



HAL
open science

Lois limites uniformes et estimation non-paramétrique de la régression

David Blondin

► **To cite this version:**

David Blondin. Lois limites uniformes et estimation non-paramétrique de la régression. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2004. Français. NNT: . tel-00011943

HAL Id: tel-00011943

<https://theses.hal.science/tel-00011943>

Submitted on 14 Mar 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.

Lois limites uniformes et estimation non-paramétrique de la régression

THÈSE

présentée et soutenue publiquement le 10 décembre 2004

pour l'obtention du

Doctorat de l'Université Paris 6
(spécialité statistique)

par

David Blondin

Composition du jury

<i>Directeur de thèse :</i>	Paul Deheuvels	Université Paris 6
<i>Président :</i>	Alain Berlinet	Université Montpellier 2
<i>Rapporteurs :</i>	Alain Berlinet Uwe Einmahl	Université Montpellier 2 Vrije Universiteit Brussel
<i>Examineurs :</i>	Michel Broniatowski Armelle Guillou Marc Hoffmann	Université Paris 6 Université Paris 6 Université Marne la Vallée

Mis en page avec la classe thloria.

Remerciements

Je veux tout d'abord remercier vivement le Professeur Paul Deheuvels qui a su diriger mes travaux de recherches vers des sujets passionnants, de la théorie des processus empiriques à la statistique fonctionnelle. Je tiens à lui exprimer ma plus profonde gratitude pour son soutien moral, ses encouragements et son extrême patience. Toujours positif, sa rigueur et son dynamisme furent les facteurs principaux conduisant à l'aboutissement de cette thèse.

Je suis reconnaissant aux professeurs Alain Berlinet et Uwe Einmahl qui ont accepté la tâche fastidieuse de rapporteur de thèse. Leurs commentaires et leurs questions m'ont permis de clarifier ma rédaction et m'ont donné de nouvelles pistes de réflexion. Je remercie également les membres du jury Michel Broniatowski, Armelle Guillou et Marc Hoffmann, de me faire l'honneur d'assister à ma soutenance.

Qu'il me soit permis de remercier toute l'équipe du LSTA pour leur soutien et leur gentillesse. Je remercie les doctorants de mon bureau qui m'ont accompagné et épaulé au cours de ces trois dernières années : Pierre Ribereau, sa disponibilité et sa générosité sans pareil, Anne Massiani et son exquise aménité, Jean-Baptiste Aubin et sa bonne humeur communicative. Je souhaite également saluer et encourager mes collègues, Samuela Leoni-Aubin, Davit Varron, Segolen Geffray, Driss Driouchi, Omar El Dakkak, Alexandre Depire, Myriam Maumy, Emmanuel Delafosse, Vivian Vallon ... Je tiens à exprimer ma sympathie à notre bibliothécaire émérite Pascal Epron qui m'a aidé dans le travail de consultation et de recherche d'ouvrages.

Mes pensées vont enfin à tous mes proches, ma famille et tout particulièrement Mitra, pour m'avoir aidé et supporté.

*Je dédie cette thèse
à mes parents et à mon frère*

Table des matières

Chapitre 1

Régression non-paramétrique par la méthode du noyau

1.1	Introduction	1
1.2	L'estimateur de Nadaraya-Watson	3
1.3	Consistance de l'estimateur [NW]	9
1.3.1	Calcul de la variance	9
1.3.2	Calcul du biais	12
1.4	Optimalité asymptotique et choix des paramètres	16
1.5	La validation croisée	21
1.6	Normalité asymptotique	24
1.7	Estimation par la méthode des polynômes locaux	26
1.7.1	Construction et définition des estimateurs localement polynomiaux	26
1.7.2	Biais et variance des estimateurs localement polynomiaux	28

Chapitre 2

Lois uniformes du logarithme pour les dérivées de la régression

2.1	Introduction	35
2.2	Le cadre univarié	36
2.3	Théorèmes	41
2.4	Démonstration des théorèmes	43
2.4.1	Borne supérieure	43
2.4.2	Borne inférieure	56
2.4.3	Démonstration du théorème 2.3.1	61
2.4.4	Démonstration des corollaires 2.3.1 et 2.3.2	61
2.4.5	Démonstration du théorème 2.3.2	61
2.4.6	Le cas non-borné	63

2.5	Généralisation multidimensionnelle	
	du théorème 2.3.2	69
2.5.1	Le cas où $X \in \mathbb{R}^p$	69
2.5.2	Le cas strictement multivarié : $\psi(Y) \in \mathbb{R}^d$	75
2.6	Lois limites presque sûres pour les estimateurs localement polynomiaux . .	82
2.7	Applications statistiques	87
2.7.1	Un critère simple de choix de fenêtre	
	pour la convergence uniforme presque sûre	87
2.7.2	Fenêtre adaptative et intervalles de confiance	88

Chapitre 3

Maximum de vraisemblance local et régression non-paramétrique
--

3.1	Introduction	93
3.2	Hypothèses de travail	97
3.3	Résultats	100
3.4	Extension multidimensionnelle	102
3.5	Démonstration	103

Annexe A	111
-----------------	------------

A.1	Processus empirique et estimation fonctionnelle non-paramétrique	111
A.2	Le lemme de Bochner	115
A.3	Inégalités exponentielles en dimension infinie	117
A.4	La loi du logarithme itéré multidimensionnelle	132
A.5	Continuité des fonctions $r_\psi(\cdot)$, $m_\psi(\cdot)$ et $\sigma_\psi^2(\cdot)$	135
A.6	Construction des noyaux d'ordre élevés	136
A.7	Remarque sur le terme de centrage	140

Bibliographie	141
----------------------	------------

Chapitre 1

Régression non-paramétrique par la méthode du noyau

1.1 Introduction

La théorie de l'estimation est une des branches les plus basiques de la statistique. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non-paramétrique. Le problème de l'estimation non-paramétrique consiste, dans la majeure partie des cas, à estimer, à partir des observations, une fonction inconnue, élément d'une certaine classe fonctionnelle. Rappelons qu'une procédure non-paramétrique est définie indépendamment de la distribution ou loi de l'échantillon d'observations. Plus particulièrement, on parle de méthode d'estimation non-paramétrique lorsque celle-ci ne se ramène pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon. Un des problèmes centraux en statistique est celui de l'estimation de caractéristiques fonctionnelles associées à la loi des observations, telles que, par exemple, la fonction de densité ou la fonction de régression (dans un modèle multivarié).

Un des modèles le plus fréquemment rencontré en statistique paramétrique ou non-paramétrique est le modèle de régression, dont nous donnons ci-dessous une description sommaire.

On dispose d'un échantillon, composé de n couples indépendants de variables aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$, et on dénote par (X, Y) un élément générique de cet échantillon. Dans le modèle de régression non-paramétrique, on suppose typiquement l'existence d'une fonction $m(\cdot)$ qui exprime la valeur moyenne de la *variable réponse* Y en fonction de la *variable d'entrée* X :

$$Y_i = m(X_i) + \epsilon_i, \quad \text{pour } 1 \leq i \leq n, \quad \text{avec } \epsilon_i \stackrel{d}{=} \epsilon \sim \mathcal{N}(\mu, \sigma^2). \quad (1.1)$$

L'erreur commise est, dans le cas classique, modélisée par une variable aléatoire gaussienne, qui sera généralement choisie indépendante des observations $\{X_i : 1 \leq i \leq n\}$, et de moyenne μ nulle. Cette dernière hypothèse simplifie considérablement les calculs et l'expression des propriétés asymptotiques liées à l'estimation de la fonction de régression, sous un tel modèle simplifié, ne sera pas considérée dans nos travaux. Nous considérons

le problème plus délicat posé par l'estimation de la fonction de régression, sans hypothèse particulière sur la loi du couple (X, Y) autre que celui de l'existence de $m(\cdot)$ (supposée suffisamment régulière), et de moments supérieurs d'ordre convenable de X et Y .

Il existe deux cas principaux pour le modèle (1.1), dépendants de la nature probabiliste des données $\{(X_i, Y_i) : 1 \leq i \leq n\}$. Le premier cas est le plus simple, et est appelé *dispositif expérimental à effets fixes* (ou “*fixed design*”). Il correspond à la situation où les $X_i = x_i$ sont fixés (c'est à dire, des constantes p.s., ou, de manière équivalente, déterministes ou dégénérées).

Exemple 1.1.1 Le dispositif expérimental régulier.

On suppose $X_i = x_i = i/n$ et $m(\cdot)$ une fonction de $[0, 1]$ dans \mathbb{R} telle que

$$Y_i = m(i/n) + \epsilon_i, \quad \text{pour } 1 \leq i \leq n.$$

Le deuxième cas, dit de *dispositif expérimental à effets aléatoires* (ou “*random design*”) désigne le modèle où les données $\{X_i : 1 \leq i \leq n\}$ sont strictement aléatoires (ou non-dégénérées). Nous étudierons essentiellement ce dernier modèle, qui est clairement plus général. Précisons également que seuls les modèles à observations indépendantes seront analysés, l'étude du cas de dépendance ne rajoutant que des difficultés de nature technique.

Nous allons présenter maintenant la fonction de régression de manière plus explicite, dans le cadre du modèle aléatoire univarié. Soit (X, Y) un couple de variables aléatoires réelles admettant une densité jointe sur \mathbb{R}^2 notée $f_{X,Y}$ et une densité marginale f_X . La variable Y est supposée intégrable, i.e. $\mathbb{E}[|Y|] < \infty$. Nous pouvons alors définir proprement la fonction de régression ou espérance conditionnelle de Y sachant $X = x$, par

$$m(x) := \mathbb{E}[Y|X = x] = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{\int_{\mathbb{R}} f_{X,Y}(x, y) dy} =: \frac{r(x)}{f_X(x)}, \quad (1.2)$$

lorsque la densité $f_X(x)$ est différente de zéro. Le problème de l'estimation de $m(\cdot)$ est du type non-paramétrique, i.e. la fonction de régression appartient à un ensemble non-paramétrique (infini-dimensionnel). Par exemple, nous pouvons supposer que $m(\cdot)$ appartient à la classe de fonctions \mathcal{F} constituée des fonctions continues sur $[0, 1]$ (cf. exemple 1.1.1 ci-dessus), lorsque le support de la densité est l'intervalle $[0, 1]$. Pour l'étude des propriétés minimax des estimateurs de la fonction de régression, les classes non-paramétriques de fonctions rencontrées sont de type Hölder, Sobolev ou Besov. La fonction de régression $m(x)$ définie ci-dessus en (1.2) réalise (pour tout x fixé) la meilleure approximation de Y sachant $X = x$, au sens des moindres carrés, en supposant Y de carré intégrable. Dans ce premier chapitre, nous discuterons de quelques méthodes de construction des estimateurs de la régression par la méthode du noyau. Puis, on concentrera nos travaux sur les propriétés statistiques des estimateurs (convergence, vitesse de convergence) ainsi que leur optimalité.

Les estimateurs que nous considérons appartiennent à la vaste classe des **estimateurs linéaires** (i.e. linéaires en tant que fonction des observations Y_i) :

Définition 1.1.1 Un estimateur $\hat{m}_n(x)$ de $m(x)$ est dit **estimateur linéaire** de la régression non-paramétrique si

$$\hat{m}_n(x) = \sum_{i=1}^n Y_i W_{ni}(x),$$

où la fonction de poids $W_{ni}(\cdot)$ ne dépend pas des observations Y_i .

La classe des estimateurs linéaires regroupe la majorité des estimateurs de la régression, c'est à dire les estimateurs par fonctions splines, par projection ou séries orthogonales, par ondelettes, et par la méthode du noyau. Dans la section suivante, nous présenterons le célèbre estimateur à noyau de la régression introduit par Nadaraya et Watson et quelques unes de ses propriétés essentielles. Nous nous intéresserons ensuite à l'optimalité asymptotique de cet estimateur, puis, à l'estimation localement polynomiale de la régression, qui constitue une des approches les plus performantes actuellement. Pour une revue bibliographique des travaux plus anciens concernant la régression non-paramétrique, nous citons les articles de Collomb [19] et Stone [129].

1.2 L'estimateur de Nadaraya-Watson

Supposons que l'on dispose d'un n -échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ de variables aléatoires à valeurs réelles, de même loi que le couple (X, Y) . On se propose de construire un estimateur $\hat{m}_n(x)$ de la fonction de régression à partir des couples d'observations $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. Le premier estimateur rencontré dans la littérature est l'estimateur à noyau de Nadaraya-Watson (cf. [108] et [149]), noté **estimateur [NW]**. Il est construit à partir d'une fonction **noyau** $K(\cdot)$ et d'une **fenêtre** h , de manière analogue à l'estimateur à noyau de la fonction de densité $f_X(\cdot)$ introduit par Parzen [111] et Rosenblatt [116], noté estimateur [PR]. On rappelle la définition de l'estimateur [PR],

$$\hat{f}_{X;n}(x) := \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}. \quad (1.3)$$

Dans un premier temps, nous désignons par fenêtre une suite $\{h_n : n \geq 1\}$ (possiblement aléatoire) de nombres strictement positifs vérifiant

$$h_n \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty.$$

La fenêtre $h = h_n$ dénote une suite indexée par $n = 1, 2, \dots$, mais la dépendance en n ne sera pas toujours précisée afin d'alléger les notations.

La fonction noyau $K : \mathbb{R} \rightarrow \mathbb{R}$ sera supposée mesurable et satisfaisant certaines hypothèses basiques parmi celles énoncées ci-dessous :

$$(K.1) \quad K \text{ est bornée, i.e. } \sup_{u \in \mathbb{R}} |K(u)| < \infty;$$

$$(K.2) \quad \lim_{|u| \rightarrow \infty} |u|K(u) = 0;$$

$$(K.3) \quad K(\cdot) \in L_1(\mathbb{R}), \text{ i.e. } \int_{\mathbb{R}} |K(u)| du < \infty;$$

$$(K.4) \quad \int_{\mathbb{R}} K(u) du = 1.$$

L'estimateur [NW] se présente sous la forme d'une moyenne locale pondérée des valeurs Y_i et est défini par,

$$\hat{m}_n^{NW}(x) := \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \times \mathbb{I}\left\{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \neq 0\right\}, \quad (1.4)$$

où $\mathbb{I}\{\cdot\} := \mathbb{I}_{\{\cdot\}}$ désigne la fonction indicatrice. On rappelle que, pour tout événement A Borel-mesurable,

$$\mathbb{I}(A) := \begin{cases} 1, & \text{si } A \text{ est vérifié,} \\ 0, & \text{sinon.} \end{cases}$$

De manière similaire, nous pouvons définir l'estimateur [NW] par,

$$\hat{m}_n^{NW}(x) := \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}, & \text{lorsque } \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i, & \text{sinon.} \end{cases} \quad (1.5)$$

Le noyau K détermine la forme du voisinage autour du point x et la fenêtre h contrôle la taille de ce voisinage, c'est à dire le nombre d'observations prises pour effectuer la moyenne locale. Intuitivement, il est naturel que la fenêtre h soit prépondérante pour la consistance de l'estimateur [NW]. Cette observation sera confirmée dans la prochaine section et dans le paragraphe suivant la remarque 1.2.1 ci-dessous.

En posant

$$\hat{r}_n(x) := \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right) \quad \{\text{estimateur à noyau de } r(x) \text{ (cf. (1.2))}\}, \quad (1.6)$$

nous remarquons que l'estimateur [NW] peut s'écrire $\hat{m}_n^{NW}(x) = \hat{r}_n(x) / \hat{f}_{X;n}(x)$. Cette dernière formulation est courante dans la littérature et consiste en une bonne première approche de l'estimateur [NW]. De facto, on traitera séparément le numérateur et le dénominateur aléatoires afin d'obtenir les propriétés asymptotiques usuelles de l'estimateur [NW], car il est difficile de travailler directement avec un quotient aléatoire. La méthode consiste

alors à linéariser la déviation $\hat{m}_n^{NW}(x) - m(x)$ en fonction de $\hat{f}_{X;n}(x) - f_X(x)$ et $\hat{r}_n(x) - r(x)$. Cette technique est centrale (voire systématique) en régression non-paramétrique, elle sera développée en détails dans les sections suivantes et particulièrement lors des démonstrations (cf. section 1.2).

- **Premières observations sur l'estimateur [NW] :**

L'estimateur [NW] (1.4) est bien linéaire au sens de la définition 1.1.1 avec comme fonction de poids $W_{ni}^{NW}(\cdot)$ définie par,

$$W_{ni}^{NW}(x) := \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \mathbb{I}\left\{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \neq 0\right\}.$$

Remarque 1.2.1 Pour une discussion plus générale sur la fonction de poids dans le cadre de la régression non-paramétrique et une exposition de certaines conditions nécessaires à sa consistance, nous citerons l'article pionnier de Stone (1977) [129]. Notons aussi que, en restreignant notre étude aux noyaux positifs (c'est à dire, tels que $K \geq 0$), la fonction indicatrice, présente dans (1.4), disparaît.

