_i, x_j) > \varepsilon$ pour tout x_i, x_j de la collection). Formellement, on obtient

$$D(\varepsilon, S, d) = \max\{m : \exists y_1, \dots, y_m \in S : \forall i, j \ d(y_i, y_j) > \varepsilon\}. \quad (\text{A.2})$$

Remarque A.1.1. *En l'absence de confusion possible, nous noterons $N(\varepsilon, d)$ la quantité $N(\varepsilon, S, d)$.*

Ces deux définitions sont similaires, et conduisent généralement à des résultats équivalents, ce qui peut être résumé par la proposition suivante (voir pp. 83-84 de van der Vaart et Wellner (1996)).

Proposition A.1.1. Soient ε un réel strictement positif et (S, d) un espace semi-métrique. Les quantités $N(\varepsilon, S, d)$ et $D(\varepsilon, S, d)$ définies ci-dessus vérifient les inégalités

$$N(\varepsilon, S, d) \leq D(\varepsilon, S, d) \leq N\left(\frac{\varepsilon}{2}, S, d\right). \quad (\text{A.3})$$

Ayant introduit les notions de nombres de recouvrement, nous allons pouvoir définir les classes d'ensembles, puis de fonctions, dites de Vapnik-Červonenkis.

A.2 Classes d'ensembles de Vapnik-Červonenkis

A.2.1 Définitions et notations

Soit \mathcal{C} une collection de sous-ensembles d'un ensemble \mathcal{X} . Soit par ailleurs A un sous-ensemble de \mathcal{X} . Par la suite, nous dirons que \mathcal{C} *picks out* A , ou, de manière équivalente, A est *picked out* par \mathcal{C} , si $A = C \cap \mathcal{X}$, pour un certain $C \in \mathcal{C}$.

Introduisons maintenant les quantités suivantes.

$$C \sqcap A := \mathcal{C}_A := \{C \cap A, C \in \mathcal{C}\} \quad (\text{A.4})$$

$$\text{et } \Delta^{\mathcal{C}}(A) := \text{card}(\mathcal{C}_A) =: |\mathcal{C}_A|. \quad (\text{A.5})$$

Définition A.2.1. Si $\mathcal{C}_A = 2^A$, alors \mathcal{C} est dit *pulvériser* A .

Remarque A.2.1. Si A est fini, alors \mathcal{C} pulvérise A si et seulement si $\Delta^{\mathcal{C}}(A) = 2^{|A|}$.

Soit $n \leq |\mathcal{X}|$, on appelle coefficient de pulvérisation la quantité

$$m^{\mathcal{C}}(n) := \max\{\Delta^{\mathcal{C}}(F), F \subset \mathcal{X}, |F| = n\}. \quad (\text{A.6})$$

Pour $n \geq |\mathcal{X}|$, on adoptera la convention $m^{\mathcal{C}}(n) = m^{\mathcal{C}}(|\mathcal{X}|)$.

On a $m^{\mathcal{C}}(n) \leq 2^n$, pour tout $n \geq 1$.

Remarque A.2.2. Remarquons que si $|F| = n$ et \mathcal{C} pulvérise F , alors $m^{\mathcal{C}}(n) = 2^n$. Par ailleurs, une des propriétés fondamentales de la théorie de Vapnik-Červonenkis est le résultat du corollaire A.2.1 présenté plus bas : si $m^{\mathcal{C}}(n) \leq 2^n$, alors $m^{\mathcal{C}}(n)$ est majoré par un polynôme en n . Nous reviendrons sur ce point par la suite.

Deux quantités supplémentaires sont classiquement utilisées pour traiter des classes d'ensembles.

$$V(\mathcal{C}) := \begin{cases} \inf\{n : m^{\mathcal{C}}(n) < 2^n\}, & \text{si l'infimum est fini,} \\ +\infty, & \text{si } m^{\mathcal{C}}(n) = 2^n \text{ pour tout } n, \end{cases}$$

et

$$S(\mathcal{C}) := \begin{cases} \sup\{n : m^{\mathcal{C}}(n) = 2^n\}, & \text{si } \mathcal{C} \text{ est non vide,} \\ +\infty, & \text{si } \mathcal{C} \text{ est vide.} \end{cases}$$

La quantité $V(\mathcal{C})$ est appelée **VC-index** (ou index VC) de la classe \mathcal{C} .

Remarquons que $S(\mathcal{C}) = V(\mathcal{C}) - 1$, dès lors que ces deux quantités sont finies.

Au regard de ces définitions, nous pouvons introduire la notion de classes d'ensembles de Vapnik-Červonenkis.

Définition A.2.2. Si $S(\mathcal{C})$, ou de manière équivalente $V(\mathcal{C})$, est finie, alors la classe \mathcal{C} est dite de Vapnik-Červonenkis, ou VC .

A.2.2 Propriétés fondamentales des classes VC

Soit la quantité

$${}_N C_{\leq k} := \sum_{j=0}^k C_N^j = \sum_{j=0}^k \frac{N!}{j!(N-j)!}. \quad (\text{A.7})$$

Vapnik et Červonenkis ont obtenu une majoration fine de ${}_N C_{\leq k}$, correspondant pratiquement à une équivalence à l'infini.

Proposition A.2.1. *Pour tous entiers k et n , si $n \geq k + 2$, alors ${}_n C_{\leq k} \leq \frac{3}{2}n^k/k!$.*

Démonstration. Procédons par récurrence sur k et n . Pour $k = 1$, l'inégalité devient $n + 1 \leq \frac{3}{2}n$ et est ainsi trivialement vérifié pour $n \geq 2$. Pour $n = k + 2$, l'inégalité devient

$$2^n - n - 1 \leq \frac{3n^{n-2}}{2(n-2)!} = \frac{3(n-1)n^{n-1}}{2n!}. \quad (\text{A.8})$$

Or, d'après la formule de Stirling,

$$n! \leq \left(\frac{n}{e}\right)^n (2\pi n)^{1/2} e^{1/12n}.$$

Ainsi, il suffit d'établir le résultat suivant pour obtenir le résultat pour $n = k + 2$,

$$(2n/e)^n (2\pi n)^{1/2} e^{1/12} \leq \frac{3}{2}(n-1)n^{n-1}. \quad (\text{A.9})$$

En posant $f(x) = (e/2)^x$ et $g(x) = 2x^{1/2}$, on a pour tout $x \geq 7$, $f(x) \geq g(x)$, $f'(x) \geq g'(x)$, $f''(x) > 0$ et $g''(x) < 0$. Comme, de plus, $\sqrt{\frac{2\pi}{4}} \leq \frac{3}{2}$ et $ne^{1/12n} \leq n - 1$, pour $n \geq 7$, l'inégalité (A.9) (et donc (A.8)) est vérifiée pour $n \geq 7$. D'autre part, l'inégalité (A.8) est vérifiée pour $n=3, 4, 5$ et 6 .