Parmi les deux paramètres K (fonctionnel) et h (numérique) à sélectionner, la fenêtre h détermine le degré de lissage de l'estimateur [NW]. Supposons que l'estimateur soit seulement évalué aux points d'observations $\{X_i : 1 \leq i \leq n\}$, alors, lorsque K est à support compact, nous obtenons

$$\lim_{h \rightarrow 0} \hat{m}_n^{NW}(X_i) = K(0)Y_i / K(0) = Y_i.$$

Plus précisément, nous avons

$$\lim_{h \rightarrow 0} \hat{m}_n^{NW}(x) = \begin{cases} Y_i, & \text{lorsque } x = X_i, \quad \forall 1 \leq i \leq n, \\ 0, & \text{sinon.} \end{cases}$$

Lorsque h tend vers zéro, l'estimateur [NW] a donc tendance à reproduire les données, la courbe obtenue est proche d'une interpolation des points $\{(X_i, Y_i) : 1 \leq i \leq n\}$. C'est un phénomène de sous-lissage, la variance de l'estimateur est trop grande. De l'autre côté,

$$\lim_{h \rightarrow \infty} \hat{m}_n^{NW}(x) = \frac{\sum_{i=1}^n K(0)Y_i}{\sum_{i=1}^n K(0)} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Lorsque h tend vers l'infini, nous avons un phénomène de surlissage, l'estimateur $\hat{m}_n^{NW}(x)$ tend vers $n^{-1} \sum_{i=1}^n Y_i$ qui est une fonction indépendante de x . L'erreur déterministe (ou

biais) est trop grande. Ce constat nous indique que les propriétés statistiques de l'estimateur [NW] dépendent bien de la fenêtre ou paramètre de lissage h , qu'il faudra choisir afin d'équilibrer le biais et la variance.

A présent, nous allons aborder une des multiples façons de construire l'estimateur de la fonction de régression introduit par Nadaraya et Watson. Pour une justification intuitive de l'estimateur [NW], rappelons la définition de l'estimateur à noyau de la densité bivariée, extension naturelle de (1.3),

$$\hat{f}_{X,Y;n}(x, y) := \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right). \quad (1.7)$$

En remplaçant dans (1.2) la densité jointe $f_{X,Y}$ et la densité marginale f_X par leurs estimateurs à noyaux [PR] respectifs, nous retrouvons l'estimateur [NW] défini en (1.4) ou (1.5). Il s'ensuit la proposition suivante.

Proposition 1.2.1 *Si le noyau K est symétrique (ou d'ordre 1), nous obtenons les égalités suivantes*

$$\hat{m}_n^{NW}(x) = \frac{\int_{\mathbf{R}} y \hat{f}_{X,Y;n}(x, y) dy}{\int_{\mathbf{R}} \hat{f}_{X,Y;n}(x, y) dy} = \int_{\mathbf{R}} y \hat{f}_{X,Y;n}(x, y) dy / \hat{f}_{X;n}(x). \quad (1.8)$$

D'après (1.7), nous avons

$$\begin{aligned} \int_{\mathbf{R}} \hat{f}_{X,Y;n}(x, y) dy &= \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \int_{\mathbf{R}} K\left(\frac{y - Y_i}{h}\right) dy \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \times \int_{\mathbf{R}} K(u) du = \hat{f}_{X;n}(x). \end{aligned}$$

De même,

$$\begin{aligned} \int_{\mathbf{R}} y \hat{f}_{X,Y;n}(x, y) dy &= \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \int_{\mathbf{R}} y K\left(\frac{y - Y_i}{h}\right) dy \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \times \left\{ \int_{\mathbf{R}} \left(\frac{y - Y_i}{h}\right) K\left(\frac{y - Y_i}{h}\right) dy \right. \\ &\quad \left. + \frac{Y_i}{h} \times \int_{\mathbf{R}} K\left(\frac{y - Y_i}{h}\right) dy \right\} \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \times \left\{ h \int_{\mathbf{R}} u K(u) du + Y_i \int_{\mathbf{R}} K(u) du \right\} \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i = \hat{r}_n(x), \end{aligned}$$

ce qui démontre (1.8). \square

La définition (1.7) nous conduit à introduire l'estimateur [NW] dans le cadre multivarié. Lorsque la variable explicative ou prédictive X est à valeurs dans \mathbb{R}^p , pour un certain $p \in \mathbb{N}$ fixé, les estimateurs [PR] et [NW] sont définis par,

$$\hat{f}_{X;n}(\mathbf{x}) := \frac{1}{nh^p} \sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right), \quad \mathbf{x} \in \mathbb{R}^p, \quad (1.9)$$

et

$$\hat{m}_n^{NW}(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i K\left(\frac{\mathbf{x} - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right)} \times \mathbb{I}\left\{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right) \neq 0\right\}. \quad (1.10)$$

Ci-dessus, $K : \mathbb{R}^p \rightarrow \mathbb{R}$ désigne une fonction multivariée définie comme le produit de noyaux univariés K_j (possiblement identiques pour $1, \dots, j$, cf. (1.7)), tels que

$$K(\mathbf{u}) = K(u_1, \dots, u_p) := \prod_{j=1}^p K_j(u_j), \quad \mathbf{u} \in \mathbb{R}^p.$$

Remarque 1.2.2 Soit

$$\mathcal{H} := \{\mathbf{h} = (h_1, \dots, h_p) : \min_{1 \leq j \leq p} h_j > 0\},$$

un sous-ensemble de \mathbb{R}^p correspondant à l'espace de toutes les fenêtres possibles. La définition (1.9) de l'estimateur à noyau [PR] de la densité est un cas particulier de l'estimateur suivant :

$$\frac{1}{n} \sum_{i=1}^n K_{\mathbf{h}}(\mathbf{x} - X_i), \quad (\text{cf. [33], chapitre 12}),$$

avec

$$K_{\mathbf{h}}(\mathbf{x}) = \prod_{j=1}^p \frac{1}{h_j} K_j\left(\frac{x_j}{h_j}\right).$$

Il est possible de présenter l'estimateur [PR] multivarié dans un contexte encore plus général. Soit \mathbf{H} une matrice $p \times p$ non-singulière (i.e. n'admettant pas de valeur propre nulle et donc inversible) appartenant à l'espace des matrices carrés $\mathcal{M}_p(\mathbb{R})$. On utilise un noyau multivarié $\mathbf{K} : \mathbb{R}^p \rightarrow \mathbb{R}$ qui satisfait les conditions suivantes :

$$\begin{aligned} \int_{\mathbb{R}^p} \mathbf{K}(\mathbf{u}) d\mathbf{u} &= 1, \\ \int_{\mathbb{R}^p} \mathbf{u} \mathbf{K}(\mathbf{u}) d\mathbf{u} &= 0 \quad \text{propriété de symétrie.} \end{aligned}$$

Alors, l'estimateur à noyau de la densité est défini, sous sa forme la plus générale, par

$$\hat{f}_{X;n}(\mathbf{x}) := \frac{1}{n|\mathbf{H}|} \sum_{i=1}^n \mathbf{K}\{\mathbf{H}^{-1}(\mathbf{x} - X_i)\}, \quad \mathbf{x} \in \mathbb{R}^p, \quad (1.11)$$

où $|\mathbf{H}|$ dénote le déterminant de la matrice \mathbf{H} . En reprenant les notations ci-dessus dans les définitions (1.9) et (1.10), la matrice fenêtre est de la forme $\mathbf{H} = h\mathbf{I}_p$, où \mathbf{I}_p désigne la matrice $p \times p$ identité. En d'autres termes, nous avons choisi dans chaque direction la même fenêtre $h = h_i$, $i = 1, \dots, p$. Le noyau \mathbf{K} peut être également à support sphérique, c'est à dire tel que

$$\mathbf{K}(\mathbf{u}) = W(\|\mathbf{u}\|_p),$$

où W dénote un noyau univarié à support compact et $\|\cdot\|_p$ est la norme Euclidienne sur \mathbb{R}^p . Par contre, lorsqu'on se base sur la définition (1.9), le support du noyau est plutôt de forme rectangulaire. On se réfère à Scott (1992), p. 152-155, [122], pour plus de détails sur l'estimation de la densité et de la régression dans le cadre multivarié.

-Estimateurs alternatifs

Le dénominateur aléatoire dans (1.4) est un inconvénient majeur, notamment pour l'étude des dérivées de l'estimateur [NW]. Dans le cadre du dispositif expérimental où les variables X_i sont ordonnées, Gasser et Müller (1979) [52] ont proposé l'estimateur suivant :

$$\hat{m}_n^{GM}(x) := \sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K\left(\frac{x-t}{h}\right) dt \times Y_i \right\}, \quad (1.12)$$

avec $s_i = (X_i + X_{i+1})/2$, $X_0 = -\infty$ et $X_{n+1} = +\infty$. Cet estimateur est bien linéaire au sens de la définition 1.1.1, avec une fonction de poids sans dénominateur et sommable à 1. D'après (1.12), la fonction de poids est définie par,

$$W_{ni}^{GM}(x) := \int_{s_{i-1}}^{s_i} K\left(\frac{x-t}{h}\right) dt.$$

L'estimateur [GM] de Gasser et Müller est une modification d'une version antérieure développée par Priestley et Chao (1972) [113]. Pour une étude complète de l'estimateur [GM], nous citons l'ouvrage de Müller (1988) [104].

Lorsque la fonction de densité marginale f_X est connue, il existe une version légèrement différente de l'estimateur [NW], proposée par Johnston (cf. [81] et [82]),

$$\hat{m}_n^J(x) := \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right) / f_X(x). \quad (1.13)$$

L'estimateur $\hat{m}_n^J(\cdot)$ se réfère également au dispositif expérimental à effets fixes car la fonction de densité f_X est connue. Le biais de l'estimateur \hat{m}_n^J est proche de l'estimateur [NW] (cf. proposition 1.3.4, sous-section 1.3.2, ci-après). En suivant Wand et Jones (1995), p. 152, [148], nous présentons l'estimateur

$$\hat{m}_n^*(x) := \frac{1}{nh} \sum_{i=1}^n f_X(X_i)^{-1} Y_i K\left(\frac{x-X_i}{h}\right), \quad (1.14)$$

qui a un meilleur biais que l'estimateur [NW] ou l'estimateur $\hat{m}_n^J(x)$. Le biais de l'estimateur défini en (1.14) est équivalent à celui de l'estimateur localement linéaire, défini ultérieurement dans la section 1.7.

La restriction de notre présentation des estimateurs de la régression à la méthode du noyau peut être excuser par la remarque suivante : deux autres classes importantes d'estimateurs, les splines et les plus proches voisins correspondent à des estimateurs à noyaux construits avec des fenêtres particulières, de la forme $f_X^{-\alpha}$, $0 \leq \alpha \leq 1$ (cf. Jennen-Steinmetz et Gasser (1988), [80], pour des références appropriées).

1.3 Consistance de l'estimateur [NW]

L'estimateur à noyau de la régression est donc dépendant du choix de deux paramètres, la fenêtre h et le noyau K . Nous verrons dans les sections suivantes que le paramètre crucial est la fenêtre pour obtenir de bonnes propriétés asymptotiques. Toutefois le noyau ne doit pas être négligé, il permet de réduire le biais de notre estimateur en s'appuyant sur les propriétés de régularité de la courbe de régression. Dans cette section, nous déterminerons les conditions sur la fenêtre et le noyau nécessaires à la consistance de l'estimateur [NW].

Nous obtenons la consistance des estimateurs du type [NW], via la décomposition biais-variance suivante,

$$\mathbb{E}\left[\{\hat{m}_n^{NW}(x) - m(x)\}^2\right] = \text{Var}[\hat{m}_n^{NW}(x)] + \left\{\mathbb{E}[\hat{m}_n^{NW}(x)] - m(x)\right\}^2. \quad (1.15)$$

On dénote par $\xrightarrow{L^2}$ (respectivement $\xrightarrow{\mathbf{P}}$) la convergence en norme L^2 (resp. en probabilité). Lorsque (1.15) tend vers zéro, il s'ensuit

$$\hat{m}_n^{NW}(x) \xrightarrow{L^2} m(x), \quad \text{ce qui implique,} \quad \hat{m}_n^{NW}(x) \xrightarrow{\mathbf{P}} m(x). \quad (1.16)$$

En vue de (1.16), une simple étude des critères de convergence vers zéro du biais et de la variance ci-dessus nous précisera les conditions nécessaires à la consistance de l'estimateur [NW]. On note également que la perte L^2 caractérisée ci-dessus est une mesure très pratique de la performance de notre estimateur, elle sera utilisée afin de déterminer les paramètres optimaux asymptotiquement (cf. section 1.4).

1.3.1 Calcul de la variance

Nous débutons l'étude de l'estimateur [NW] par le calcul de sa variance et son expression asymptotique. Le noyau K est supposé vérifier les hypothèses (K.1-4). On note que (K.1) et (K.3) impliquent le fait que $K(\cdot)$ soit de carré intégrable. Nous posons, par convenance,

$$\sigma^2(x) := \text{Var}[Y|X = x] =: \frac{1}{f_X(x)} \int y^2 f_{X,Y}(x, y) dy - \{m(x)\}^2,$$

lorsque cette expression est bien définie.

Proposition 1.3.1 *On suppose $\mathbb{E}[Y^2] < \infty$. A chaque point de continuité des fonctions $m(x)$, $f_X(x)$ et $\sigma^2(x)$, tel que $f_X(x) > 0$,*

$$\text{Var}[\hat{m}_n^{NW}(x)] = \frac{1}{nh} \times \left\{ \frac{\sigma^2(x)}{f_X(x)} \int_{\mathbf{R}} K^2(u) du \right\} (1 + o(1)), \quad (1.17)$$

où le terme $o(1)$ tend vers 0 lorsque $h \rightarrow 0$.

En utilisant le lemme de Bochner (cf. résultat A.2.1, situé en annexe), nous obtenons aisément

$$\begin{aligned}\text{Var}[\hat{f}_{X;n}(x)] &= \frac{1}{nh^2} \left\{ \mathbb{E} \left[K^2 \left(\frac{x-X}{h} \right) \right] - \mathbb{E} \left[K \left(\frac{x-X}{h} \right) \right]^2 \right\} \\ &= \frac{1}{nh} \left\{ \int_{\mathbf{R}} K^2(u) f_X(x-hu) du - h \left\{ \int_{\mathbf{R}} K(u) f_X(x-hu) du \right\}^2 \right\} \\ &= \frac{1}{nh} f_X(x) \int_{\mathbf{R}} K^2(u) du (1 + o(1)),\end{aligned}$$

lorsque $h \rightarrow 0$. Soit la fonction $s(x) := \int y^2 f_{X,Y}(x,y) dy$. Nous avons,

$$\begin{aligned}\text{Var}[\hat{r}_n(x)] &= \frac{1}{nh^2} \left\{ \mathbb{E} \left[Y^2 K^2 \left(\frac{x-X}{h} \right) \right] - \mathbb{E} \left[Y K \left(\frac{x-X}{h} \right) \right]^2 \right\} \\ &= \frac{1}{nh} \left\{ \int_{\mathbf{R}} K^2(u) s(x-hu) du - h \left\{ \int_{\mathbf{R}} K(u) r(x-hu) du \right\}^2 \right\} \\ &= \frac{1}{nh} s(x) \int_{\mathbf{R}} K^2(u) du (1 + o(1)).\end{aligned}$$

De même,

$$\mathbb{E} \left[\left\{ \hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)] \right\} \left\{ \hat{r}_n(x) - \mathbb{E}[\hat{r}_n(x)] \right\} \right] = \frac{1}{nh} r(x) \int_{\mathbf{R}} K^2(u) du (1 + o(1)).$$

Soit le vecteur

$$A_n(x) := \begin{pmatrix} \hat{f}_{X;n}(x) \\ \hat{r}_n(x) \end{pmatrix},$$

et $\Sigma[A_n(x)]$ sa matrice de variance covariance. Il s'ensuit

$$\Sigma[A_n(x)] = \frac{1}{nh} \begin{pmatrix} f_X(x) & r(x) \\ r(x) & s(x) \end{pmatrix} \int_{\mathbf{R}} K^2(u) du (1 + o(1)).$$

En remarquant que,

$$\begin{pmatrix} -\frac{r(x)}{\{f_X(x)\}^2} & \frac{1}{f_X(x)} \end{pmatrix} \begin{pmatrix} f_X(x) & r(x) \\ r(x) & s(x) \end{pmatrix} \begin{pmatrix} -\frac{r(x)}{\{f_X(x)\}^2} \\ \frac{1}{f_X(x)} \end{pmatrix} = \frac{s(x)}{\{f_X(x)\}^2} - \frac{\{r(x)\}^2}{\{f_X(x)\}^3},$$

on obtient alors,

$$\begin{aligned}\text{Var}[\hat{m}_n^{NW}(x)] &= \frac{1}{nh} \left\{ \frac{s(x)}{\{f_X(x)\}^2} - \frac{\{r(x)\}^2}{\{f_X(x)\}^3} \right\} \int_{\mathbf{R}} K^2(u) du (1 + o(1)) \\ &= \frac{1}{nh} \times \left\{ \frac{\sigma^2(x)}{f_X(x)} \int_{\mathbf{R}} K^2(u) du \right\} (1 + o(1)).\end{aligned}$$

□

Remarque 1.3.1 Dans l'expression asymptotique des termes de variance d'estimateurs à noyau, nous retrouvons invariablement la quantité :

$$\int_{\mathbb{R}} K^2(u)du = \|K\|_2^2. \quad (1.18)$$

Pour s'assurer de la finitude de cette intégrale, nous pouvons choisir la fonction noyau $K(\cdot)$ à variation bornée sur \mathbb{R} et à support compact, en remarquant que ces dernières hypothèses impliquent clairement (K.1–3). En vue d'optimalité asymptotique, la variance minimale sera obtenue en minimisant (1.18) suivant K dans une certaine classe de noyaux fixée. Pour une expression explicite des noyaux de variance minimale, nous citons l'article de Gasser, Müller et Mammitzsch (1985) [54]. Notons également que l'hypothèse $\mathbb{E}[Y^2] < \infty$ est indispensable, afin de s'assurer de l'existence de la variance conditionnelle $\sigma^2(x)$.

En conclusion, si la fenêtre h_n satisfait les conditions

$$h_n \rightarrow 0 \quad \text{et} \quad nh_n \rightarrow \infty \quad \text{lorsque} \quad n \rightarrow \infty,$$

la variance de l'estimateur [NW] tend vers zéro.

-Extension multidimensionnelle : $X \in \mathbb{R}^p$

Soient \mathbf{x} et \mathbf{u} des vecteurs de \mathbb{R}^p . La variance asymptotique a une expression similaire au cas univarié. On rappelle que

$$\hat{m}_n^{NW}(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\mathbf{x} - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right)} \times \mathbb{I}\left\{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right) \neq 0\right\},$$

où $K : \mathbb{R}^p \rightarrow \mathbb{R}$ fonction noyau, produit de noyaux univariés vérifiant (K.1–4).

Proposition 1.3.2 Nous supposons $\mathbb{E}[Y^2] < \infty$. Alors, à chaque point de continuité des fonctions $m(\mathbf{x})$, $f_X(\mathbf{x})$ et $\sigma^2(\mathbf{x})$, tel que $f_X(\mathbf{x}) > 0$, nous avons,

$$\text{Var}[\hat{m}_n^{NW}(\mathbf{x})] = \frac{1}{nh^p} \times \left\{ \frac{\sigma^2(\mathbf{x})}{f_X(\mathbf{x})} \int_{\mathbb{R}^p} K^2(\mathbf{u})d\mathbf{u} \right\} (1 + o(1)), \quad (1.19)$$

où le terme $o(1)$ tend vers 0 lorsque $h \rightarrow 0$.

Nous obtenons,

$$\text{Var}[\hat{f}_{X;n}(\mathbf{x})] = \frac{1}{nh^p} \left\{ f_X(\mathbf{x}) \int_{\mathbb{R}^p} K^2(\mathbf{u})d\mathbf{u} \right\} (1 + o(1)),$$

et

$$\text{Var}[\hat{r}_n(\mathbf{x})] = \frac{1}{nh^p} \left\{ s(\mathbf{x}) \int_{\mathbb{R}^p} K^2(\mathbf{u})d\mathbf{u} \right\} (1 + o(1)).$$

Le reste de la démonstration est similaire au cadre univarié et ne sera pas présenté par souci de concision. \square

Lorsque la fenêtre h_n satisfait

$$h_n \rightarrow 0 \quad \text{et} \quad nh_n^p \rightarrow \infty \quad \text{lorsque} \quad n \rightarrow \infty,$$

la variance de l'estimateur [NW] multivarié tend vers zéro.

1.3.2 Calcul du biais

Le traitement du biais est purement analytique et repose essentiellement sur le développement de Taylor. Il nous faut supposer certaines conditions de régularités sur les fonctions $m(\cdot)$ et $f_X(\cdot)$ qui détermineront l'ordre du biais asymptotique en fonction du paramètre de lissage h . L'estimateur [NW] se présente sous la forme d'un quotient aléatoire, c'est pourquoi on utilise généralement comme terme de centrage l'approximation suivante

$$\tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)] := \frac{\mathbb{E}[\hat{r}_n(x)]}{\mathbb{E}[\hat{f}_{X;n}(x)]}. \quad (1.20)$$

La formule (1.20) est plus facile à manipuler et permet notamment la linéarisation de la déviation $\tilde{d}_n(x) := \hat{m}_n^{NW}(x) - \tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)]$. Nous avons, par exemple,

$$\tilde{d}_n(x) = \{\hat{r}_n(x) - \mathbb{E}[\hat{r}_n(x)]\} \times \frac{1}{\mathbb{E}[\hat{f}_{X;n}(x)]} - \{\hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)]\} \times \frac{\hat{r}_n(x)\hat{f}_{X;n}(x)}{\mathbb{E}[\hat{f}_{X;n}(x)]}.$$

La proposition ci-dessous démontrée par Nadaraya (cf. p. 116-117, [109]) justifie le choix du terme de centrage (1.20).

Proposition 1.3.3 *Lorsque Y est bornée et $nh \rightarrow \infty$,*

$$\mathbb{E}[\hat{m}_n^{NW}(x)] = \tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)] + O((nh)^{-1}). \quad (1.21)$$

Lorsque $\mathbb{E}[Y^2] < \infty$ et $nh^2 \rightarrow \infty$,

$$\mathbb{E}[\hat{m}_n^{NW}(x)] = \tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)] + O((n^{1/2}h)^{-1}). \quad (1.22)$$

Nous utilisons l'identité suivante,

$$\frac{1}{\hat{f}_{X;n}(x)} = \frac{1}{\mathbb{E}[\hat{f}_{X;n}(x)]} - \frac{\hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)]}{\{\mathbb{E}[\hat{f}_{X;n}(x)]\}^2} + \frac{\{\hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)]\}^2}{\hat{f}_{X;n}(x)\{\mathbb{E}[\hat{f}_{X;n}(x)]\}^2}$$

On multiplie par $\hat{r}_n(x)$ des deux côtés, puis on passe à l'espérance,

$$\begin{aligned} \mathbb{E}[\hat{m}_n^{NW}(x)] &= \tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)] - \frac{\mathbb{E}\left[\{\hat{r}_n(x) - \mathbb{E}[\hat{r}_n(x)]\}\{\hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)]\}\right]}{\{\mathbb{E}[\hat{f}_{X;n}(x)]\}^2} \\ &\quad + \mathbb{E}\left[\frac{\hat{r}_n(x)\{\hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)]\}^2}{\hat{f}_{X;n}(x)\{\mathbb{E}[\hat{f}_{X;n}(x)]\}^2}\right] \\ &=: \tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)] + \frac{a_n(x) + b_n(x)}{\{\mathbb{E}[\hat{f}_{X;n}(x)]\}^2}. \end{aligned}$$

Soit $s(x) = \int_{\mathbf{R}} y^2 f_{X,Y}(x, y) dy$. Nous calculons la variance asymptotique de $\hat{r}_n(x)$ puis $\hat{f}_{X;n}(x)$, via le lemme de Bochner (cf. résultat A.2.1 en annexe),

$$\text{Var}[\hat{r}_n(x)] = \frac{1}{nh} \int_{\mathbf{R}} K^2(u) s(x - uh) du - \frac{1}{n} \left\{ \int_{\mathbf{R}} K(u) r(x - uh) du \right\}^2$$

$$\begin{aligned}
 &\approx \frac{1}{nh} s(x) \int_{\mathbf{R}} K^2(u) du. \\
 \text{Var}[\hat{f}_{X;n}(x)] &= \frac{1}{nh} \int_{\mathbf{R}} K^2(u) f_X(x - uh) du - \frac{1}{n} \left\{ \int_{\mathbf{R}} K(u) f_X(x - uh) du \right\}^2 \\
 &\approx \frac{1}{nh} f_X(x) \int_{\mathbf{R}} K^2(u) du.
 \end{aligned}$$

En utilisant l'inégalité de Cauchy-Schwartz combinée aux formules ci-dessus, on obtient

$$a_n(x) = O\left(\frac{1}{nh}\right) \quad (1.23)$$

Lorsque la variable Y est bornée, i.e. $|Y| \leq M$ pour une certaine constante M fixée, nous remarquons que l'estimateur [NW] est lui aussi naturellement borné,

$$\frac{\hat{r}_n(x)}{\hat{f}_{X;n}(x)} = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \leq \frac{\sum_{i=1}^n M \times K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} = M \quad (1.24)$$

Cette dernière inégalité (1.24) permet de borner $b_n(x)$,

$$\begin{aligned}
 b_n(x) &\leq M \times \mathbb{E} \left[\left\{ \hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)] \right\}^2 \right] \\
 &\approx \frac{M}{nh} f_X(x) \int_{\mathbf{R}} K^2(u) du = O\left(\frac{1}{nh}\right).
 \end{aligned} \quad (1.25)$$

Les relations (1.23) et (1.25) entraînent (1.21).

Lorsque $\mathbb{E}[Y^2] < \infty$, nous avons

$$\begin{aligned}
 b_n(x) &\leq \mathbb{E} \left[\max_{1 \leq i \leq n} |Y_i| \left\{ \hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)] \right\}^2 \right] \\
 &\leq \left\{ \sum_{i=1}^n Y_i^2 \right\}^{1/2} \times \left\{ \mathbb{E} \left[\left\{ \hat{f}_{X;n}(x) - \mathbb{E}[\hat{f}_{X;n}(x)] \right\}^4 \right] \right\}^{1/2} \\
 &= \sqrt{n} \left\{ \mathbb{E}[Y^2] \right\}^{1/2} \times O\left(\frac{1}{nh}\right) = O\left(\frac{1}{n^{1/2}h}\right).
 \end{aligned} \quad (1.26)$$

Les relations (1.23) et (1.26) impliquent (1.22), la démonstration est achevée. \square

Nous sommes maintenant prêts pour énoncer le biais asymptotique de l'estimateur [NW]. Nous supposons la variable Y bornée, de telle sorte que (1.21) soit vérifiée. Nous verrons que le biais de l'estimateur [NW], suivant les propriétés de régularité de la courbe de régression, est une fonctionnelle des dérivées de la régression.

Proposition 1.3.4 *Supposons que $m(\cdot)$ et $f_X(\cdot)$ sont de classe $C^2(\mathbb{R})$ et que le noyau K est d'ordre 2, i.e. tel que*

$$\int_{\mathbb{R}} K(u)du = 1, \quad \int_{\mathbb{R}} uK(u)du = 0 \quad \text{et} \quad \int_{\mathbb{R}} u^2K(u)du < \infty.$$

Nous avons alors, lorsque $h \rightarrow 0$ et $nh \rightarrow \infty$,

$$\mathbb{E}[\hat{m}_n^{NW}(x)] - m(x) = \frac{h^2}{2} \times \left\{ \left\{ m''(x) + 2m'(x) \frac{f'_X(x)}{f_X(x)} \right\} \int_{\mathbb{R}} u^2K(u)du \right\} (1 + o(1)). \quad (1.27)$$

Remarque 1.3.2 Notons que le terme $o(1)$ dans (1.27) ci-dessus se décompose comme suit $\{O(h) + O((nh)^{-1})\}$, d'après (1.21).

$$\begin{aligned} \tilde{\mathbb{E}}[\hat{m}_n^{NW}(x)] - m(x) &= \left\{ \mathbb{E}[K((x - X)/h)] \right\}^{-1} \left\{ \int \frac{1}{h} K\left(\frac{x - t}{h}\right) r(t) dt - r(x) \right. \\ &\quad \left. + r(x) - m(x) \int \frac{1}{h} K\left(\frac{x - t}{h}\right) f_X(t) dt \right\} \\ &\approx \frac{h^2}{2} \times \{f_X(x)\}^{-1} \times \{r''(x) - m(x)f''_X(x)\} \times \int_{\mathbb{R}} u^2K(u)du \\ &= \frac{h^2}{2} \times \left\{ m''(x) + 2m'(x) \frac{f'_X(x)}{f_X(x)} \right\} \times \int_{\mathbb{R}} u^2K(u)du. \end{aligned} \quad (1.28)$$

Le signe \approx ci-dessus dénote une erreur de l'ordre $O(h)$ ou $o(1)$ d'après le lemme de Bochner. La proposition 1.3.3 et (1.28) impliquent (1.27). \square

Le terme de biais asymptotique fait apparaître la dérivée des fonctions $m(\cdot)$ et $f_X(\cdot)$. Ceci est dû au fait que l'estimateur [NW] réalise une approximation des moindres carrés localement constante des valeurs Y_i (cf. section 1.7). L'estimateur [NW] souffre donc d'un biais élevé dans la région où la dérivée de la vraie fonction de régression est grande. Le biais peut également être grand lorsque $f'_X(x)/f_X(x)$ est grand. En comparaison, sous des hypothèses similaire à celles de la proposition 1.3.4, l'estimateur [GM] a un meilleur biais :

$$\mathbb{E}[\hat{m}_n^{GM}(x)] - m(x) = \frac{h^2}{2} \times \left\{ m''(x) \times \int_{\mathbb{R}} u^2K(u)du \right\} (1 + o(1)). \quad (1.29)$$

La forme du biais asymptotique ci-dessus est préférable d'un point de vue statistique, car elle ne dépend pas de la densité f_X et de sa dérivée. Par exemple, si la courbe de régression est une droite, le terme de biais principal disparaît quelque soit la forme de la densité marginale f_X . De nombreuses techniques ont été développées dans la dernière décennie pour remédier à ce mauvais biais de l'estimateur [NW], nous citons les articles de Müller et Song (1993) [107], Linton et Nielsen (1994) [93] Mammen et Marron [96], Müller (1997) [105], Choi, Hall et Rousson (2000) [15] et Hall et Müller (2003) [64], pour une exposition des différentes méthodologies existantes.

Lorsque la fonction de régression admet des conditions de régularité supplémentaires, il est possible de réduire le biais asymptotique de l'estimateur [NW] en utilisant un **noyau d'ordre supérieur**. Soit q un entier naturel fixé.

Définition 1.3.1 Le noyau K est appelé **noyau d'ordre q** si il vérifie les conditions suivantes :

$$\int_{\mathbb{R}} K(u)du = 1, \quad \int_{\mathbb{R}} u^j K(u)du = 0, \quad j = 1, \dots, q-1, \quad \text{et} \quad \int_{\mathbb{R}} u^q K(u)du < \infty.$$

Pour illustrer l'utilité des noyaux d'ordre supérieurs, nous considérons l'exemple simple de l'estimation de la densité. Le biais de l'estimateur à noyau de la densité s'écrit

$$\mathbb{E}[\hat{f}_{X;n}(x)] - f_X(x) = \int \{f_X(x - hu) - f_X(x)\}K(u)du.$$

A présent, supposons que la densité $f_X(x)$ admet des dérivées bornées jusqu'à l'ordre q dans un voisinage du point x . Nous obtenons donc, via le développement de Taylor,

$$\mathbb{E}[\hat{f}_{X;n}(x)] - f_X(x) = \sum_{k=1}^{q-1} \left\{ h^k \frac{(-1)^k}{k!} f_X^{(k)}(x) \int u^k K(u)du \right\} + O(h^q). \quad (1.30)$$

La formule (1.30) ci-dessus montre clairement l'importance des noyaux dont les premiers moments sont nuls : un noyau d'ordre q permet de réduire le biais à l'ordre $O(h^q)$ modulo quelques hypothèses de régularité.

Dans le cadre multivarié, nous avons les conditions d'orthogonalité suivantes,

$$\int_{\mathbb{R}^p} \left\{ \prod_{i=1}^n u_i^{s_i} \right\} \times K(u_1, \dots, u_p) du_1 \dots du_p = 0, \quad \text{lorsque} \quad \sum_{i=1}^n s_i = 1, 2, \dots, q-1. \quad (1.31)$$

Si (1.31) est vérifiée ainsi que

$$\int_{\mathbb{R}^p} \|\mathbf{u}\|^q |K(\mathbf{u})| d\mathbf{u} < \infty,$$

le noyau multivarié $K(\cdot)$ est appelé noyau multivarié d'ordre q , c'est à dire tous ses moments jusqu'à l'ordre $q-1$ sont nuls.

Par convenance, nous dénotons par $[\mu_j(K)]$ le moment d'ordre j associé à la fonction noyau $K(\cdot)$, lorsque $j \in \mathbb{N}$.

Proposition 1.3.5 *Supposons que $m(\cdot)$ et $f_X(\cdot)$ sont de classe $C^q(\mathbb{R})$ et que le noyau K est d'ordre q , c'est à dire tel que*

$$[\mu_0(K)] = 1, \quad [\mu_j(K)] = 0, \quad 1 \leq j \leq q-1, \quad \text{et} \quad [\mu_q(K)] < \infty.$$

Lorsque $h \rightarrow 0$ et $nh \rightarrow \infty$, nous avons

$$\mathbb{E}[\hat{m}_n^{NW}(x)] - m(x) = \frac{h^q}{q!} \times \left\{ \left\{ m^{(q)}(x) + q \times m^{(q-1)}(x) \frac{f'_X(x)}{f_X(x)} \right\} [\mu_q(K)] \right\} (1 + o(1)). \quad (1.32)$$

Dans un premier temps, on considère l'espérance de $\hat{r}_n(x)$:

$$\begin{aligned}\mathbb{E}[\hat{r}_n(x)] &= \int_{\mathbf{R}} \int_{\mathbf{R}} yK\left(\frac{x-t}{h}\right) f_{X,Y}(t,y) dt dy = \int_{\mathbf{R}} K\left(\frac{x-t}{h}\right) r(t) dt \\ &= \int_{\mathbf{R}} K(u)r(x-hu) du = r(x) + \frac{h^q}{q!} r^{(q)}(x) \times [\mu_q(K)](1 + o(1)).\end{aligned}$$

Puis,

$$\mathbb{E}[\hat{f}_{X;n}(x)] = f_X(x) + \frac{h^q}{q!} f_X^{(q)}(x) \times [\mu_q(K)](1 + o(1)).$$

Le reste de la démonstration est similaire à la démonstration de la proposition 1.3.4 et ne sera pas présenté par souci de concision. \square

-Extension multidimensionnelle : $X \in \mathbb{R}^p$

On précise quelques notations nécessaires à la présentation du biais asymptotique dans le cadre multivarié. Soit $f : \mathbb{R}^p \rightarrow \mathbb{R}$ une fonction multivariée quelconque. Nous désignons par Q l'opérateur sur f défini par,

$$Q[f](x) := \int_{\mathbf{R}^p} [\mathbf{u}^T (\nabla^2 f(\mathbf{x})) \mathbf{u}] K(\mathbf{u}) d\mathbf{u},$$

où $\nabla^2 f(\mathbf{x})$ dénote la matrice Hessienne des dérivées partielles d'ordre 2 de la fonction $f(\cdot)$ au point \mathbf{x} .