On suppose maintenant que la proposition est vraie pour $n = k + i, i = 2, \dots, j$ et pour $n = k + J, J := j + 1, k = 1, \dots, K$. Pour achever la démonstration, il suffit d'obtenir l'inégalité pour $n = k + J$ et $k = K + 1$. Remarquons que, d'après le théorème binomial, on a $(n + 1)^k \geq n^{k-1}(n + k)$, soit

$$\frac{n^{k-1}}{(k-1)!} + \frac{n^k}{k!} \leq \frac{(n+1)^k}{k!}. \quad (\text{A.10})$$

Alors, par le triangle de Pascal et d'après les hypothèses de récurrence,

$$\begin{aligned} {}_n C_{\leq k} &= {}_{k+J} C_{\leq K+1} \\ &= {}_{k+j} C_{\leq K+1} + {}_{K+J} C_{\leq K} \\ &\leq \frac{3(k+j)^k}{2k!} + \frac{3(K+J)^K}{2K!} \\ &\leq \frac{3n^k}{2k!}, \end{aligned}$$

où nous avons utilisé (A.10), en notant que $k + j = K + J$. \square

Le résultat suivant est connu sous le nom de lemme de Sauer (cf. Sauer (1972), ou encore Dudley (1999)).

Proposition A.2.2. *Pour tout entier $k \geq 1$, si $m^{\mathcal{C}}(n) > {}_n C_{\leq k-1}$, alors $m^{\mathcal{C}}(k) = 2^k$. Ainsi, si $S(\mathcal{C}) \leq \infty$, alors $m^{\mathcal{C}}(n) \leq {}_n C_{\leq S(\mathcal{C})}$, pour tout n .*

Démonstration. La preuve est empruntée à van der Vaart et Wellner (1996). Elle repose essentiellement sur le résultat suivant.

Lemme A.2.1. *Soit $\{x_1, \dots, x_n\}, n \geq 1$ un ensemble de points. Le nombre total de sous-ensembles picked out par \mathcal{C} , une collection de parties, dans $\{x_1, \dots, x_n\}$, est majoré par le nombre de sous-ensembles de $\{x_1, \dots, x_n\}$ pulvérisés par \mathcal{C} .*

Démonstration. On se place, sans perte de généralité, dans le cas où \mathcal{C} est constituée de parties $C \subset \{x_1, \dots, x_n\}$. On a alors

$$\Delta^{\mathcal{C}}(\{x_1, \dots, x_n\}) = |\mathcal{C}|. \quad (\text{A.11})$$

Si \mathcal{C} est *héréditaire* (i.e. $(B \subset C \text{ et } C \in \mathcal{C}) \Rightarrow B \in \mathcal{C}$), alors le résultat est immédiat. Le but est maintenant de montrer que l'on peut toujours se ramener à une classe héréditaire. Pour tout indice $1 \leq i \leq n$ et tout $C \in \mathcal{C}$, on définit

$$T_i(C) = \begin{cases} C - \{x_i\} & \text{si } C - \{x_i\} \notin \mathcal{C} \\ C & \text{sinon.} \end{cases}$$

Si $x_i \notin C$, alors $T_i(C) = C$, et on ôte x_i à C seulement si cela crée une nouvelle partie dans \mathcal{C} . Nous allons montrer que si $T_i(\mathcal{C})$ vérifie le résultat du lemme, alors \mathcal{C} aussi. Soit $A \subset \{x_1, \dots, x_n\}$, un sous-ensemble pulvérisé par $T_i(\mathcal{C})$.

- Si $x_i \notin A$, alors pour tout $a \subset 2^A$, il existe $C \in \mathcal{C}$ tel que $a = A \cap T_i(C)$. Or $A \cap (C - \{x_i\}) = A \cap C$, donc $a = A \cap C$, et finalement \mathcal{C} pulvérise A .
- Si $x_i \in A$, alors pour tout $a \subset 2^A$, il existe $C \in \mathcal{C}$ tel que $a \cup \{x_i\} = T_i(C) \cap A$ (car il existe $b \subset A$ tel que $a \cup \{x_i\} = b$). On en déduit que $x_i \in T_i(C)$, et, par suite, que $T_i(C) = C$. Ainsi, $a \cup \{x_i\}$ est *picked out* par \mathcal{C} . D'autre part $T_i(C) = C \Rightarrow C - \{x_i\} \in \mathcal{C}$, et $a - \{x_i\} = (C - \{x_i\}) \cap A$ est lui aussi pulvérisé par \mathcal{C} . Finalement, a est *picked out* par \mathcal{C} (car soit $a = a - \{x_i\}$, soit $a = a \cup \{x_i\}$), donc \mathcal{C} pulvérise A .

De plus, l'application T_i étant bijective, on a $|\mathcal{C}| = |T_i(\mathcal{C})|$. En combinant cette dernière égalité et (A.11), il vient $\Delta^{\mathcal{C}}(\{x_1, \dots, x_n\}) = |\mathcal{C}| = \Delta^{T_i(\mathcal{C})}(\{x_1, \dots, x_n\})$. Ainsi, le lemme est vérifié par \mathcal{C} s'il est vérifié par $T_i(\mathcal{C})$.