Proposition 1.3.6 *Lorsque Y bornée et $nh^p \rightarrow \infty$,*

$$\mathbb{E}[\hat{m}_n^{NW}(\mathbf{x})] = \tilde{\mathbb{E}}[\hat{m}_n^{NW}(\mathbf{x})] + O((nh^p)^{-1}). \quad (1.33)$$

Supposons que $m(\cdot)$ et $f_X(\cdot)$ sont de classe $C^2(\mathbb{R}^p)$ et que le noyau K est d'ordre 2. Nous avons alors, lorsque $h \rightarrow 0$ et $nh^p \rightarrow \infty$,

$$\mathbb{E}[\hat{m}_n^{NW}(\mathbf{x})] - m(\mathbf{x}) = \frac{h^2}{2} \left\{ \frac{Q[r](\mathbf{x}) - m(\mathbf{x})Q[f_X](\mathbf{x})}{f_X(\mathbf{x})} \right\} (1 + o(1)). \quad (1.34)$$

On peut également formuler le biais asymptotique (1.34) de manière plus explicite mais moins compacte :

$$\frac{h^2}{2} \left\{ \sum_{j=1}^p \left\{ \frac{\partial^2}{\partial x_j^2} m(\mathbf{x}) + 2 \left\{ \frac{\partial}{\partial x_j} m(\mathbf{x}) \right\} \left\{ \frac{\partial}{\partial x_j} f_X(\mathbf{x}) \right\} \frac{1}{f_X(\mathbf{x})} \right\} \times \int_{\mathbf{R}^p} u_j^2 K(\mathbf{u}) d\mathbf{u} \right\}. \quad (1.35)$$

1.4 Optimalité asymptotique et choix des paramètres

Dans la section 1.2, nous avons établi les conditions nécessaires et suffisantes sur la fenêtre h_n pour obtenir la consistance de l'estimateur [NW] :

$$h_n \rightarrow 0 \quad \text{et} \quad nh_n \rightarrow \infty \quad \text{lorsque} \quad n \rightarrow \infty.$$

On se propose à présent de déterminer la fenêtre optimale, au sens d'un certain critère d'efficacité asymptotique. Nous chercherons la fenêtre qui minimise la perte L_2 associée à l'estimateur [NW] en fixant le noyau K dans une certaine classe. Puis, on s'intéressera à l'optimalité du noyau.

On désigne par $\mathcal{K}[q]$ la classe des noyaux d'ordre q à support compact et bornés. Nous supposons, tout au long de cette section, que le noyau $K \in \mathcal{K}[q]$. L'hypothèse K borné et à support compact est très classique en régression non-paramétrique, elle implique notamment l'intégrabilité des divers moments de la fonction noyau $K(\cdot)$.

Sous les hypothèses de la proposition 1.3.5, nous avons,

$$\begin{aligned} \mathbb{E}[\hat{m}_n^{NW}(x)] - m(x) &= \frac{h^q}{q!} \times \left\{ \left\{ m^{(q)}(x) + q \times m^{(q-1)}(x) \frac{f'_X(x)}{f_X(x)} \right\} [\mu_q(K)] \right\} (1 + o(1)) \\ &=: \frac{h^q}{q!} \times [b(x; q)] (1 + o(1)). \end{aligned} \quad (1.36)$$

Sous les hypothèses de la proposition 1.3.1, via (1.17), il s'ensuit

$$\begin{aligned} \text{Var}[\hat{m}_n^{NW}(x)] &= \frac{1}{nh} \times \left\{ \frac{\sigma^2(x)}{f_X(x)} [\mu_0(K^2)] \right\} (1 + o(1)) \\ &=: \frac{1}{nh} \times [v^2(x)] (1 + o(1)). \end{aligned} \quad (1.37)$$

Ces développements asymptotiques sont récurrents en optimisation asymptotique, car la fenêtre optimale équilibre le biais et la variance. On distingue essentiellement deux types de procédures pour la sélection du paramètre de lissage : l'approche locale et l'approche globale. En vue de résultats ponctuels ou uniformes, nous choisirons la procédure adéquate, c'est à dire l'approche locale pour les résultats de type convergence ponctuelle et l'approche globale pour les résultats de type convergence uniforme.

-critère de sélection local : AMSE

Nous considérons comme critère d'efficacité la célèbre **erreur quadratique moyenne** ou MSE (“*mean squared error*”). D'après les formules (1.36) et (1.37), nous pouvons présenter le théorème spécifiant le comportement asymptotique exact du risque quadratique de l'estimateur [NW] \hat{m}_n^{NW} au point x .

Théorème 1.4.1 *Sous les hypothèses des propositions 1.3.5 et 1.3.1, nous obtenons,*

$$\begin{aligned} [\text{MSE}]\{\hat{m}_n^{NW}(x)\} &:= \mathbb{E}\left[\{\hat{m}_n^{NW}(x) - m(x)\}^2\right] \\ &= \left\{ \frac{h^{2q}}{(q!)^2} \times \left\{ [b(x; q)] \right\}^2 + \frac{1}{nh} \times [v^2(x)] \right\} (1 + o(1)). \end{aligned} \quad (1.38)$$

D'après (1.36) et (1.37),

$$[\text{MSE}]\{\hat{m}_n^{NW}(x)\} = \left\{ \mathbb{E}[\hat{m}_n^{NW}(x) - m(x)]^2 + \text{Var}[\hat{m}_n^{NW}(x)] \right\}$$

$$= \frac{h^{2q}}{(q!)^2} \times \left\{ [b(x; q)] \right\}^2 (1 + o(1)) + \frac{1}{nh} \times [v^2(x)] (1 + o(1)).$$

□

D'après le théorème 1.4.1 et la formule (1.38), nous obtenons l'expression de l'erreur quadratique moyenne asymptotique ou AMSE ("asymptotic mean squared error") :

$$[\text{AMSE}]\{\hat{m}_n^{\text{NW}}(x)\} = \frac{h^{2q}}{(q!)^2} \times \left\{ [b(x; q)] \right\}^2 + \frac{1}{nh} \times [v^2(x)] =: [\text{AMSE}](h, K). \quad (1.39)$$

Notons que le risque quadratique asymptotique (1.39) dépend du noyau K et de la fenêtre h associés à l'estimateur [NW]. Nous supposons, dans un premier temps, le noyau K fixé. La fenêtre optimale, au sens du critère local de minimisation de l'AMSE au point x , est alors obtenue en minimisant suivant h la quantité (1.39), c'est à dire

$$h_{n,\text{opt}}^{\text{MSE}}(x) = h^{\text{MSE}}(K) = \arg \min_h [\text{AMSE}](h, K).$$

La fenêtre $h^{\text{MSE}}(K)$ est solution de l'équation suivante :

$$\frac{2q}{(q!)^2} h^{2q-1} \times \left\{ [b(x; q)] \right\}^2 - \frac{1}{nh^2} \times [v^2(x)] = 0.$$

Lorsque $[b(x; q)] \neq 0$, nous obtenons

$$\begin{aligned} h^{\text{MSE}}(K) &= n^{-1/(2q+1)} \left\{ \frac{(q!)^2 [v^2(x)]}{2q \{ [b(x; q)] \}^2} \right\}^{1/(2q+1)} \\ &= n^{-1/(2q+1)} \left\{ \frac{q!(q-1)! \left\{ \frac{\sigma^2(x)}{f_X(x)} [\mu_0(K^2)] \right\}}{2 \left\{ m^{(q)}(x) + q \times m^{(q-1)}(x) \frac{f'_X(x)}{f_X(x)} \right\}^2 [\mu_q(K)]^2} \right\}^{1/(2q+1)} \end{aligned} \quad (1.40)$$

La fenêtre $h^{\text{MSE}}(K)$ minimise donc asymptotiquement la MSE de l'estimateur [NW] au point x (critère local). Après calculs, il s'ensuit

$$\begin{aligned} \min_h [\text{AMSE}](h, K) &= \left\{ (q!)^{-2(q+1)/2q+1} \left\{ \frac{(q-1)!}{2} \right\}^{2q/2q+1} + \left\{ \frac{q!(q-1)!}{2} \right\}^{-1/2q+1} \right\} \times \\ &\quad \left\{ [v^2(x)] \right\}^{2q/2q+1} \left| [b(x; q)] \right|^{2/2q+1} n^{-2q/2q+1}. \end{aligned}$$

Pour simplifier notre écriture, on peut considérer le cas particulier $q = 2$, qui correspond au cadre d'étude où le noyau est positif ou d'ordre 2. D'après (1.40), lorsque $q = 2$,

$$h^{\text{MSE}}(K) = n^{-1/5} \left\{ \frac{\left\{ \frac{\sigma^2(x)}{f_X(x)} [\mu_0(K^2)] \right\}}{\left\{ m''(x) + 2 \times m'(x) \frac{f'_X(x)}{f_X(x)} \right\}^2 [\mu_2(K)]^2} \right\}^{1/5}.$$

Nous obtenons, en conséquence,

$$\min_h [\text{AMSE}](h, K) = \frac{5}{4} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\}^{4/5} \left| m''(x) + 2 \times m'(x) \frac{f'_X(x)}{f_X(x)} \right|^{2/5} [\mu_0(K^2)]^{4/5} [\mu_2(K)]^{2/5} n^{4/5}.$$

Par convenance, nous introduisons les notations :

$$G[q] := \left\{ (q!)^{-2(q+1)/2q+1} \left\{ \frac{(q-1)!}{2} \right\}^{2q/2q+1} + \left\{ \frac{q!(q-1)!}{2} \right\}^{-1/2q+1} \right\},$$

$$C[K, q] := [\mu_0(K^2)]^{2q/(2q+1)} [\mu_q(K)]^{2/(2q+1)}.$$

Il s'ensuit le corollaire suivant.

Corollaire 1.4.1 *On suppose les hypothèses du Théorème 1.4.1 vérifiées. Nous avons, si $\hat{m}_n^{NW}(x)$ est construit avec la fenêtre $h = h^{\text{MSE}}(K)$ (estimateur oracle),*

$$\lim_{n \rightarrow \infty} n^{2q/2q+1} \mathbb{E} \left[\left\{ \hat{m}_n^{NW}(x) - m(x) \right\}^2 \right] = G[q] \left| m^{(q)}(x) + q \times m^{(q-1)}(x) \frac{f'_X(x)}{f_X(x)} \right|^{2/2q+1} \times \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\}^{2q/2q+1} C[K, q].$$

Remarque 1.4.1 La fonction aléatoire $\hat{m}_n^{NW}(\cdot)$ définie en (1.5) avec la fenêtre $h^{\text{MSE}}(K)$ n'est plus un estimateur, stricto sensu, car elle dépend de la fonction de régression à estimer. Ce type de fonction est appelée pseudo-estimateur ou estimateur oracle dans la littérature. Le corollaire ci-dessus n'a donc aucun intérêt en pratique car il ne permet pas de construire un estimateur. Il est possible toutefois de remplacer les quantités inconnues par des estimateurs préliminaires consistants. Cette procédure, dite **plug-in**, conduit à des algorithmes itératifs tels le plug-in itéré (voir, Biau [10]). Pour d'autres procédures conduisant au choix de la fenêtre dans le cadre de l'estimation de la densité f_X , nous citons l'ouvrage de Eggermont et LaRiccia (2001), chapitre 7, [34].

La fenêtre optimale $h^{\text{MSE}}(K)$ permet de déterminer la vitesse de convergence optimale du risque quadratique (proche de $1/n$) lorsque le noyau est fixé dans la classe de fonctions $\mathcal{K}[q]$. On s'intéresse à présent à l'optimalité du noyau sur $\mathcal{K}[q]$. Il faut remarquer que le choix du noyau n'a d'impact que sur la constante limite, par l'intermédiaire de $C[K, q]$. Le problème du choix optimal du noyau K se résume ainsi :

$$K_{opt}^{\text{MSE}} := \arg \min_{K \in \mathcal{K}[q]} \left\{ [\mu_0(K^2)]^{2q/(2q+1)} [\mu_q(K)]^{2/(2q+1)} \right\}. \quad (1.41)$$

On note que le noyau d'Epanechnikov (ou Bartlett-Epanechnikov, [6] et [44]) est solution de la problématique (1.41) lorsque $q = 2$ et le support du noyau $[-1, 1]$. On rappelle la définition du noyau d'Epanechnikov,

$$K^E(u) := \frac{3}{4} \{1 - u^2\} \mathbb{I}\{|u| \leq 1\},$$

qui fournit la valeur minimale $C[K^E, 2] = 3^{4/5}5^{-6/5}$. Nous pouvons alors donner l'expression de la fenêtre optimale correspondante :

$$h^{\text{MSE}}(K^E) = n^{-1/5} \left\{ \frac{15 \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\}}{\left\{ m''(x) + 2 \times m'(x) \frac{f'_X(x)}{f_X(x)} \right\}^2} \right\}^{1/5}.$$

Pour d'autres développements et perspectives autour de l'optimalité des noyaux d'ordre élevés, on cite les travaux de Granovsky, Müller et Pfeifer (1995) [59] ainsi que l'article récent de Mammitzsch (2001) [97].

-critère de sélection global : AMISE

A présent, on s'intéresse à l'estimation de la fonction de régression sur un intervalle $I \subseteq \mathbb{R}$ et au risque global de l'estimateur [NW] sur cet intervalle. On introduit pour cela l'**erreur quadratique intégrée moyenne** ou MISE (“*mean integrated squared error*”),

$$\begin{aligned} [\text{MISE}]\{\hat{m}_n^{NW}(x)\} &:= \mathbb{E} \left[\int_I \{\hat{m}_n^{NW}(x) - m(x)\}^2 dx \right] \\ &= \int_I [\text{MSE}]\{\hat{m}_n^{NW}(x)\} dx \\ &= \int_I \mathbb{E} \left[\{\hat{m}_n^{NW}(x) - m(x)\}^2 \right] dx, \end{aligned}$$

d'après le théorème de Tonelli-Fubini.

Théorème 1.4.2 *Supposons les hypothèses des propositions 1.3.5 et 1.3.1.*

$$[\text{MISE}]\{\hat{m}_n^{NW}(x)\} = \left\{ \frac{h^{2q}}{(q!)^2} \int_I \{[b(x; q)]\}^2 dx + \frac{1}{nh} \int_I [v^2(x)] dx \right\} (1 + o(1)). \quad (1.42)$$

La fenêtre optimale, au sens du critère global de minimisation de l'AMISE (“*asymptotic mean integrated squared error*”) sur l'intervalle I , est donnée par,

$$h_{n, \text{opt}}^{\text{MISE}}(x) = n^{-1/(2q+1)} \left\{ \frac{q!(q-1)! \int_I [v^2(x)] dx}{2 \int_I \{[b(x; q)]\}^2 dx} \right\}^{1/(2q+1)}. \quad (1.43)$$

De nouveau, la fenêtre optimale dépend de paramètres inconnus et n'est donc pas utilisable en pratique. On se propose de remédier à cet obstacle via une méthode de référence, la validation croisée, présentée dans la section suivante.

Le choix optimal de la fenêtre dans le cadre multivarié est fondé sur les formules asymptotiques (1.35) et (1.19), en supposant vérifiées les hypothèses des propositions 1.3.2 et 1.3.6. Nous citons les articles de Mack et Müller (1987) [94] ainsi que Müller et Prewitt (1993) [106], qui ont démontré la consistance d'estimateurs à noyaux de type [NW], construits avec une fenêtre asymptotiquement optimale, via la méthode plug-in. Les estimateurs proposés sont alors asymptotiquement efficaces (au sens MSE) et on note que la technique de démonstration repose sur l'étude de la convergence faible d'un certain processus d'erreur (cf. [1] et [87] dans le cadre de l'estimation de la densité). A ce titre, nous rappelons que la convergence faible d'un processus stochastique s'appuie sur deux arguments, l'étude de la convergence faible en dimension finie combinée à une hypothèse d'équicontinuité.

1.5 La validation croisée

Dans cette section, nous supposons le noyau K fixé, et on ne s'intéresse qu'au choix de la fenêtre h . Nous avons observé dans les précédents paragraphes que l'efficacité de l'estimateur [N-W] est liée au paramètre de lissage, la fenêtre h . Il faut choisir la fenêtre afin d'équilibrer un terme stochastique (la variance) et un terme déterministe (le biais), si possible indépendamment des propriétés de régularité de la courbe de régression. Dans la précédente section, la fenêtre optimale qui minimise le risque quadratique intégré (MISE) est obtenue sous des hypothèses de régularité spécifiques et dépend alors de quantités inconnues, fonctionnelles de la distribution du couple (X, Y) . Afin de construire un estimateur non oracle qui minimise l'erreur quadratique, il faut utiliser d'autres méthodes dont la plus commune est appelée la procédure de validation croisée. L'idée principale de la validation croisée consiste à minimiser, par rapport à h , l'estimé d'une mesure de la MISE. La fenêtre h n'est alors plus déterministe, elle dépend des observations, à l'instar des méthodes plug-in dont nous reparlerons dans le paragraphe suivant.

Cadre de travail

Soient (X, Y) , (X_1, Y_1) , $(X_2, Y_2), \dots$, des variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^p \times \mathbb{R}$. Nous considérons des estimateurs à noyaux, avec fenêtre aléatoire (ou “*data-driven bandwidth*”) de la forme,

$$\hat{h} = \hat{h}_n := h_n\{(X_1, Y_1), \dots, (X_n, Y_n); \mathbf{x}\} \in H_n, \quad \mathbf{x} \in \mathbb{R}^p,$$

lorsque H_n désigne un sous-ensemble de \mathbb{R}_+^n (i.e., la zone de variation de \hat{h}_n). Soit $d(\cdot, \cdot)$ une certaine distance, utilisée pour définir le risque, qui servira à mesurer l'efficacité d'un certain estimateur \hat{m}_n de la fonction de régression. Afin de simplifier l'exposition de la procédure de validation croisée, nous travaillerons avec l'estimateur [NW] de la régression, qui sera noté \hat{m}_h pour souligner sa dépendance en h ,

$$\hat{m}_h(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i K\left(\frac{\mathbf{x} - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right)}, \quad \text{lorsque } \sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right) \neq 0.$$

La méthode de sélection de la fenêtre \hat{h} est dite **asymptotiquement optimale par rapport à la distance d** lorsque nous avons

$$\lim_{n \rightarrow \infty} \left[\frac{d(\hat{m}_{\hat{h}}, m)}{\inf_{h \in H_n} d(\hat{m}_h, m)} \right] \stackrel{p.s.}{=} 1, \quad (1.44)$$

où la notation $\stackrel{p.s.}{=}$ désigne une égalité presque sûre. Par la suite, nous désignons par $w(\cdot)$ une fonction de poids positive et arbitraire. Les différentes distances considérées dans cette section sont :

-l'Erreur Moyenne Quadratique :

$$d_M(\hat{m}, m) = \frac{1}{n} \sum_{i=1}^n \{\hat{m}(X_i) - m(X_i)\}^2 w(X_i);$$

-l'Erreur Quadratique Intégrée :

$$d_I(\hat{m}, m) = \int_{\mathbf{R}^p} \{\hat{m}(\mathbf{x}) - m(\mathbf{x})\}^2 w(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x};$$

-l'Erreur Quadratique Intégrée Moyenne Conditionnelle :

$$d_C(\hat{m}, m) = \mathbb{E}[d_I(\hat{m}, m) | X_1, \dots, X_n].$$

Remarque 1.5.1 Chacune de ces mesures d'erreur d_M , d_I ou d_C se décompose en un terme de biais au carré et un terme de variance. Par exemple, la variance de l'estimateur $\hat{m}_n^{NW}(\mathbf{x})$ est proportionnelle à $f_X(\mathbf{x})^{-1}$ d'après (1.19). Il s'ensuit un choix naturel de $w(\mathbf{x}) = f_X(\mathbf{x})$ lorsque l'on travaille avec l'estimateur [NW] (voir Nadaraya (1982) ou Härdle et Kelly (1987) [71]).

Maintenant, nous allons présenter la procédure de sélection de la fenêtre aléatoire \hat{h} pour la distance d_I . On peut décomposer $d_I(\hat{m}_h, m)$ de la manière suivante,

$$\begin{aligned} d_I(\hat{m}_h, m) &= \int_{\mathbf{R}^p} \{\hat{m}_h(\mathbf{x}) - m(\mathbf{x})\}^2 w(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{R}^p} \hat{m}_h^2(\mathbf{x}) w(\mathbf{x}) f_X(\mathbf{x}) d(\mathbf{x}) - 2 \int_{\mathbf{R}^p} \hat{m}_h(\mathbf{x}) m(\mathbf{x}) w(\mathbf{x}) f_X(\mathbf{x}) d(\mathbf{x}) \\ &\quad + \int_{\mathbf{R}^p} m^2(\mathbf{x}) w(\mathbf{x}) f_X(\mathbf{x}) d(\mathbf{x}). \end{aligned}$$

Comme la dernière intégrale est indépendante de h , pour minimiser la perte associée à la distance d_I en fonction de h , il suffit de minimiser

$$\int_{\mathbf{R}^p} \hat{m}_h^2(\mathbf{x}) w(\mathbf{x}) f_X(\mathbf{x}) d(\mathbf{x}) - 2 \int_{\mathbf{R}^p} \hat{m}_h(\mathbf{x}) m(\mathbf{x}) w(\mathbf{x}) f_X(\mathbf{x}) d(\mathbf{x}). \quad (1.45)$$

Cependant, ceci n'est pas réalisable en pratique car cette dernière quantité dépend de fonctions inconnues $m(\cdot)$ et $f_X(\cdot)$. La méthode classique pour contourner cette difficulté consiste à remplacer ces termes par leur versions empiriques. Nous remarquons que le deuxième terme de l'intégrale

$$\int_{\mathbf{R}^p} \hat{m}_h(\mathbf{x}) m(\mathbf{x}) w(\mathbf{x}) f_X(\mathbf{x}) d(\mathbf{x}) = \mathbb{E}[\hat{m}_h(X) Y w(X)].$$

Il s'ensuit comme estimateur naturel,

$$\frac{1}{n} \sum_{i=1}^n \{\hat{m}_i(X_i) Y_i w(X_i)\},$$

où $\hat{m}_i(\cdot)$ est l'estimateur dénommé "leave-one-out", défini par,

$$\hat{m}_i(\mathbf{x}) := \frac{\sum_{j \neq i} Y_j K\left(\frac{\mathbf{x} - X_j}{h}\right)}{\sum_{j \neq i} K\left(\frac{\mathbf{x} - X_j}{h}\right)}$$

L'estimateur "leave-one-out" est simplement l'estimateur [N-W] construit avec les $(n-1)$ couples aléatoires $\{(X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), \dots, (X_n, Y_n)\}$. De même, il est possible d'approximer le premier terme intégrale de (1.45) par,

$$\frac{1}{n} \sum_{i=1}^n \{\hat{m}_i^2(X_i) w(X_i)\}.$$

En somme, il paraît raisonnable de choisir la fenêtre h qui minimise la version empirique de (1.45), c'est à dire h qui minimise :

$$\frac{1}{n} \sum_{i=1}^n \{\hat{m}_i^2(X_i) w(X_i)\} - \frac{2}{n} \sum_{i=1}^n \{\hat{m}_i(X_i) Y_i w(X_i)\}.$$

Cette dernière quantité est égale à

$$\frac{1}{n} \sum_{i=1}^n \{\hat{m}_i(X_i) - Y_i\}^2 w(X_i) - \frac{1}{n} \sum_{i=1}^n \{Y_i\}^2 w(X_i).$$

où le deuxième terme ne dépend pas de h et n'intervient donc pas dans la minimisation. Le critère de sélection de la fenêtre se réduit à :

-choisir \hat{h} qui minimise

$$CV(h) := \frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}_i(X_i)\}^2 w(X_i). \quad (1.46)$$

Cette méthode est bien connue dans la littérature statistique et est appelée **procédure par validation croisée**. Les références principales à ce sujet sont Hall (1984) [61], Härdle et Marron (1985) [73], Härdle et Kelly (1987) [71], concernant l'estimation non-paramétrique de la régression. La procédure de validation croisée peut s'interpréter comme étant le meilleur choix de h qui fait de $\hat{m}_i(X_i)$ un estimateur efficace de Y_i au sens de (1.46). Sous les hypothèses (A.1-6), p. 1467-1468, [73], nous avons le théorème suivant :

Théorème 1.5.1 Härdle et Marron (1985)

La procédure de validation croisée, choisir \hat{h} qui minimise $CV(h)$, est asymptotiquement optimale, au sens de (1.44), par rapport aux distances d_M , d_I et d_C .

Autres méthodes de sélection

En supposant le noyau K fixé, sélectionner la meilleure fenêtre h à partir des données consiste à définir une fenêtre aléatoire de la forme $h((X_1, Y_1), \dots, (X_n, Y_n))$ approchant au mieux la fenêtre optimale (cette fois-ci déterministe) au sens d'un certain critère (généralement minimiser une erreur liée à une distance ou une norme). Il est important de remarquer que l'optimalité n'est pas un concept absolu mais est liée au choix d'une fonction de risque (MSE ou MISE par exemple). Si la procédure de sélection de la fenêtre ne requiert aucun choix de paramètre a priori, nous dirons qu'une telle méthode est **automatique**. Le praticien qui cherche à trouver le paramètre de lissage optimal en fonction des données a le choix entre deux méthodologies principales. La première méthodologie comprend des procédures de sélection traditionnelles, essentiellement des variations de la validation croisée définie dans la section précédente. Ces procédures sont automatiques et asymptotiquement équivalentes d'après l'article de Härdle, Hall and Marron (1988) [69]. Un des problèmes majeurs (ou point faible) lié à la validation croisée est son manque de robustesse par rapport aux changements de taille de l'échantillon. Plus précisément, le paramètre de lissage optimisant une certaine mesure d'erreur ne peut être approché qu'à la vitesse $n^{1/10}$. La principale alternative à la validation croisée est d'utiliser une procédure de sélection de fenêtre dite "de deuxième génération". Ces procédures, développées principalement dans les années 90, sont de type plug-in et donnent de meilleurs résultats théoriques et pratiques. Il convient de citer l'article de Jones, Marron and Sheather (1996) [83] pour des références complètes et une étude comparative des différentes méthodes de sélection, mettant en exergue l'avantage des méthodes plug-in sur la validation croisée classique. Par exemple, dans l'article de Härdle, Hall and Marron (1992) [70], la vitesse de convergence de la fenêtre aléatoire vers la fenêtre théorique est optimale, c'est à dire de l'ordre $O(n^{-1/2})$, telle que

$$n^{1/2} \left\{ \frac{\hat{h} - h_0}{h_0} \right\} \xrightarrow{\mathcal{L}} \mathcal{N},$$

où h_0 dénote la fenêtre optimale, \hat{h} son estimé et \mathcal{N} une variable aléatoire gaussienne. Pour des travaux plus récents concernant les méthodes de sélection de fenêtre avec vitesse de convergence \sqrt{n} dans un cadre multivarié, nous citons enfin Wu et Tsai (2004) [150].

1.6 Normalité asymptotique

La première démonstration de la normalité asymptotique de l'estimateur [NW] est due à Schuster (1972) [120]. On se réfère également aux théorèmes 1.3 et 1.4 p. 117-120 de Nadaraya [109] et au théorème 4.2.1 p. 99 de Härdle (1990) [66], qui proposent d'autres méthodes de démonstration. Le noyau K est supposé borné, à support compact et d'ordre 2. La fenêtre h_n est choisie égale à $cn^{-1/5}$.

Théorème 1.6.1 Härdle (1990)

Supposons Y bornée ou admettant un moment d'ordre $l > 2$. Les fonctions $f_X(\cdot)$ et $m(x)$

sont supposées deux fois continûment dérivables sur \mathbb{R} . A chaque point de continuité de $\sigma^2(x)$, tel que $f_X(x) > 0$,

$$(nh)^{1/2} \{ \hat{m}_n^{NW}(x) - m(x) \} \xrightarrow{\mathcal{L}} \mathcal{N}(B(x), v^2(x)), \quad (1.47)$$

avec

$$v^2(x) := \frac{\sigma^2(x)}{f_X(x)} \int_{\mathbb{R}} K^2(u) du, \quad (\text{la variance asymptotique}),$$

et

$$B(x) := \left\{ m''(x) + 2m'(x) \frac{f'_X(x)}{f_X(x)} \right\} \times \int_{\mathbb{R}} u^2 K(u) du, \quad (\text{le biais asymptotique}).$$

Pour un nombre d de points x_1, \dots, x_d de continuité, nous avons,

$$\left\{ (nh)^{1/2} \left\{ \frac{\hat{m}_n^{NW}(x_i) - m(x_i)}{v(x_i)} \right\} \right\}_{i=1}^d \xrightarrow{\mathcal{L}} \mathcal{N}_d \left(\{ B(x_i) \}_{i=1}^d, \mathbf{I}_d \right), \quad (1.48)$$

où \mathbf{I}_d dénote la matrice identité d -dimensionnelle.

-Extension multidimensionnelle : $X \in \mathbb{R}^p$

Afin d'énoncer proprement le théorème concernant la normalité asymptotique, nous récapitulons certaines hypothèses essentielles, liées au contrôle du biais et de la variance dans le cadre multivarié.

Soit $V_{\mathbf{x}}$ un voisinage du point \mathbf{x} . On suppose les conditions suivantes sur la distribution du couple (X, Y) .

- Toutes les dérivées partielles d'ordre 2 de $m(\cdot)$ existent sur $V_{\mathbf{x}}$;
- toutes les dérivées partielles d'ordre 2 de $f_X(\cdot)$ existent et sont continues sur $V_{\mathbf{x}}$, de plus $f_X(\mathbf{u}) > 0$, pour tout $\mathbf{u} \in V_{\mathbf{x}}$;
- la densité jointe $f_{X,Y}(\mathbf{u}, y)$ est continue sur $V_{\mathbf{x}} \times \mathbb{R}$, et toutes les dérivées partielles d'ordre 2 par rapport aux composantes du vecteur \mathbf{u} existent et sont continues sur $V_{\mathbf{x}} \times \mathbb{R}$.

Dans le cadre multivarié, la fonction noyau $K : \mathbb{R}^p \rightarrow \mathbb{R}$ satisfait :

- celles - K est à support compact tel que $\int_{\mathbb{R}^p} K^2(\mathbf{u}) d\mathbf{u} < \infty$;
- K est d'ordre 2.

La fenêtre $h = h_n$ vérifie $h \rightarrow 0$ et $nh^p \rightarrow \infty$. Plus précisément, en vue d'un équilibre biais-variance, nous choisissons h de l'ordre $n^{-1/(4+p)}$. On rappelle l'expression de la variance et du biais asymptotiques de l'estimateur $\hat{m}_n^{NW}(\mathbf{x})$: via (1.19),

$$\frac{1}{nh^p} \times \left\{ \frac{\text{Var}[Y|X = \mathbf{x}]}{f_X(\mathbf{x})} \int_{\mathbb{R}^p} K^2(\mathbf{u}) d\mathbf{u} \right\} =: \frac{1}{nh^p} v^2(\mathbf{x}),$$

et d'après (1.34),

$$\frac{h^2}{2} \left\{ \frac{Q[r](\mathbf{x}) - m(\mathbf{x})Q[f_X](\mathbf{x})}{f_X(\mathbf{x})} \right\} =: h^2 B(\mathbf{x}).$$

En supposant les hypothèses ci-dessus vérifiées, il s'ensuit la normalité asymptotique dans le cadre multivarié. D'après la proposition 3, p. 243, de Müller et Song (1993) [107] :

Théorème 1.6.2 Müller et Song (1993)

$$\{nh^p\}^{1/2} \left\{ \hat{m}_n^{NW}(\mathbf{x}) - m(\mathbf{x}) \right\} \xrightarrow{\mathcal{L}} \mathcal{N}(B(\mathbf{x}), v^2(\mathbf{x})).$$

1.7 Estimation par la méthode des polynômes locaux

L'estimation de la fonction de régression par la méthode des polynômes locaux est fondée sur une simple généralisation de l'estimateur [NW]. L'idée maîtresse de l'approche localement polynomiale est de considérer le problème de la régression sous l'angle des moindres carrés. Intuitivement, cette démarche est pleine de bon sens, en dénotant que la fonction de régression $m(\cdot)$ est elle-même solution d'un problème de moindres carrés. Par convenance, nous rappelons la définition de l'estimateur [NW] : lorsque $K \geq 0$,

$$\hat{m}_n^{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} = \frac{\hat{r}_n(x)}{\hat{f}_{X;n}(x)}.$$

Nous avons, lorsque $K \geq 0$,

$$\left\{ \hat{r}_n(x) - \hat{m}_n^{NW}(x) \hat{f}_{X;n}(x) \right\} = 0.$$

L'estimateur de la régression $\hat{m}_n^{NW}(x)$ peut donc être regardé comme la solution du problème de moindres carrés pondérés suivant :

$$\arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^n \{Y_i - \theta\}^2 K\left(\frac{x - X_i}{h_n}\right). \tag{1.49}$$

En d'autres termes, l'estimateur $\hat{m}_n^{NW}(x)$ est obtenu par une approximation des moindres carrés localement constante. Le principe de l'estimation localement polynomiale consiste en l'ajustement local d'un polynôme de degré p aux données $\{(X_i, Y_i) : 1 \leq i \leq n\}$. Le but de cette section est de présenter les estimateurs localement polynomiaux ainsi que leurs propriétés statistiques fondamentales.

1.7.1 Construction et définition des estimateurs localement polynomiaux

Soit p un entier naturel fixé. Nous cherchons à ajuster le polynôme

$$\beta_0 + \beta_1(\cdot - x) + \beta_2(\cdot - x)^2 + \dots + \beta_p(\cdot - x)^p$$

aux données (X_i, Y_i) , via la méthode des moindres carrés pondérés .