Le même type de résultat peut être obtenu pour l'opérateur $\mathbf{T}_n = T_1 \circ \dots \circ T_n$. On introduit alors la suite d'opérateurs définie par

$$\begin{cases} \mathbf{T}_{n,1} &= \mathbf{T}_n \\ \mathbf{T}_{n,k} &= \mathbf{T}_{n,k-1} \circ \mathbf{T}_n \text{ pour } k \geq 2. \end{cases}$$

Comme $\sum_{C \in \mathcal{C}} |T_i(C)| < \sum_{C \in \mathcal{C}} |C|$ dès que $T_i(C) \neq C$, il existe un entier ℓ tel que $\mathbf{T}_{n,\ell+1}(\mathcal{C}) = \mathbf{T}_{n,\ell}(\mathcal{C})$. De plus, $\ell \leq \sum_{C \in \mathcal{C}} |C|$. Soit alors $\mathcal{D} = \mathbf{T}_{n,\ell}(\mathcal{C})$. Pour tout $1 \leq i \leq n$, et tout $D \in \mathcal{D}$, $D - \{x_i\} \in \mathcal{D}$, donc \mathcal{D} est héréditaire, et vérifie le résultat du lemme. Finalement, le résultat étant vérifié par $\mathbf{T}_{n,\ell}(\mathcal{C})$, il l'est aussi par \mathcal{C} . \square

Le résultat de la proposition A.2.2 est maintenant immédiat. En effet, une classe \mathcal{C} d'index $V(\mathcal{C})$ ne pulvérisant aucun ensemble de $V(\mathcal{C})$ points, tous les ensembles de points pulvérisés par \mathcal{C} font partie des ensembles de cardinal au plus $V(\mathcal{C}) - 1$. Or il y a exactement ${}_n C_{\leq V(\mathcal{C})-1}$

ensembles de cardinal $V(\mathcal{C})$ dans $\{x_1, \dots, x_n\}$, ce qui permet de conclure au vu du lemme A.2.1. \square

En combinant les résultats des propositions A.2.1 et A.2.2, le corollaire suivant est immédiat.

Corollaire A.2.1. *Soit \mathcal{C} une classe VC, alors, pour $n \geq S(\mathcal{C}) + 2$, on a*

$$m^{\mathcal{C}}(n) \leq \frac{3n^{S(\mathcal{C})}}{2S(\mathcal{C})!}. \quad (\text{A.12})$$

A.2.3 Construction de classes VC

Dans ce paragraphe, nous montrons comment obtenir des classes VC (les construire, ou, « symétriquement », montrer qu'une classe donnée est VC). Le premier principe sera, dans les deux cas, de se ramener à une classe « simple » et en obtenir la propriété VC. Ensuite, la stabilité par union, intersection et produit cartésien (cf. les propositions présentées ci-après) des classes VC permet de construire des classes plus « complexes ». Notons cependant que ce premier principe n'est évidemment pas suffisant pour générer l'ensemble des classes VC, ou obtenir la propriété VC de l'ensemble des classes VC (voir les ouvrages de Pollard (1984), van der Vaart et Wellner (1996) et Dudley (1999) pour de plus amples détails).

Classes VC d'index 1 et autres classes VC triviales

Proposition A.2.3. *Une classe \mathcal{C} de sous-ensembles d'un ensemble \mathcal{X} présente un index $V(\mathcal{C}) = 0$ (ou de manière équivalente $S(\mathcal{C}) = -1$) si et seulement si \mathcal{C} est vide. De plus, $V(\mathcal{C}) = 1$ (ou de manière équivalente $S(\mathcal{C}) = 0$) si et seulement si \mathcal{C} ne contient qu'un ensemble. Ainsi, $S(\mathcal{C}) \geq 1$ si et seulement si \mathcal{C} contient au moins deux ensembles.*

Démonstration. \mathcal{C} pulvérise l'ensemble vide si et seulement si \mathcal{C} contient au moins un ensemble. Par ailleurs, si \mathcal{C} contient au moins deux ensembles, alors il existe A, B dans \mathcal{C} et $x \in X$ tels que $x \in A \setminus B$. Donc \mathcal{C} pulvérise $\{x\}$, et $S(\mathcal{C}) \geq 1$. Réciproquement, si $S(\mathcal{C}) \geq 1$, alors \mathcal{C} contient au moins deux ensembles. \square

Proposition A.2.4. *Soit \mathcal{C} une collection d'au moins deux sous-ensembles d'un ensemble \mathcal{X} . $S(\mathcal{C}) = 1$, si*

1. \mathcal{C} est linéairement ordonné pour l'inclusion,
ou si
2. les sous-ensembles de \mathcal{C} sont deux à deux disjoints.

Démonstration. Compte tenu de la proposition précédente, on a dans les deux cas $S(\mathcal{C}) \geq 1$. Raisonnons alors par l'absurde.

Si \mathcal{C} est linéairement ordonné pour l'inclusion, soit $\{x, y\}$ un ensemble pulvérisé par \mathcal{C} . On sélectionne $A, B \in \mathcal{C}$ tels que $A \cap \{x, y\} = \{x\}$ et $B \cap \{x, y\} = \{y\}$. Alors l'hypothèse $A \subset B$ ou $B \subset A$ conduit à une contradiction.

De même, si les ensembles de \mathcal{C} sont disjoints, en raisonnant comme dans le cas précédent, et en prenant $C \in \mathcal{C}$ tel que $\{x, y\} = C \cap \{x, y\}$, l'hypothèse C et A disjoints (ou C et B disjoints) conduit à une contradiction. \square

Considérons maintenant deux exemples de classes d'ensembles VC simples. Lorsque $\mathcal{X} = \mathbb{R}$ et $\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}$, alors $V(\mathcal{C}) = 2$ puisqu'aucun couple de points n'est pulvérisé par \mathcal{C} (alors que les singletons le sont). En effet,

$$\forall F = \{x_1, x_2\}, x_1 < x_2 \Rightarrow F' = x_2 \neq F \cap (-\infty, t], \forall t \in \mathbb{R}.$$

Lorsque $\mathcal{X} = \mathbb{R}$ et $\mathcal{C} = \{(s, t] : (s, t) \in \mathbb{R}^2\}$, alors $V(\mathcal{C}) = 3$ puisqu'aucun triplet de points n'est pulvérisé par \mathcal{C} (alors que les couples le sont). En effet,

$$\forall F = \{x_1, x_2, x_3\}, x_1 < x_2 < x_3 \Rightarrow F' = x_1, x_3 \neq F \cap (s, t], \forall (s, t) \in \mathbb{R}^2.$$

Propriétés de stabilité des classes VC

Nous présentons ici des propriétés permettant notamment de construire des classes VC « complexes » à partir de classes VC initiales plus simples.