Premièrement, on suppose l'existence de la $(p+1)$ -ième dérivée de la fonction de régression $m(\cdot)$ au point x . Cette hypothèse, bien que difficile à vérifier en pratique, est essentielle

pour valider théoriquement la construction de l'estimateur localement polynomial. Nous pouvons alors approximer localement la fonction de régression $m(x)$ par un polynôme d'ordre p . Il s'ensuit, via le développement de Taylor autour du point x ,

$$\begin{aligned} m(z) &\approx m(x) + m'(x)(z-x) + \frac{m''(x)}{2}(z-x)^2 + \dots + \frac{m^{(p)}(x)}{p!}(z-x)^p \\ &\approx \sum_{j=0}^p \frac{m^{(j)}(x)}{j!}(z-x)^j =: \sum_{j=0}^p \beta_j (z-x)^j, \end{aligned} \quad (1.50)$$

lorsque z est situé dans un voisinage du point x .

A présent, nous ajustons localement le polynôme (1.50) aux données $\{(X_i, Y_i) : 1 \leq i \leq n\}$ par la méthode des moindres carrés pondérés avec comme fonction de poids $K\{(\cdot-x)/h_n\}$. Il faut minimiser par rapport au vecteur $\beta = (\beta_0, \dots, \beta_p)^T \in \mathbb{R}^{p+1}$ la quantité suivante

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x)^j \right\}^2 K\left(\frac{X_i - x}{h_n}\right). \quad (1.51)$$

Comme pour l'estimateur [NW], les paramètres K et h_n déterminent la forme et la taille du voisinage autour du point x . Soit $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^T \in \mathbb{R}^{p+1}$, le vecteur qui minimise l'expression (1.51). D'après l'égalité en (1.50), la dérivée k -ième $m^{(k)}(x)$ peut être donc estimer par $\hat{\beta}_k \times k!$, pour $k = 0, 1, \dots, p$. Il s'ensuit la définition suivante :

Définition 1.7.1 La statistique

$$\hat{m}_n^{(k)}(x; p) = \hat{\beta}_k \times k!, \quad 0 \leq k \leq p, \quad (1.52)$$

est l'**estimateur localement polynomial d'ordre p** de la dérivée k -ième de la régression $m^{(k)}(x)$, et noté estimateur [LP](p) de $m^{(k)}(x)$.

Lorsque $k = p = 0$, on retrouve bien l'estimateur [NW], i.e. $\hat{m}_n(x; 0) = \hat{m}_n^{NW}(x)$. Un exemple particulièrement intéressant est le cas $p = 1$ et $k = 0$. L'estimateur $\hat{m}_n(x; 1)$ de la fonction de régression est appelé l'**estimateur localement linéaire** et noté $\hat{m}_n^{LL}(x)$. D'après (1.51) et (1.52), il est égal à $\hat{\beta}_0$ lorsque $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ désigne le vecteur solution de l'équation des moindres carrés suivante :

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n \left\{ Y_i - \beta_0 - \beta_1 (X_i - x) \right\}^2 K\left(\frac{X_i - x}{h_n}\right).$$

Plus explicitement, l'estimateur [LL] est défini par :

$$\hat{m}_n^{LL}(x) := \frac{\hat{r}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{r}_{n,1}(x)\hat{f}_{n,1}(x)}{\hat{f}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{f}_{n,1}(x)\hat{f}_{n,1}(x)}, \quad (1.53)$$

où

$$\hat{f}_{n,j}(x) := \frac{1}{nh_n} \sum_{i=1}^n \left\{ \frac{X_i - x}{h_n} \right\}^j K\left(\frac{X_i - x}{h_n}\right), \quad j = 0, 1, 2,$$

$$\hat{r}_{n,j}(x) := \frac{1}{nh_n} \sum_{i=1}^n Y_i \left\{ \frac{X_i - x}{h_n} \right\}^j K\left(\frac{X_i - x}{h_n}\right), \quad j = 0, 1.$$

Nous constaterons, par la suite, que les estimateurs [LP] sont supérieurs aux estimateurs à noyaux [NW] (1.4) et [GM] (1.12) dans le cadre du dispositif expérimental aléatoire. D'après Fan (1992) [45], l'estimateur [LL] ou [LP](1) a un meilleur biais que l'estimateur [NW] et une meilleure variance que l'estimateur [GM]. De plus, l'estimateur [LL] a de bonnes propriétés minimax, il est le meilleur estimateur sur la classe des fonctions de régression à dérivée seconde bornée, parmi tous les estimateurs linéaires (cf. Fan (1993), [46]). On se réfère aux ouvrages de Wand et Jones (1995) [148] et Fan et Gijbels (1996) [49] pour une exposition complète des propriétés des estimateurs [LP] avec de nombreuses applications statistiques.

1.7.2 Biais et variance des estimateurs localement polynomiaux

Les estimateurs localement polynomiaux sont issus d'un problème de moindres carrés. Il est préférable d'adopter une notation matricielle dans ce contexte. Soit \mathbf{X}_x la matrice associée à notre dispositif expérimental :

$$\mathbf{X}_x = \mathbf{X} = \begin{pmatrix} 1 & (X_1 - x) & \dots & (X_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x) & \dots & (X_n - x)^p \end{pmatrix}_{n \times (p+1)}.$$

Nous posons

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1} \quad \text{et} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1) \times 1}.$$

On désigne par \mathbf{W}_x la matrice diagonale $n \times n$ de poids :

$$\mathbf{W}_x = \mathbf{W} = \text{diag} \left\{ K\left(\frac{X_i - x}{h_n}\right) \right\}.$$

La problématique des moindres carrés (1.51) peut se résumer ainsi :

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

où le signe T dénote la transposition, pour un vecteur ou une matrice. On suppose dorénavant l'inversibilité de la matrice carré $\mathbf{X}^T \mathbf{W} \mathbf{X} \in \mathcal{M}_{p+1}(\mathbb{R})$.

Remarque 1.7.1 Plus généralement, si la matrice $\mathbf{X}^T \mathbf{W} \mathbf{X} \in \mathcal{M}_{p+1}(\mathbb{R})$ est définie positive l'estimateur [LP](p) appartient à la classe des estimateurs linéaires (cf. (1.60) ci-après).

D'après la théorie des moindres carrés, le vecteur de solution est donné par

$$\hat{\beta} = \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}. \quad (1.54)$$

Cette dernière égalité (1.54) permet de formuler aisément le biais et la variance conditionnels de l'estimateur $\hat{\beta}$. Nous rappelons la définition du vecteur β ,

$$\beta = \left\{ m(x), \dots, \frac{m^{(p)}(x)}{p!} \right\}^T,$$

d'après (1.50). Soit \mathbb{X} l'ensemble des variables X_i , $1 \leq i \leq n$. Nous définissons,

$$\mathbf{m} = \{m(X_1), \dots, m(X_n)\}^T \quad \text{et} \quad \mathbf{r} = \mathbf{m} - \mathbf{X}\beta \quad \text{le vecteur des résidus.}$$

Il s'ensuit, d'après (1.54),

$$\begin{aligned} \mathbb{E}[\hat{\beta} | \mathbb{X}] &= \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{W} \mathbf{m} \\ &= \beta + \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{W} \mathbf{r}. \end{aligned} \quad (1.55)$$

Soit

$$\Sigma = \text{diag} \left\{ K^2 \left(\frac{X_i - x}{h_n} \right) \right\} \sigma^2(X_i) \in \mathcal{M}_n(\mathbb{R}),$$

où $\sigma^2(x) = \text{Var}[Y|X=x]$. La matrice de variance-covariance conditionnelle est

$$\text{Var}[\hat{\beta} | \mathbb{X}] = \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \{\mathbf{X}^T \Sigma \mathbf{X}\} \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \quad (1.56)$$

Les expressions (1.55) et (1.56) ne sont pas directement utilisables, car elles dépendent de quantités inconnues : le vecteur des résidus \mathbf{r} et la matrice Σ . Ruppert et Wand (1994) [118] ont obtenu des développements asymptotiques pour le biais et la variance de l'estimateur localement polynomial $\hat{m}_n^{(k)}(x; p)$ défini en (1.52). Avant d'énoncer leur théorème, on rappelle quelques notations utiles. Les moments de K et K^2 sont dénotés par

$$[\mu_j(K)] = \int_{\mathbb{R}} u^j K(u) du \quad \text{et} \quad [\mu_j(K^2)] = \int_{\mathbb{R}} u^j K^2(u) du \quad \text{respectivement,}$$

avec $j \in \mathbb{N}$. Soient

$$\begin{aligned} S &= \left([\mu_{j+l}(K)] \right)_{0 \leq j, l \leq p} \in \mathcal{M}_{p+1}(\mathbb{R}) \\ \tilde{S} &= \left([\mu_{j+l+1}(K)] \right)_{0 \leq j, l \leq p} \in \mathcal{M}_{p+1}(\mathbb{R}) \\ \bar{S} &= \left([\mu_{j+l}(K^2)] \right)_{0 \leq j, l \leq p} \in \mathcal{M}_{p+1}(\mathbb{R}) \\ c_p &= \left([\mu_{p+l}(K)], \dots, [\mu_{2p+1}(K)] \right)^T \in \mathbb{R}^{p+1} \\ \tilde{c}_p &= \left([\mu_{p+2}(K)], \dots, [\mu_{2p+2}(K)] \right)^T \in \mathbb{R}^{p+1}. \end{aligned}$$

Nous désignons par $e_{k+1} = (0, \dots, 0, 1, 0, \dots, 0)^T$ le $(k+1)$ -ième vecteur unité dans \mathbb{R}^{p+1} .

Théorème 1.7.1 Ruppert et Wand (1994)

Nous supposons $f_X(x) > 0$ et les fonctions $f_X(\cdot)$, $m^{p+1}(\cdot)$ et $\sigma^2(\cdot)$ continues dans un voisinage du point x . La fenêtre h vérifie $h \rightarrow 0$ et $nh \rightarrow \infty$. Alors, nous obtenons,

$$\text{Var} [\hat{m}_n^{(k)}(x; p) | \mathbb{X}] = (k!)^2 \times e_{k+1}^T S^{-1} \bar{S} S^{-1} e_{k+1} \frac{\sigma^2(x)}{nh^{1+2k} f_X(x)} + o_{\mathbf{P}} \left(\frac{1}{nh^{1+2k}} \right). \quad (1.57)$$

Lorsque $p - k$ est impair,

$$\text{Biais} [\hat{m}_n^{(k)}(x; p) | \mathbb{X}] = k! \times e_{k+1}^T S^{-1} \frac{c_p}{(p+1)!} m^{(p+1)}(x) h^{p+1-k} + o_{\mathbf{P}}(h^{p+1-k}). \quad (1.58)$$

Lorsque $p - k$ est pair, en supposant $f'_X(\cdot)$ et $m^{(p+2)}(\cdot)$ continues dans un voisinage du point x ainsi que $nh^3 \rightarrow \infty$, le biais conditionnel asymptotique est donné par,

$$k! \times e_{k+1}^T \tilde{S}^{-1} \frac{\tilde{c}_p}{(p+2)!} \left\{ m^{(p+2)}(x) + (p+2)m^{(p+1)}(x) \frac{f'_X(x)}{f_X(x)} \right\} h^{p+2-k} + o_{\mathbf{P}}(h^{p+2-k}).$$

D'après le théorème ci-dessus, il apparaît clairement une différence entre le cas $p - k$ pair et le cas $p - k$ impair. Lorsque $p - k$ pair, le terme de biais principal en $O(h^{p+1})$ s'annule via la symétrie de noyau K . Par contre, lorsque $p - k$ impair, le terme de biais asymptotique a une expression simple où ne figure pas de termes de dérivées tels $f'_X(x)$. On remarque que lorsque $p = k = 0$, on retrouve bien le biais asymptotique de l'estimateur [NW]. D'un point de vue pratique et théorique, nous privilégierons le cas $p - k$ impair (cf. la section 3.3 de [49]), où la forme du biais est plus appréciable d'un point de vue théorique.

La meilleure représentation des estimateurs [LP] est obtenue par la méthode des "noyaux équivalents", c'est à dire en réécrivant asymptotiquement les estimateurs [LP] sous une forme plus classique proche de l'estimateur [NW]. Nous introduisons la notation suivante :

$$S_{n,j} = \sum_{i=1}^n (X_i - x)^j K\left(\frac{X_i - x}{h_n}\right). \quad (1.59)$$

Soit $S_n = \mathbf{X}^T \mathbf{W} \mathbf{X}$ la matrice carré de dimension $p + 1$ définie également par,

$$S_n = \{S_{n,j+l}\}_{0 \leq j, l \leq p}.$$

D'après (1.54),

$$\begin{aligned} \hat{\beta}_k &= e_{k+1}^T \hat{\beta} = e_{k+1}^T S_n^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \\ &= \sum_{i=1}^n W_k^n \left(\frac{X_i - x}{h_n} \right) Y_i. \end{aligned} \quad (1.60)$$

On remarque que

$$\mathbf{X}^T \mathbf{W} = \begin{pmatrix} K\left(\frac{X_1 - x}{h_n}\right) & \dots & K\left(\frac{X_n - x}{h_n}\right) \\ \vdots & \vdots & \vdots \\ (X_1 - x)^p K\left(\frac{X_1 - x}{h_n}\right) & \dots & (X_n - x)^p K\left(\frac{X_n - x}{h_n}\right) \end{pmatrix}_{(p+1) \times n}.$$

Il s'ensuit

$$\mathbf{X}^T \mathbf{W} \mathbf{y} = \begin{pmatrix} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h_n}\right) \\ \vdots \\ \sum_{i=1}^n Y_i (X_i - x)^p K\left(\frac{X_i - x}{h_n}\right) \end{pmatrix}_{(p+1) \times 1}.$$

Nous obtenons finalement,

$$W_k^n\left(\frac{X_i - x}{h_n}\right) = e_{k+1}^T S_n^{-1} \times \begin{pmatrix} K\left(\frac{X_i - x}{h_n}\right) \\ \vdots \\ (X_i - x)^p K\left(\frac{X_i - x}{h_n}\right) \end{pmatrix}$$

ou

$$W_k^n(t) = e_{k+1}^T S_n^{-1} \{1, th, \dots, (th)^p\}^T K(t). \quad (1.61)$$

L'estimateur $\hat{\beta}_k$ a donc une forme conventionnelle, excepté que le noyau W_k^n dépend des points X_i et de leur localisation. Ceci explique intuitivement pourquoi l'estimation localement polynomiale s'adapte aux différents dispositifs expérimentaux ainsi qu'à l'estimation aux bornes du support de la densité. Nous énonçons à présent une propriété fondamentale des estimateurs [LP](p).

Lemme 1.7.1 *La fonction de poids $W_k^n(\cdot)$ satisfait la condition suivante :*

$$\sum_{i=1}^n (X_i - x)^q W_k^n\left(\frac{X_i - x}{h_n}\right) = \delta_{k,q}, \quad 0 \leq k, q \leq p.$$

Ci-dessus $\delta_{k,q}$ dénote le symbole de Kronecker.

$$\begin{aligned} \sum_{i=1}^n (X_i - x)^q W_k^n\left(\frac{X_i - x}{h_n}\right) &= e_{k+1}^T S_n^{-1} \sum_{i=1}^n (X_i - x)^q \begin{pmatrix} 1 \\ (X_i - x) \\ \vdots \\ (X_i - x)^p \end{pmatrix} K\left(\frac{X_i - x}{h_n}\right) \\ &= e_{k+1}^T S_n^{-1} S_n e_{q+1} = e_{k+1}^T \times e_{q+1} = \delta_{k,q}. \end{aligned}$$

□

Comme conséquence du lemme 1.7.1, le biais à distance finie de l'estimateur $\hat{\beta}_k$ est nul lorsque la fonction $m^{(k)}(\cdot)$ à estimer est un polynôme de degré inférieur ou égal à p . Cette propriété met en exergue un des avantages pratiques de l'estimation par la méthode des polynômes locaux pour la réduction du biais, en comparaison avec l'utilisation de noyaux d'ordres élevés. En effet, le biais est nul à n fixé et non asymptotiquement. En d'autres termes, l'estimateur [LP](p) possède la propriété de reproduire les polynômes de degré $q \leq p$ (cf. proposition 1.12, p. 32, [142]).