Lemme A.2.2. *Soient \mathcal{C} et \mathcal{D} deux classes VC de sous-ensembles d'un ensemble \mathcal{X} , $\phi : \mathcal{X} \mapsto Y$. Soient, de plus, $\psi : Z \mapsto \mathcal{X}$ deux fonctions, où Y et Z sont deux ensembles donnés. Alors*

1. $\mathcal{C}^c = \{C^c, C \in \mathcal{C}\}$ est VC ;
2. $\mathcal{C} \sqcup \mathcal{D} = \{C \cup D, C \in \mathcal{C}, D \in \mathcal{D}\}$ est VC ;
3. $\mathcal{C} \cap \mathcal{D} = \{C \cap D, C \in \mathcal{C}, D \in \mathcal{D}\}$ est VC ;
4. $\phi(\mathcal{C})$ est VC si ϕ est bijective ;
5. $\psi^{-1}(\mathcal{C})$ est VC ;
6. l'adhérence (au sens de la convergence ponctuelle des fonctions indicatrices) de \mathcal{C} est VC .

Pour des classes VC , \mathcal{C} et \mathcal{D} , de deux ensembles X et Y , on a

7. $\mathcal{C} \boxtimes \mathcal{D} := \{C \times D, C \in \mathcal{C}, D \in \mathcal{D}\}$ est VC dans $X \times Y$.

A.3 Les classes de fonctions VC

La théorie développée par Vapnik et Červonenkis sur les classes d'ensembles s'étend naturellement aux classes de fonctions. Il existe plusieurs définitions, non équivalentes, des classes VC de fonctions. Nous n'évoquerons ici que les classes VC dites d'hypographe (pour les définitions des classes de fonctions VC dites *major* et *hull*, nous renvoyons notamment à van der Vaart et Wellner (1996)). La définition la plus intéressante, dans notre cadre de travail, est celle des classes VC d'hypographe, puisque ces dernières présentent la propriété de nombre de recouvrement polynomial. Les autres types de classes de fonctions VC ne présentent généralement que des *nombre d'entropie*¹ polynômiaux, ce qui est essentiellement suffisant pour montrer qu'une classe est GC ou Donsker par exemple.

Après avoir introduit la notion d'hypographe, nous définirons les classes de fonctions VC . Enfin, nous verrons que les classes VC d'hypographe sont incluses dans l'ensemble des classes présentant un nombre de recouvrement polynomial.

¹Les *nombre d'entropie* correspondent au *logarithme* des nombres de recouvrement.

A.3.1 Définitions

Définition A.3.1. *L'hypographe d'une fonction $f : \mathcal{X} \rightarrow \mathbb{R}$ est le sous-ensemble de $\mathcal{X} \times \mathbb{R}$*

$$\{(x, t) : t < f(x)\}. \quad (\text{A.13})$$

Une définition menant à des résultats équivalents est obtenue en remplaçant formellement ' $<$ ' par ' \leq ' dans (A.13) (voir le problème 2.6.10 de van der Vaart et Wellner (1996)).

Définition A.3.2. *Une classe \mathcal{F} de fonctions mesurables sur un espace \mathcal{X} est appelée VC subgraph class (ou classe VC, ou classe VC d'hypographe), si la classe de tous les hypographe des fonctions de \mathcal{F} forme une classe d'ensembles VC dans $\mathcal{X} \times \mathbb{R}$.*

A.3.2 Propriétés-Exemples

Soit \mathcal{F} , une classe de fonctions VC. Soit, de plus, $V(\mathcal{F})$, le VC-index de la classe (d'ensembles) des hypographe des fonctions composant \mathcal{F} . La proposition A.3.1 qui suit est fondamentale, puisqu'elle assure que les classes de fonctions VC présentent un nombre de recouvrement polynomial (voir p. 141 dans van der Vaart et Wellner (1996)).

Proposition A.3.1. *Soit \mathcal{F} une classe de fonctions VC présentant une fonction enveloppe F mesurable, avec $F(x) \geq \sup_{f \in \mathcal{F}} f(x)$, $x \in \mathcal{X}$. Soit, de plus, un entier $r \geq 1$. Pour toute mesure de probabilité Q telle que $\|F\|_{Q,r} > 0$, on a, pour tout réel $0 < \varepsilon < 1$,*

$$\mathcal{N}(\varepsilon \|F\|_{Q,r}, \mathcal{F}, \mathbb{L}_r(Q)) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{r(V(\mathcal{F})-1)}, \quad (\text{A.14})$$

où K est une constante universelle.

Citons un exemple utile de classes de fonctions VC (voir le lemme 2.6.15 dans van der Vaart et Wellner (1996)). Un espace vectoriel \mathcal{F} de dimension finie, composé de fonctions mesurables de $\mathcal{X} \times \mathbb{R}$, est une classe de fonctions VC, avec $V(\mathcal{F}) \leq \dim(\mathcal{F}) + 2$.

Comme autres exemples de classes de fonctions VC simples, nous pouvons citer les classes de fonctions indicatrices du type $\mathbb{I}_{(-\infty, t]}$, pour $t \in \mathbb{R}$, ou $\mathbb{I}_{(s, t]}$, pour $(s, t) \in \mathbb{R}^2$.

A.4 Un résultat utile

Pour conclure cette annexe, nous citons un résultat établi par Einmahl et Mason (2000). Celui-ci est notamment très utile dans nos preuves. Il assure la « stabilité par produit cartésien » des classes de fonctions présentant un nombre de recouvrement polynomial.

Lemme A.4.1. *Soient \mathcal{F} et \mathcal{G} , deux classes de fonctions réelles mesurables définies sur un ensemble \mathcal{X} vérifiant,*

- (a) $|f(x)| \leq F(x)$, $f \in \mathcal{F}$, $x \in \mathcal{X}$, F étant une fonction enveloppe bornée sur \mathcal{X} ;
- (b) $\sup_{x \in \mathcal{X}} |g(x)| \leq M < \infty$, $g \in \mathcal{G}$.

Supposons que, pour toute mesure de probabilité Q telle que $0 < Q(F^2) < \infty$,

$$\mathcal{N}(\varepsilon(Q(F^2))^{1/2}, \mathcal{F}, d_Q) \leq C_1 \varepsilon^{-\nu_1}, \quad 0 < \varepsilon < 1,$$

et pour toute mesure de probabilité Q ,

$$\mathcal{N}(\varepsilon M, \mathcal{G}, d_Q) \leq C_2 \varepsilon^{-\nu_2}, \quad 0 < \varepsilon < 1,$$

où C_1, C_2, ν_1 et ν_2 sont des constantes convenables.

Alors, pour toute mesure de probabilité Q vérifiant $0 < Q(F^2) < \infty$, il existe une constante universelle $0 < C_3 < \infty$, telle que, pour tout réel $0 < \varepsilon < 1$,

$$N(\varepsilon M(Q(F^2))^{1/2}, \mathcal{FG}, d_Q) \leq C_3 \varepsilon^{-\nu_1 \nu_2},$$

où $\mathcal{FG} := \{fg, f \in \mathcal{F}, g \in \mathcal{G}\}$.

Perspectives de recherche

Dans cette partie, nous présentons brièvement certains travaux en cours.

A.5 Fonctionnelles de la loi conditionnelle en données censurées

A.5.1 Estimateurs des lois conditionnelles avec censure

Dans le **chapitre 4**, nous avons obtenu, parmi d'autres résultats, une loi du logarithme pour un estimateur non paramétrique de la fonction de répartition conditionnelle en données censurées, et ce en sélectionnant la classe $\mathcal{F} = \{\mathbb{I}_{(-\infty, t]}, t \leq \tau\}$, $\tau < T_H$. Ce résultat peut être généralisé en considérant les classes de fonctions, aléatoires, $\mathcal{F}_n = \{\mathbb{I}_{(-\infty, t]}, t \leq Z_{N(n), n}\}$, où $Z_{i, n}$, $1 \leq i \leq n$, est la i -ème statistique d'ordre associée à l'échantillon de durées observées (Z_1, \dots, Z_n) , et $N(n) < n$ est une suite convenable. En utilisant essentiellement la loi du logarithme de Giné et Guillou (1999), on peut étendre les résultats du **chapitre 4** à la classe \mathcal{F}_n .

D'autre part, en considérant d'autres types de classes de fonctions, il est possible de traiter le cas de la densité conditionnelle. Des lois analogues portant sur les quantiles – par exemple la médiane – conditionnels en données censurées peuvent alors être déduites. Ces résultats étendent notamment ceux de Ould-Saïd et Cai (2005).

Enfin, en combinant les lois du logarithme obtenues pour les estimateurs de la fonction de répartition conditionnelle et pour les estimateurs de la densité conditionnelle (à la manière de ce qui a été fait dans le **chapitre 2**), on peut obtenir des lois du logarithme pour des estimateurs du taux de mortalité conditionnel en données censurées.

L'article relatif à ces différents travaux est actuellement en cours de rédaction (Maillot et Viallon (2006)).

A.5.2 Le cas « âge en échelle de temps »

Dans les études épidémiologiques (et notamment les *études de cohorte prospectives*), on cherche à établir le lien entre le développement d'une pathologie et divers *facteurs de risque*. Contrairement au cas de l'*essai thérapeutique*, les individus entrant dans ce type d'enquête ne deviennent pas à risque à l'inclusion, mais le sont depuis leur naissance. Ainsi, sélectionner l'inclusion dans l'étude comme origine des temps peut conduire à des résultats biaisés (voir par exemple Thiebaut et Bénichou (2004)). La bonne modélisation doit prendre la date de naissance comme origine des temps. On parle alors de modèles avec *âge en échelle de temps*. Ils induisent une *troncature à gauche*, en plus de la traditionnelle censure à droite. Leur étude théorique fait notamment intervenir les *processus de comptage* (Andersen *et al.* (1992)). L'élaboration d'un estimateur de la régression additive en données censurées avec l'âge en échelle de temps est actuellement à l'étude. Une des applications

directes de ce type d'estimateurs est la *construction de scores de risque de pathologie*, autrement dit l'estimation de la probabilité de développer une maladie étant donné un ensemble de facteurs de risque (voir §A.7 ci-dessous).

A.6 Modèles additifs et régression censurées

En complément des résultats du **chapitre 5**, où nous avons établi la vitesse de convergence uniforme presque sûre d'un estimateur de la fonction de régression censurée additive, divers travaux sont en cours. Nous avons notamment établi la vitesse de convergence en moyenne quadratique associée à cet estimateur (Debbbarh et Viallon (2006b)). Nous avons également établi la normalité asymptotique des estimateurs des composantes additives (Debbbarh et Viallon (2006a)). Ce dernier résultat entraîne naturellement la normalité asymptotique de l'estimateur intégré. Enfin, nous travaillons actuellement sur des lois du logarithme portant sur l'estimateur intégré (Debbbarh et Viallon (2006d)); ces lois sont similaires à celles présentées au **chapitre 4** pour l'estimateur de la fonction de régression multivariée en données censurées.

A.7 Constructions de scores de risques de pathologies

Les chercheurs, les médecins, mais aussi le grand public sont de plus en plus demandeurs d'outils statistiques permettant de prédire des risques absolus individuels de maladie. Dans cette optique, et en collaboration avec Stéphane Ragusa (L.S.T.A.), Françoise Clavel-Chepelon (Équipe ERI20-INSERM) et Jacques Bénichou (CHU Rouen), nous avons entrepris l'élaboration d'un score de risque de cancer du sein pour les femmes françaises (Viallon *et al.* (2006b)). Pour ce faire, nous avons appliqué un modèle de Cox, avec âge en échelle de temps, sur la cohorte E3N (qui est une étude prospective portant sur 100 000 femmes de la Mutuelle Générale de l'Éducation Nationale, suivies depuis 1990). Des scores traitant d'autres pathologies (cancers de l'ovaire, de l'endomètre, du colon, maladies cardio-vasculaires, ostéoporose, etc.) sont envisagés. De plus, d'autres modèles statistiques pourraient être utilisés pour construire ces scores, et notamment des modèles additifs, incorporant eux aussi l'âge comme échelle des temps.