Nous continuons l'investigation des propriétés de la fonction de poids W_k^n . Nous notons que, lorsque $h \rightarrow 0$ et $nh \rightarrow \infty$,

$$\begin{aligned} S_{n,j} &= \mathbb{E}[S_{n,j}] + O_{\mathbf{P}}\left(\sqrt{\text{Var}[S_{n,j}]}\right) \\ &= nh^{j+1} \int_{\mathbf{R}} u^j K(u) f_X(x+hu) du + O_{\mathbf{P}}\left(\sqrt{\mathbb{E}[S_{n,j}^2]}\right) \\ &= nh^{j+1} \left\{ f_X(x)[\mu_j](K) + o(1) + O_{\mathbf{P}}(1/\sqrt{nh}) \right\} \\ &= nh^{j+1} f_X(x)[\mu_j](K) \{1 + o_{\mathbf{P}}(1)\}, \end{aligned}$$

via une application du lemme de Bochner et de la loi des grands nombres. Il s'ensuit

$$S_n = n f_X(x) H S H \{1 + o_{\mathbf{P}}(1)\}, \quad (1.62)$$

où $H = \text{diag}\{1, h, \dots, h^p\}$. En substituant la formule (1.62) dans la définition (1.61) de $W_k^n(\cdot)$, nous obtenons

$$W_k^n(t) = \frac{1}{nh^{k+1} f_X(x)} e_{k+1}^T S^{-1} \{1, h, \dots, h^p\}^T K(t) \{1 + o_{\mathbf{P}}(1)\}.$$

Il en découle,

$$\hat{\beta}_k = \frac{1}{nh^{k+1} f_X(x)} \sum_{i=1}^n Y_i K_k^* \left(\frac{X_i - x}{h_n} \right) \{1 + o_{\mathbf{P}}(1)\}, \quad (1.63)$$

avec

$$K_k^*(t) := e_{k+1}^T S^{-1} \{1, h, \dots, h^p\}^T K(t). \quad (1.64)$$

Le noyau en (1.64) est appelé noyau équivalent (“*equivalent kernel*”) et est très utile pour exprimer les propriétés asymptotiques de l'estimateur $[\text{LP}](p)$. Le noyau (1.64) vérifie les conditions de moments suivantes :

$$\int_{\mathbf{R}} u^q K_k^*(u) du = \delta_{k,q}, \quad 0 \leq k, q \leq p. \quad (1.65)$$

Le noyau équivalent $K_k^*(u)$ est donc simplement un noyau d'ordre $(k, p+1)$ (cf. (A.24) ou définition A.6.2 en annexe). On le note en conséquence $K_{k,p}^*(u)$ afin de souligner la dépendance en p . Pour plus de détails concernant les noyaux équivalents et le lien entre l'estimation $[\text{LP}]$ et les autres méthodes d'estimation ($[\text{NW}]$ et $[\text{GM}]$), nous nous référons aux articles de Lejeune (1985) [90] et Müller (1987) [103].

La variance et le biais conditionnels de l'estimateur $\hat{m}_n^{(k)}(x; p)$, spécifiés en (1.57) et (1.58) respectivement, peuvent être exprimés en fonction du noyau équivalent $K_{k,p}^*(\cdot)$, nous conduisant aux expressions asymptotiques suivantes :

$$\text{Var} [\hat{m}_n^{(k)}(x; p) | \mathbb{X}] = \left\{ \frac{1}{nh^{1+2k}} \times \frac{\sigma^2(x)}{f_X(x)} \left\{ (k!)^2 \int_{\mathbf{R}} \{K_{k,p}^*(u)\}^2 du \right\} \right\} \{1 + o_{\mathbf{P}}(1)\}, \quad (1.66)$$

et

$$\text{Biais} [\hat{m}_n^{(k)}(x; p) | \mathbb{X}] = \left\{ h^{p+1-k} \times \frac{m^{(p+1)}(x)}{(p+1)!} \left\{ k! \int_{\mathbb{R}} u^{p+1} K_{k,p}^*(u) du \right\} \right\} \{1 + o_{\mathbf{P}}(1)\}. \quad (1.67)$$

Ces développements asymptotiques sont obtenues aisément en s'appuyant sur les formules (1.63) et (1.65).

La fenêtre optimale au sens du critère local de minimisation de l'AMSE est obtenue à partir de (1.66) et (1.67) :

$$h^{\text{MSE}}(x) = C_{k,p}(K) \left\{ \frac{\sigma^2(x)}{f_X(x) \{m^{(p+1)}(x)\}^2} \right\}^{1/(2p+3)} n^{-1/(2p+3)},$$

avec

$$C_{k,p}(K) := \left\{ \frac{\{(p+1)!\}^2 (2k+1) [\mu_0(\{K_{k,p}^*(u)\}^2)]}{2(p+1-k) [\mu_{p+1}(K_{k,p}^*(u))]} \right\}^{1/(2p+3)}.$$

A partir de ces formules, il existe différentes procédures pour choisir la fenêtre optimale à partir des données. Citons l'article de Fan et Gijbels (1995) [48] qui combine les notions de la validation croisée et du "plug-in" ainsi que le papier de Ruppert, Sheather et Wand (1995) [117] qui propose plusieurs méthodes de sélection globale de la fenêtre, adaptations de techniques de type plug-in développées dans le cadre de l'estimation de la densité. Leurs travaux sont fondés sur la minimisation de la MISE conditionnelle, définie par,

$$\text{MISE}[\hat{m}_n(x; p) | \mathbb{X}] = \mathbb{E} \left[\int_{\mathbb{R}} \left\{ \hat{m}_n(x; p) - m(x) \right\}^2 f_X(x) dx | \mathbb{X} \right].$$

La fenêtre MISE-optimale a donc pour expression asymptotique, d'après (1.66) et (1.67),

$$h^{\text{MISE}} = C_{0,p}(K) \left\{ \frac{\int_{\mathbb{R}} \sigma^2(x) dx}{\int_{\mathbb{R}} f_X(x) \{m^{(p+1)}(x)\}^2 dx} \right\}^{1/(2p+3)} n^{-1/(2p+3)}. \quad (1.68)$$

On rappelle que les stratégies de type plug-in sont basées sur le remplacement dans (1.68) des intégrales inconnues par des estimateurs consistants. On peut citer également l'article de Wand et Gutierrez (1997) [147] qui proposent une approche intéressante, fondée sur l'expression du risque exact (i.e. à distance finie) et non des formulations asymptotiques.

La question du choix du noyau optimale est traité dans Fan, Gasser, Gijbels, Brockmann et Engel (1995) [47]). La normalité asymptotique est discutée dans l'article de Tenreiro (1997) [139] notamment. Pour une étude du cadre multivarié, on se réfère au chapitre 7 de Fan et Gijbels (1996) [49] et à l'article de Ruppert et Wand (1994) [118].

Enfin, pour un état de l'art comparatif des différentes techniques d'estimation de la fonction de régression par la méthode du noyau, nous citons Chu et Marron (1991) [16] et Hastie et Loader (1993) [75].

Chapitre 2

Lois uniformes du logarithme pour les dérivées de la régression

2.1 Introduction

L'objet central de cette thèse est de présenter de nouvelles lois uniformes du logarithme concernant une large classe d'estimateurs non-paramétriques de la fonction de régression ainsi que ses dérivées.

La méthodologie la plus ancienne permettant l'obtention de lois du logarithme itéré concernant les estimateurs à noyau est fondée sur un principe d'invariance fort démontré par Komlós, Major et Tusnády [85] et noté [KMT]. Par principe d'invariance fort, nous entendons l'approximation presque sûre du processus empirique par un certain processus gaussien. Le [KMT] consiste donc en l'approximation du processus empirique uniforme par une suite de ponts browniens et est utilisé par Hall (1981) [60] pour démontrer une loi du logarithme itéré pour l'estimateur [PR] de la densité. En s'appuyant sur la théorie des processus empiriques et certains résultats (cf. [132]) concernant le module d'oscillation du processus empirique uniforme, Stute (1982) [133] a établi la première loi du logarithme uniforme concernant l'estimateur [PR] de la densité. Ces résultats seront raffinés par Deheuvels et Mason en (1992) [26]. En ce qui concerne l'estimation [NW] de la régression, il existe une version bivariée du [KMT], développée par Tusnády (1977) [143] en s'appuyant sur la transformation de Rosenblatt (1952), qui pourrait aider à déterminer la vitesse de convergence optimale de l'estimateur [NW]. Toutefois, cette approximation n'est pas vraiment appropriée à l'étude du comportement asymptotique de l'estimateur [NW] (cf. remarque 5, p. 81, [41]). En effet, l'approximation de Tusnády n'est valide que pour un échantillon de taille fixée et ne permet pas l'écriture rigoureuse de lois limites. Il est possible cependant de préciser la vitesse de convergence exacte et la constante limite associée en utilisant des techniques de démonstration plus sophistiquées. Tout au long de ce chapitre, nous utiliserons la notation suivante $\hat{m}_n^{NW}(x) = \hat{m}_n(x)$ (cf. (1.4)). Pour éviter toute valeur négative du logarithme, nous introduisons la convention $\log(u) = \log(u \vee e)$. D'après les travaux récents de Deheuvels et Mason (2004) [29] (voir les travaux de Einmahl et Mason (2000) [42] pour la convergence p.s.), sous certaines hypothèses classiques,

nous avons la loi uniforme du logarithme de l'estimateur [NW] :

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{m}_n(x) - \tilde{\mathbb{E}}[\hat{m}_n(x)] \right\} - \sigma_m(I) \right| \stackrel{\mathbf{P}}{=} o(1),$$

où $\sigma_m(I)$ (cf. (2.5) ci-après) désigne la borne limite qui dépend de la variance asymptotique de l'estimateur [NW] et I dénote un intervalle compact arbitraire. Cet intervalle sera supposé contenu dans le support de la distribution afin d'éviter les valeurs nulles de la densité marginale f_X (cf. (1.17)). La méthodologie employée repose sur la théorie moderne des processus empiriques et, plus particulièrement, l'étude du processus empirique indexé par des classes de fonctions vérifiant certaines propriétés combinatoires. En effet, après linéarisation, nous remarquons que la déviation $\hat{m}_n(x) - \tilde{\mathbb{E}}[\hat{m}_n(x)]$ a le même comportement limite que le processus empirique indexé par une certaine classe de fonctions uniformément bornée. Il existe alors des inégalités exponentielles pour la déviation par rapport à l'espérance de la norme du supremum du processus empirique indexé par des classes de fonctions uniformément bornées (cf. section 2.14.3 dans [145], [88] et [137]). Nous verrons en annexe, section A.3, que ces résultats font appels à une borne de moment, c'est à dire pour majorer la norme L_1 du processus empirique indexé par certaines classes de fonctions particulières dites de Vapnik-Chervonenkis ou, plus explicitement, des classes de fonctions à nombre de recouvrement uniformément polynomial. Ces inégalités exponentielles, issues de la théorie moderne des processus empiriques, permettent de trouver la borne limite supérieure ci-dessus lorsque la norme du supremum est bornée ou plus précisément de contrôler les oscillations du processus empirique, étape classique de la technique de chaînage en vue de démontrer des résultats de nature uniforme. La méthodologie employée n'utilise pas de principe d'invariance fort, contrairement à certains résultats ponctuels de type loi du logarithme itéré (cf. Einmahl et Mason (1997)-(1998), [40] et [41]).

L'objet de ce chapitre est d'établir des lois limites, similaires à des lois du logarithme itéré, pour la déviation uniforme des estimateurs [NW] des dérivées de la régression et de généraliser ces résultats au modèle multivarié. En fin de chapitre, nous présenterons une extension de ces résultats à l'estimation localement polynomiale ainsi que certaines applications statistiques.

2.2 Le cadre univarié

Soient $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$, des couples aléatoires indépendants et identiquement distribués (i.i.d.) à valeurs dans \mathbb{R}^2 . Le couple de variables aléatoires (X, Y) est supposé admettre une densité jointe sur \mathbb{R}^2 notée $f_{X,Y}(\cdot, \cdot)$ et nous désignons par $f_X(\cdot)$ la densité marginale (par rapport à la mesure de Lebesgue sur \mathbb{R}) associée à la variable aléatoire X . Dans ce chapitre, nous considérons une version plus générale de la fonction de régression qui permettra de traiter diverses fonctionnelles de la densité conditionnelle de Y sachant $X = x$, telle la fonction de répartition conditionnelle $F(\cdot | x) = \mathbb{P}\{Y \leq \cdot | X = x\}$. Soit

$$m_\psi(x) := \mathbb{E}[\psi(Y) | X = x] =: \frac{1}{f_X(x)} \int_{\mathbf{R}} \psi(y) f_{X,Y}(x, y) dy, \quad (2.1)$$

où $\psi(\cdot)$ dénote une fonction à valeurs réelles, supposée mesurable et bornée sur tout intervalle compact dans \mathbb{R} . Cette hypothèse peu restrictive sert à borner $\psi(Y)$, lorsque nous travaillerons sous l'hypothèse (F.3), présentée ci-dessous. Nous définissons deux intervalles compacts $I = [a, b]$ et $J = [c, d]$, contenus dans \mathbb{R} , tels que

$$-\infty < c < a < b < d < \infty.$$

Nous supposons certaines conditions sur la distribution du couple (X, Y) parmi les hypothèses (F.1–5) énoncées ci-dessous. Soit k un entier naturel fixé désignant le degré de dérivation, tout au long de ce chapitre.

(F.1) $f_{X,Y}(\cdot, \cdot)$ est continue sur $J \times \mathbb{R}$;

(F.2) $f_X(\cdot)$ est continue et strictement positive sur J ;

(F.3) $Y \mathbb{I}\{X \in J\}$ est bornée.

Pour le cas non-borné, c'est à dire lorsque (F.3) n'est plus vérifiée, il nous faut une condition de moment liée à la troncation,

(F.4) $\sup_{x \in J} \mathbb{E} \left[|\psi(Y)|^s | X = x \right] < \infty$, pour un certain $s > 2$.

Enfin, en vue du traitement du biais ou de la construction d'intervalles de confiance pour la dérivée k -ième de la régression, nous supposerons

(F.5) f_X et $f_{X,Y}$ sont k -fois continûment différentiables sur $J \times \mathbb{R}$.

Nous avons clairement (F.3) implique (F.4). L'hypothèse (F.3) permet de borner les variables $\{Y_i : 1 \leq i \leq n\}$ et nous sera très utile pour la démonstration de nos prochains résultats. On note que cette hypothèse de bornitude est récurrente en régression non-paramétrique. Elle entraîne l'existence des divers moments de la distribution conditionnelle, notamment celui d'ordre deux. En fin de section nous traiterons le cas non-borné, en s'appuyant sur l'hypothèse (F.4), équivalente à un moment d'ordre strictement supérieur à deux. Cette fois-ci, (F.4) requiert une hypothèse supplémentaire, liée à s , concernant la fenêtre h_n associée à notre estimateur.

Remarque 2.2.1 Sous (F.1–3), la fonction de régression $m_\psi(\cdot)$ est proprement définie, $\forall x \in J$, par

$$m_\psi(x) = \frac{1}{f_X(x)} \int_{\mathbb{R}} \psi(y) f_{X,Y}(x, y) dy =: \frac{r_\psi(x)}{f_X(x)},$$

où nous notons par convenance,

$$r_\psi(x) = \int_{\mathbb{R}} \psi(y) f_{X,Y}(x, y) dy.$$

Sous (F.1–3), la variance conditionnelle de $\psi(Y)$ sachant $X = x$ est également bien définie,

$$\sigma_\psi^2(x) := \text{Var} [\psi(Y) | X = x] = \frac{1}{f_X(x)} \int_{\mathbb{R}} \left\{ \psi(y) - m_\psi(x) \right\}^2 f_{X,Y}(x, y) dy. \quad (2.2)$$

Sous les hypothèses (F.1–3), nous pouvons démontrer la continuité uniformément sur l'intervalle I des fonctions $r_\psi(\cdot)$, $m_\psi(\cdot)$ et $\sigma_\psi^2(x)$, via une application du lemme de Scheffé

ou du théorème de convergence dominée de Lebesgue (voir la section A.5 en annexe). Si, en outre, l'hypothèse additionnelle (F.5) est vérifiée, nous obtenons également la continuité des dérivées k -ièmes de $r_\psi(\cdot)$ et $m_\psi(\cdot)$, uniformément sur I .

Les estimateurs des dérivées de la régression seront de la même forme que ceux du précédent chapitre. Ils sont construits à partir de l'estimateur [NW]. On rappelle qu'une fonction mesurable $H : \mathbb{R} \rightarrow \mathbb{R}$ est à variation bornée sur \mathbb{R} lorsque,

$$\int_{\mathbb{R}} |dH(u)| := |H|_v < \infty,$$

où $|H|_v$ désigne la variation totale de la fonction $H(\cdot)$ sur \mathbb{R} . En vue d'une estimation des dérivées de la régression, le noyau K et ses dérivées successives jusqu'à l'ordre k satisfont,

(K.1) $K(\cdot)$ est à variation bornée et continue (à droite) sur \mathbb{R} ;

(K.2) $K(u) = 0$ pour $u \notin [-\xi/2; \xi/2]$, pour un certain $0 < \xi < \infty$;

(K.3) $\int_{\mathbb{R}} K(u)du = 1$;

(K.4) K est k -fois dérivable avec $|K^{(k)}|_v < \infty$.

Remarque 2.2.2 L'hypothèse (K.2) nous assure que le noyau K et ses dérivées successives jusqu'à l'ordre k sont à support compact. La valeur du nombre réel ξ est arbitraire et sera choisie égale à 1 sans perte de généralité. Autrement, la valeur de ξ n'interviendrait que dans des constantes ou pour régler le pas de la discrétisation. Dans certains cas, il est possible de s'affranchir de cette hypothèse, d'après Deheuvels (2000) [22]. On note que la condition (K.1), i.e. $K(\cdot)$ à variation bornée sur \mathbb{R} , est impliquée par les hypothèses (K.2) et (K.4), lorsque $k \geq 2$ (voir, par exemple, [76]). L'assertion (K.3) intervient uniquement lorsqu'on considère l'estimation de la densité, ou pour obtenir des estimateurs asymptotiquement sans biais, via le lemme de Bochner. Notons que sous l'hypothèse (K.3) l'estimateur [PR] est une fonction de densité. Enfin, l'hypothèse de continuité en (K.1) sera utile pour des problèmes liés à la mesurabilité (cf. lemme A.3.3 en annexe) et le prolongement par continuité de classes de fonctions dénombrables.