Par ailleurs, nous avons développé une méthode simple permettant de tester la *calibration*² des scores de risques (Viallon *et al.* (2006a)). Différentes méthodes sont communément utilisées dans les études de validation de tels modèles. Cependant, elles présentent généralement un biais (dû principalement à la présence de censures). Dans ce travail, nous listons les méthodes existantes, soulignons leurs défauts respectifs de façon théorique, puis proposons notre méthode. Des simulations ainsi qu'un exemple de validation, sur la cohorte E3N, d'un score de risque américain de cancer du sein viennent enfin illustrer notre propos.

²La *calibration* est une mesure de type « *goodness-of-fit* ».

Bibliographie

- Akaike, H. (1954). An approximation to the density function. *Ann. Inst. Stat. Math.*, **6**, 127–132.
- Andersen, P., Borgan, O., Gill, R., et Keiding, N. (1992). *Statistical models based on counting processes*. Springer, New York.
- Ango-Nze, P. et Rios, R. (2000). Density estimation in L^∞ norm for mixing processes. *J. Statist. Plann. Inference*, **83**(1), 75–90.
- Bahadur, R. (1966). A note on quantiles in large samples. *Ann. Math. Stat.*, **37**, 577–580.
- Beran, R. (1981). Nonparametric regression with randomly censored data. In *Technical report*. Univ. California Press, Berkeley.
- Bosq, D. et Lecoutre, J. (1987). *Théorie de l'estimation fonctionnelle*. Economica, Paris.
- Breiman, L. et Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Am. Stat. Assoc.*, **80**, 580–619.
- Brunel, E. et Comte, F. (2006a). Adaptive nonparametric regression estimation in presence of right censoring. *Preprint. Available at <http://www.math-info.univ-paris5.fr/map5/publis/PUBLIS05/2005-16-rev.pdf>*.
- Brunel, E. et Comte, F. (2006b). Model selection for additive regression in presence of right censoring. *Preprint. Available at <http://www.math-info.univ-paris5.fr/map5/publis/PUBLIS06/2006-5.pdf>*.
- Camlong, C. (1999). Convergence presque sûre de l'estimateur à noyau d'une fonction de régression additive sous une hypothèse de mélange. *C. R. Acad. Sci. Paris Sér. I Math.*, **329**(1), 75–78.
- Cantelli, F. P. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giorn. Ist. Ital. Attuari*, **4**, 421–424.
- Carbonez, A., Györfi, L., et van der Meulen, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Statist. Decisions*, **13**(1), 21–37.
- Chen, R., Härdle, W., Linton, O., et Severance-Lossin, E. (1996). Nonparametric estimation of additive separable regression models. *Statistical Theory and Computational Aspects of Smoothing*, Physica Verlag, Berlin, pages 247–265.
- Chung, K.-L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans. Am. Math. Soc.*, **67**, 36–50.

- Collomb, G. (1981). Estimation non-paramétrique de la regression : Revue bibliographique. *Int. Stat. Rev.*, **49**, 75–93.
- Csörgő, M. (1983). *Quantile processes with statistical applications*. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.
- Csörgő, M. et Révész, P. (1978). Strong approximations of the quantile process. *Ann. Stat.*, **6**, 882–894.
- Csörgő, M. et Révész, P. (1981). *Strong Approximation in Probability and Statistics*. Academic Press, New York.
- Dabrowska, D. M. (1987). Nonparametric regression with censored survival data. *J. Scand. Statist.*, **14**, 181–197.
- Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Stat.*, **17**(3), 1157–1167.
- Dabrowska, D. M. (1995). Nonparametric regression with censored covariates. *J. Multivariate Anal.*, **54**(2), 253–283.
- Debbbarh, M. et Viallon, V. (2006a). Asymptotic normality of the additive regression model components under random censorship. *Submitted to J. Nonparametric Stat.*
- Debbbarh, M. et Viallon, V. (2006b). Mean square convergence for an estimator of the additive regression function under random censorship. *In revision in C. R., Math., Acad. Sci. Paris.*
- Debbbarh, M. et Viallon, V. (2006c). Uniform convergence for an estimator of the additive regression function under random censorship. *Submitted to C. R., Math., Acad. Sci. Paris (in French).*
- Debbbarh, M. et Viallon, V. (2006d). Uniform in bandwidth laws of the logarithm for an estimate of the additive regression function under random censorship. *In progress.*
- Deheuvels, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre des estimateurs de la densité. *C. R. Acad. Sci., Paris, Sér. A*, **278**, 1217–1220.
- Deheuvels, P. (1982). Strong limiting bounds for maximal uniform spacings. *Ann. Probab.*, **10**, 1058–1065.
- Deheuvels, P. (1991). Laws of the iterated logarithm for density estimators. In G. Roussas, editor, *Nonparametric Functional estimators and Related Topics*, NATO adv.Sci.Ins. C, pages 19–29. Kluwer Academic.
- Deheuvels, P. (1992). Functional laws of the iterated logarithm for large increments of empirical and quantile processes. *Stochastic Processes Appl.*, **43**(1), 133–163.
- Deheuvels, P. (2000). Chung type functional laws of the iterated logarithm for tail empirical processes. *Ann.I.H.P. B*, **36**, 583–616.
- Deheuvels, P. et Derzko, G. (2006). Nonparametric estimation of conditional lifetime distributions under random censorship. In *Advances in Statistical Methods for the Health Sciences : Applications to Cancer and AIDS Studies , Genome Sequence Analysis and Survival Analysis, to appear*. Springer, New York.

- Deheuvels, P. et Einmahl, J. (1996). On the strong limiting behavior of local functionals of empirical processes based upon censored data. *Ann. Prob.*, **24**, 504–525.
- Deheuvels, P. et Einmahl, J. (2000). Functional limit laws for the increments of kaplan-meier product-limit processes and applications. *Ann. Prob.*, **28**(7), 1301–1335.
- Deheuvels, P. et Mason, D. (1990). Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes. *Ann. Prob.*, **18**, 1693–1722.
- Deheuvels, P. et Mason, D. (1991). A tail empirical process approach to some nonstandard laws of the iterated logarithm. *J.Theor.Prob.*, **4**, 53–85.
- Deheuvels, P. et Mason, D. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Prob.*, **20**, 1248–1287.
- Deheuvels, P. et Mason, D. (1994). Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann.Prob.*, **22**, 1619–1661.
- Deheuvels, P. et Mason, D. (1995). Nonstandard local empirical processes indexed by sets. *J.Satist.Plann. Inf.*, **45**, 91–112.
- Deheuvels, P. et Mason, D. (2004). General confidence bounds for nonparametric functional estimators. *Stat. Inf. for Stoch. Proc.*, **7**, 225–277.
- Deheuvels, P. et Mason, D. (2006). Estimating conditional distributions under random censorship. *Private communication*.
- Delecroix, M., Lopez, O., et Patilea, V. (2006). Nonlinear censored regression using synthetic data. *Preprint. Available at <http://www.crest.fr/doctravail/document/2006-10.pdf>*.
- Derbort, S., Dette, H., et Munk, A. (2002). A test for additivity in nonparametric regression. *Ann. Inst. Statist. Math.*, **54**(1), 60–82.
- Deutsch, F. (1983). von Neumann’s alternating method : the rate of convergence. In *Approximation theory, IV (College Station, Tex., 1983)*, pages 427–434. Academic Press, New York.
- Devroye, L. et Györfi, L. (1985). *Nonparametric density estimation. The L_1 view*. Wiley, New York.
- Donsker, M. (1951). An invariance principle for certain probability theorems. *Mem. Amer. Math. Soc.*, **6**.
- Dudley, R. (1999). *Uniform central limit theorems*. Cambridge University Press.
- Efron, B. (1967). The two sample problem with censored data. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4**, pages 831–853. Univ. California Press, Berkeley.
- Eggermont, P. et La Riccia, V. (2001). *Maximum penalized likelihood estimation. Vol. 1 : Density estimation*. Springer, New York.
- Einmahl, U. et Mason, D. (2000). An empirical process approach to the uniform consistency of kernel type estimators. *Journ. Theoretic. Probab.*, **13**, 1–13.
- Einmahl, U. et Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.*, **33**(3), 1380–1403.

- Eubank, R., Hart, J., Simpson, D., et Stefanski, L. (1995). Testing for additivity in nonparametric regression. *Ann. Statist.*, **23**(6), 1896–1920.
- Fan, J. et Gijbels, I. (1994). Censored regression : Local linear approximations and their applications. *J. Am. Stat. Assoc.*, **89**(426), 560–570.
- Fan, J. et Gijbels, I. (1996). *Local polynomial modelling and its application*. Chapman and Hall, London.
- Finkelstein, H. (1971). The law of iterated logarithm for empirical distributions. *Ann. Math. Stat.*, **42**, 607–625.
- Földes, A. et Rejtő, L. (1981). A LIL type result for the product-limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**, 75–86.
- Friedman, J. H. et Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.*, **76**(376), 817–823.
- Giné, E. et Guillou, A. (1999). Laws of the iterated logarithm for censored data. *Ann. Proba*, **27**, 2042–2067.
- Glivenko, V. (1933). Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist. Ital. Attuari*, **4**, 92–99.
- Györfi, L., Kohler, M., Krzyzak, A., et Walk, H. (2002). *A distribution-free theory of nonparametric regression*. Springer, New York.
- Hajek, J., Sidak, Z., et Sen, P. (1999). *Theory of Rank Tests. 2nd ed.* Academic Press, Orlando, Florida.
- Hall, P. (1989). On projection pursuit regression. *Ann. Stat.*, **17**(2), 573–588.
- Hastie, T. et Tibshirani, R. (1986). Generalized additive models. With discussion. *Stat. Sci.*, **1**(3), 297–318.
- Hastie, T. et Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall.
- Härdle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.
- Härdle, W. et Hall, P. (1993). On the backfitting algorithm for additive regression models. *Stat. Neerl.*, **47**(1), 43–57.
- Jones, M., Davies, S., et Park, B. (1994). Versions of kernel-type regression estimators. *J. Am. Statist. Assoc.*, **89**, 825–832.
- Kaplan, E. et Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, **53**, 457–481.
- Kiefer, J. (1967). On Bahadur’s representation of sample quantiles. *Ann. Math. Stat.*, **38**, 1323–1342.
- Kiefer, J. (1972). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. In *Proc. 6th Berkeley Sympos. math. Statist. Probab.*, pages 227–244. Univ. California Press, Berkeley.

- Kim, W., Linton, O. B., et Hengartner, N. (1998). A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. *J. Comput. Graph. Statist.*, **8**(2), 278–297.
- Kohler, M., Máthé, K., et Pintér, M. (2002). Prediction from randomly right censored data. *J. Multivariate Anal.*, **80**(1), 73–100.
- Kohler, M., Kul, S., et Máthé, K. (2006). Least squares estimates for censored regression. *Preprint. Available at <http://www.mathematik.uni-stuttgart.de/mathA/lst3/kohler/hfm-pub-en.html>*.
- Linton, O. et Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**(1), 93–100.
- Linton, O. B., Chen, R., Wang, N., et Härdle, W. (1997). An analysis of transformations for additive nonparametric regression. *J. Am. Stat. Assoc.*, **92**(440), 1512–1521.
- Maillot, B. et Viallon, V. (2006). Uniform in bandwidth laws of the logarithm for functionals of the conditional law under random censorship. *In progress*.
- Mason, D. (1984). A strong limit theorem for the oscillation modulus of the uniform empirical quantile process. *Stochastic Process. Appl.*, **17**, 127–136.
- Mason, D. (2004). A uniform functional law of the iterated logarithm for the local empirical process. *Ann. Prob.*, **32**(2), 1391–1418.
- Mason, D. M., Shorack, G. R., et Wellner, J. A. (1983). Strong limit theorems for oscillation moduli of the uniform empirical process. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **65**, 83–97.
- Nadaraya, E. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141–142.
- Nadaraya, E. (1989). *Nonparametric estimation of probability densities and regression curves. Transl. from the Russian by Samuel Kotz*. Kluwer Academic Publishers, Dordrecht.
- Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, **10**(2), 233–253.
- Opsomer, J. D. (2000). Asymptotic properties of backfitting estimators. *J. Multivariate Anal.*, **73**(2), 166–179.
- Opsomer, J. D. et Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Stat.*, **25**(1), 186–211.
- Ould-Saïd, E. et Cai, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametric Stat.*, **17**(7), 797–806.
- Parzen, E. (1962). On estimation of probability density function and mode. *Ann. Math. Stat.*, **33**, 1065–1076.
- Parzen, E. (1979). Nonparametric statistical data modelling. *J. Amer. Statist. Assoc.*, **74**, 105–131.
- Pollard, D. (1984). *Convergence of stochastic processes*. Springer-Verlag, N.Y.

- Prakasa Rao, B. (1983). *Nonparametric functional estimation*. Academic Press, New York.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, **27**, 832–837.
- Roussas, G. e. (1990). *Nonparametric functional estimation and related topics. Proceedings of the NATO Advanced Study Institute held in Spetses, Greece, July 29-August 10, 1990*. NATO ASI Series 355, Kluwer Academic Publishers, Dordrecht.
- Satten, G. et Datta, S. (2001). The kaplan-meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.*, **55**, 207–210.
- Sauer, N. (1972). On the density of families of sets. *J. Comb. Theory, Ser. A*, **13**, 145–147.
- Scott, D. W. (1992). *Multivariate density estimation : theory, practice, and visualization*. Wiley, New York.
- Severance-Lossin, E. et Sperlich, S. (1999). Estimation of derivatives for additive separable models. *Statistics*, **33**(3), 241–265.
- Shorack, G. et Wellner, J. (1986). *Empirical Processes and applications to statistics*. Springer.
- Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- Sperlich, S., Tjøstheim, D., et Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econom. Theory*, **18**(2), 197–251.
- Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Stat.*, **13**, 689–705.
- Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Gebiete*, **3**, 221–226.
- Stute, W. (1982a). The law of the iterated logarithm for kernel density estimators. *Ann. Prob.*, **10**, 414–422.
- Stute, W. (1982b). The oscillation behavior of empirical processes. *Ann. Prob.*, **10**, 86–107.
- Stute, W. (1986a). Conditional empirical processes. *Ann. Prob.*, **14**(2), 638–647.
- Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Ann. Prob.*, **14**(3), 891–901.
- Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, **9**, 1089–1102.
- Susarla, V., Tsai, W., et Van Ryzin, J. (1984). A buckley-james-type estimator for the mean with censored data. *Biometrika*, **71**(3), 624–625.
- Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Prob.*, **22**, 28–76.
- Tanner, M. et Wong, W. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Stat.*, **11**, 422–439.

- Thiebaut, A. et Bénichou, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data : a simulation study. *Stat. Med.*, **23**(24), 3803–3820.
- Tjøstheim, D. et Auestad, B. H. (1994). Nonparametric identification of nonlinear time series : Projections. *J. Am. Stat. Assoc.*, **89**(428), 1398–1409.
- Tsai, W.-Y. et Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Ann. Stat.*, **13**, 1317–1334.
- Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Am. Stat. Assoc.*, **69**, 169–173.
- van der Laan, M. et Robins, J. (2002). *Unified methods for censored longitudinal data and causality*. Springer.
- van der Vaart, A. et Wellner, J. (1996). *Weak convergence and empirical processes*. Springer, New York.
- Vapnik, V. (1982). *Estimation of dependences*. Springer-Verlag, N.Y.
- Vapnik, V. et Āervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probab. Appl.*, **16**, 264–280.
- Vapnik, V. et Āervonenkis, A. Y. (1981). Necessary and sufficient conditions for uniform convergence of means to mathematical expectations. *Theor. Probab. Appl.*, **26**, 147–163.
- Varron, D. (2005). Uniformity in h in the functional limit law for the increments of the empirical process indexed by functions. (Uniformité en h dans la loi fonctionnelle limite uniforme pour les accroissements du processus empirique indexé par des fonctions). *C. R., Math., Acad. Sci. Paris*, **340**(6), 453–456.
- Varron, D. (2006). A bandwidth-uniform functional limit law for the increments of the empirical process. *Preprint*.
- Viallon, V. (2006a). A uniform in bandwidth functional limit law for the increments of the normed sample quantile process ; application to the nearest-neighbor empirical density function. *Submitted to J. Nonparametric Stat.*
- Viallon, V. (2006b). Uniform law of the logarithm for an estimate of the regression function under random censorship. *Submitted to C. R., Math., Acad. Sci. Paris (in French)*.
- Viallon, V., Bénichou, J., Clavel-Chapelon, F., et Ragusa, S. (2006a). How to evaluate the calibration of a risk prediction tool ; theory and examples. *In progress*.
- Viallon, V., Bénichou, J., Clavel-Chapelon, F., et Ragusa, S. (2006b). A new breast cancer risk prediction tool for french women ; construction and validation on the e3n cohort. *In progress*.
- Wand, M. et Jones, M. (1995). *Kernel smoothing*. Chapman and Hall, London.
- Watson, G. (1964). Smooth regression analysis. *Sankhya, Ser. A*, **26**, 359–372.
- Watson, G. et Leadbetter, M. (1964a). Hazard analysis i. *Biometrika*, **51**, 175–184.
- Watson, G. et Leadbetter, M. (1964b). Hazard analysis ii. *Sankhyā Ser. A*, **26**, 101–116.
- Winsberg, S. et Ramsay, J. (1980). Monotonic transformations to additivity using splines. *Biometrika*, **67**, 669–674.