Nous travaillerons avec une fenêtre h_n , suite de nombres réels positifs, vérifiant certaines des conditions suivantes :

(H.1) $h_n \rightarrow 0$, lorsque $n \rightarrow \infty$;

(H.2) $nh_n / \log n \rightarrow \infty$, lorsque $n \rightarrow \infty$;

(H.3) $nh_n^{2k+1} / \log(1/h_n) \rightarrow \infty$, lorsque $n \rightarrow \infty$;

(H.4) $h_n \searrow 0$ et $nh_n \nearrow \infty$, lorsque $n \rightarrow \infty$;

(H.5) $|\log h_n| / \log \log n \rightarrow \infty$, lorsque $n \rightarrow \infty$.

Remarque 2.2.3 Si nous travaillons sous l'hypothèse (F.4), c'est à dire dans le cadre où Y n'est plus bornée, nous nécessitons une hypothèse plus forte que (H.2), notée (H.2)*,

$$(H.2)^* \quad n^{1-2/s} h_n \log n \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty, \quad \text{et avec } s > 2.$$

Les hypothèses (H.1–2) sont souvent, nécessaires et suffisantes pour la convergence uniforme en probabilité de la déviation associée aux estimateurs [PR] et [NW]. Plus précisément, les hypothèses (H.1–2) sont nécessaires et suffisantes pour la consistance forte des estimateurs [PR] et [NW] mais, pour démontrer des lois limites uniformes presque sûre, il nous faut également supposer (H.4–5). La condition (H.3) est spécifique à la consistance des estimateurs à noyaux des dérivées d'ordre k des fonctions f_X , r_ψ et m_ψ . A ce sujet, nous citons le théorème D, p. 1278 et p. 1281-1282 dans [26]. On peut également consulter la proposition 3 de Collomb (1979) [18] qui spécifie une condition nécessaire et suffisante de convergence uniforme p.s. et p.co. d'un estimateur des dérivées de la régression, i.e.

$$\lim_{n \rightarrow \infty} \frac{nh_n^{2k+1}}{\log n} = \infty.$$

Cette dernière condition est équivalente à (H.3) sous (H.1–2), c'est à dire lorsque $h_n = Cn^{-\alpha}$ avec $0 < \alpha < 1$ et $C > 0$. Les deux dernières hypothèses (H.4–5) sont donc spécifiques à la convergence presque sûre. En fait, les hypothèses $nh_n/\log(1/h_n) \rightarrow \infty$ et $\log(1/h_n)/\log_2 n \rightarrow \infty$ sont indispensables afin d'établir une loi limite uniforme presque sûre (cf. Mason, Shorack et Wellner (1983) [100]). A ce propos, nous rappelons qu'une suite de constantes $\{a_n : n \geq 1\}$ est supposée satisfaire les conditions dites de Csörgö-Révész-Stute [CRS] lorsque :

$$0 < a_n < 1, \quad a_n \searrow 0 \quad \text{et} \quad na_n \nearrow \infty, \quad \text{lorsque } n \rightarrow \infty;$$

$$\log(a_n)^{-1}/\log \log n \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty;$$

$$na_n/\log n \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty.$$

Nous pouvons maintenant présenter les estimateurs à noyaux de $f_X(x)$, $r_\psi(x)$, $m_\psi(x)$ et leurs dérivées jusqu'à l'ordre deux.

$$\begin{aligned} \hat{f}_{X;n}(x) &= \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \\ \hat{r}_{\psi;n}(x) &= \frac{1}{nh_n} \sum_{i=1}^n \psi(Y_i) K\left(\frac{x - X_i}{h_n}\right), \\ \hat{m}_{\psi;n}(x) &= \begin{cases} \frac{\hat{r}_{\psi;n}(x)}{\hat{f}_{X;n}(x)}, & \text{lorsque } \hat{f}_{X;n}(x) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n \psi(Y_i), & \text{sinon,} \end{cases} \\ \hat{f}'_{X;n}(x) &= \frac{1}{nh_n^2} \sum_{i=1}^n K'\left(\frac{x - X_i}{h_n}\right), \end{aligned}$$

$$\begin{aligned}\hat{r}'_{\psi;n}(x) &= \frac{1}{nh_n^2} \sum_{i=1}^n \psi(Y_i) K' \left(\frac{x - X_i}{h_n} \right), \\ \hat{m}'_{\psi;n}(x) &= \begin{cases} \frac{\hat{r}'_{\psi;n}(x)}{\hat{f}_{X;n}(x)} - \frac{\hat{r}_{\psi;n}(x) \hat{f}'_{X;n}(x)}{\hat{f}_{X;n}^2(x)}, & \text{lorsque } \hat{f}_{X;n}(x) \neq 0, \\ 0, & \text{sinon,} \end{cases} \\ \hat{f}''_{X;n}(x) &= \frac{1}{nh_n^3} \sum_{i=1}^n K'' \left(\frac{x - X_i}{h_n} \right), \\ \hat{r}''_{\psi;n}(x) &= \frac{1}{nh_n^3} \sum_{i=1}^n \psi(Y_i) K'' \left(\frac{x - X_i}{h_n} \right), \\ \hat{m}''_{\psi;n}(x) &= \frac{\hat{r}''_{\psi;n}(x)}{\hat{f}_{X;n}(x)} - \frac{2\hat{r}'_{\psi;n}(x) \hat{f}'_{X;n}(x) + \hat{r}_{\psi;n}(x) \hat{f}''_{X;n}(x)}{[\hat{f}_{X;n}(x)]^2} + \frac{2\hat{r}_{\psi;n}(x) [\hat{f}'_{X;n}(x)]^2}{[\hat{f}_{X;n}(x)]^3}, \\ &\text{lorsque } \hat{f}_{X;n}(x) \neq 0.\end{aligned}$$

Remarque 2.2.4 Le traitement des autres dérivées pour $k > 2$ est similaire et ne sera pas présenté ici par souci de clarté. Nous avons, plus généralement,

$$\hat{f}_{X;n}^{(k)}(x) = \frac{1}{nh_n^{1+k}} \sum_{i=1}^n K^{(k)} \left(\frac{x - X_i}{h_n} \right) \quad \text{et} \quad \hat{r}_{\psi;n}^{(k)}(x) = \frac{1}{nh_n^{1+k}} \sum_{i=1}^n \psi(Y_i) K^{(k)} \left(\frac{x - X_i}{h_n} \right).$$

Pour l'estimation de la dérivée d'ordre k de la régression, via le développement de Leibniz, nous obtenons

$$\hat{m}_{\psi;n}^{(k)}(x) = \sum_{j=0}^k C_k^j \hat{r}_{\psi;n}^{(j)}(x) \{ \hat{f}_{X;n}(x)^{-1} \}^{(k-j)}, \quad \text{lorsque } \hat{f}_{X;n}(x) \neq 0.$$

Dans la présentation de nos estimateurs, nous avons choisi comme fenêtre h_n . Il est possible toutefois d'adapter la fenêtre plus précisément, c'est à dire en fonction de l'ordre de dérivation de l'estimateur considéré. Nous introduisons alors l'hypothèse suivante, quelque soit $k \in \mathbb{N}$,

$$(H.k) \quad nh_{n,k}^{2k+1} / \log(1/h_n) \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty.$$

Nous construisons alors les estimateurs des dérivées d'ordre k de $f_X(\cdot)$ et $r_\psi(\cdot)$ en utilisant la fenêtre $h_{n,k}$ appropriée. Il s'ensuit,

$$\hat{f}_{X;n}^{(k)}(x) = \frac{1}{nh_{n,k}^{1+k}} \sum_{i=1}^n K^{(k)} \left(\frac{x - X_i}{h_{n,k}} \right) \quad \text{et} \quad \hat{r}_{\psi;n}^{(k)}(x) = \frac{1}{nh_{n,k}^{1+k}} \sum_{i=1}^n \psi(Y_i) K^{(k)} \left(\frac{x - X_i}{h_{n,k}} \right).$$

Ce raffinement n'a aucune incidence sur les démonstrations et ne sera pas présenté par la suite afin d'éviter des notations trop lourdes. Par contre, ces notations rejoignent certains

résultats développés par Deheuvels (2000), p. 942-943, [23], qui permettent de conjecturer un équivalent pour les dérivées de la régression (avec d'autres méthodes de preuve). Ces travaux traitent de l'approximation forte du processus de quantile uniforme, noté $\beta_n(t)$, par une version itérative du célèbre processus de Kiefer.

Nous définissons la notation suivante, concernant les termes de centrage,

$$f_{X;n}^{(k)}(x) = \mathbb{E}[\hat{f}_{X;n}^{(k)}(x)], \quad r_{\psi;n}^{(k)}(x) = \mathbb{E}[\hat{r}_{\psi;n}^{(k)}(x)].$$

Pour les estimateurs de la régression et sa dérivée, nous posons,

$$\begin{aligned} \tilde{\mathbb{E}}[\hat{m}_{\psi;n}(x)] &:= \frac{r_{\psi;n}(x)}{f_{X;n}(x)} \\ \tilde{\mathbb{E}}[\hat{m}'_{\psi;n}(x)] &:= \frac{r'_{\psi;n}(x)}{f_{X;n}(x)} - \frac{r_{\psi;n}(x)f'_{X;n}(x)}{f_{X;n}^2(x)}, \end{aligned}$$

en procédant identiquement pour les dérivées successives de $\hat{m}_{\psi;n}(\cdot)$ jusqu'à l'ordre k .

Remarque 2.2.5 Notons que l'espérance de $\hat{m}_{\psi;n}^{(k)}(x)$ ne coïncide pas avec l'approximation ci-dessus. Toutefois leur différence est négligeable et asymptotiquement nulle, via un argument similaire à la démonstration de la proposition 1.3.3 (voir la section A.7 en annexe). De l'autre côté, cette approximation permet la linéarisation de la déviation par rapport à l'espérance modifiée :

$$\hat{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)],$$

et son expression comme une fonctionnelle linéaire du processus empirique. Cet argument est un élément essentiel de la démonstration du théorème 2.3.2 ci-dessous (cf. lemme 2.4.8 en fin de preuve). L'idée est d'exprimer la déviation $\hat{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)]$ en fonction des autres déviations,

$$\hat{r}_{\psi;n}^{(l)}(x) - r_{\psi;n}^{(l)}(x) \quad \text{et} \quad \hat{f}_{X;n}^{(l)}(x) - f_{X;n}^{(l)}(x), \quad 0 \leq l \leq k.$$

2.3 Théorèmes

Les lois concernant la déviation maximale de nos estimateurs sont obtenues à partir d'un théorème limite général énoncé dans le théorème 2.3.1 ci-dessous. Pour cela, nous introduisons le processus suivant, étant donné deux fonctions $c(\cdot)$ et $d(\cdot)$ supposées continues et bornées sur J , nous posons, pour tout $x \in J$,

$$\begin{aligned} W_{n,k}(x, \psi) &= \sum_{i=1}^n \left(c(x)\psi(Y_i) + d(x) \right) K^{(k)}\left(\frac{x - X_i}{h_n}\right) \\ &\quad - n \mathbb{E} \left\{ \left(c(x)\psi(Y) + d(x) \right) K^{(k)}\left(\frac{x - X}{h_n}\right) \right\}. \end{aligned} \tag{2.3}$$

Théorème 2.3.1 *Supposons (F.1-3), (H.1-2), (K.1-4). Lorsque $n \rightarrow \infty$, nous avons,*

$$\left| \left\{ 2nh_n \log(1/h_n) \right\}^{-1/2} \sup_{x \in I} \left\{ \pm W_{n,k}(x, \psi) \right\} - \sigma_W(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_W^2(I) = \sup_{x \in I} \mathbb{E} \left[\left(c(x)\psi(Y) + d(x) \right)^2 \middle| X = x \right] f_X(x) \int_{\mathbf{R}} [K^{(k)}(t)]^2 dt. \quad (2.4)$$

Supposons (F.1-3), (H.2-4-5), (K.1-4). Alors, nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ 2nh_n \log(1/h_n) \right\}^{-1/2} \sup_{x \in I} \left\{ \pm W_{n,k}(x, \psi) \right\} - \sigma_W(I) \right| \stackrel{p.s.}{=} o(1).$$

Corollaire 2.3.1 *Supposons (F.2), (H.1-3), (K.1-3). Lorsque $n \rightarrow \infty$, nous avons,*

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{f}_{X;n}^{(k)}(x) - f_{X;n}^{(k)}(x) \right\} - \sigma_f(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_f^2(I) = \sup_{x \in I} \left\{ f_X(x) \right\} \int_{\mathbf{R}} [K^{(k)}(t)]^2 dt.$$

Supposons (F.2), (H.3-5), (K.1-3). Alors, nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{f}_{X;n}^{(k)}(x) - f_{X;n}^{(k)}(x) \right\} - \sigma_f(I) \right| = o(1), \quad \text{presque sûrement.}$$

Corollaire 2.3.2 *Supposons (F.1-3), (H.1-3), (K.1-4). Lorsque $n \rightarrow \infty$, nous avons,*

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{r}_{\psi;n}^{(k)}(x) - r_{\psi;n}^{(k)}(x) \right\} - \sigma_r(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_r^2(I) = \sup_{x \in I} \left\{ \sigma_{\psi}^2(x) f_X(x) \right\} \int_{\mathbf{R}} [K^{(k)}(t)]^2 dt.$$

Supposons (F.1-3), (H.3-5), (K.1-4). Alors, nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{r}_{\psi;n}^{(k)}(x) - r_{\psi;n}^{(k)}(x) \right\} - \sigma_r(I) \right| = o(1), \quad \text{presque sûrement.}$$

Théorème 2.3.2 *Supposons (F.1-3), (H.1-3), (K.1-4). Lorsque $n \rightarrow \infty$, nous avons,*

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)] \right\} - \sigma_m(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_m^2(I) = \sup_{x \in I} \left\{ \frac{\sigma_\psi^2(x)}{f_X(x)} \right\} \int_{\mathbf{R}} [K^{(k)}(t)]^2 dt. \quad (2.5)$$

Supposons (F.1-3), (H.3-5), (K.1-4). Alors, nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)] \} - \sigma_m(I) \right| = o(1), \quad \text{presque sûrement.}$$

Ces théorèmes et corollaires ont été obtenus par Deheuvels et Mason (2004) [29] lorsque $k = 0$ pour la convergence en probabilité. Einmahl et Mason (2000) [42] ont également démontré les théorèmes 2.3.1 et 2.3.2 pour le mode de convergence presque sûre et $k = 0$.

2.4 Démonstration des théorèmes

La démonstration du théorème principal se divise en deux parties : la borne supérieure et la borne inférieure. Ce schéma est classique pour l'obtention de lois du type loi du logarithme itéré. La borne supérieure repose sur deux inégalités exponentielles, dont l'inégalité de Bernstein, et est scindée en deux sous-parties : discrétisation et oscillation. La borne inférieure est obtenue en approchant le processus empirique par un processus de Poisson (cf. [35] et [26] pour un exposé plus détaillé de cette méthodologie). Les résultats seront démontrés uniquement pour la convergence en probabilité qui est suffisante pour les applications statistiques. Le passage à la convergence presque sûre s'effectue en choisissant une sous-suite de nature géométrique du type $n_j = 2^j$ (cf. “*blocking argument*”), en combinaison avec le fameux lemme de Borel-Cantelli.

2.4.1 Borne supérieure

Le but de cette sous-section est de prouver que, $\forall \varepsilon > 0$, nous avons

$$\mathbb{P} \left\{ \sup_{x \in I} \frac{|W_{n,k}(x, \psi)|}{\sqrt{2nh_n \log(1/h_n)}} > (1 + \varepsilon) \sigma_W(I) \right\} = o(1), \quad (2.6)$$

avec $\sigma_W(I)$ définie en (2.4). La démonstration de ce résultat en probabilité sera divisée en deux parties : discrétisation et oscillation.

Discrétisation

Premièrement, nous examinons le comportement du processus $W_{n,k}(\cdot, \psi)$ pour un nombre fini de points appartenant à l'intervalle $I = [a, b]$. Ce procédé permet de ramener le supremum sur I à un maximum sur un nombre fini de points. (On note que ce procédé analytique est utilisé de manière systématique lors de démonstrations de résultats de nature uniforme.) Pour cela, nous exprimons $W_{n,k}(\cdot, \psi)$ comme un processus empirique fonctionnel, c'est à dire indexé par une classe de fonctions. La classe de fonctions sera elle-même indexée par les points de la discrétisation.

Soit $\alpha_n(\cdot)$ le processus empirique bivarié fondé sur les n couples de variables aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$ et indexé par une classe \mathcal{G} de fonctions $g : \mathbb{R}^2 \rightarrow \mathbb{R}$. Par définition, pour $g \in \mathcal{G}$, nous posons,

$$\alpha_n(g) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(g(X_i, Y_i) - \mathbb{E}[g(X_i, Y_i)] \right). \quad (2.7)$$

Pour une classe de fonctions \mathcal{G} arbitraire, nous désignons par

$$\|n^{1/2}\alpha_n\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |n^{1/2}\alpha_n(g)|,$$

la norme maximale de $n^{1/2}\alpha_n$ sur la classe \mathcal{G} .

Pour n'importe quelle fonction $\psi(\cdot)$ mesurable et bornée, on introduit la fonction suivante, pour tout $x \in J$,

$$\eta_{n,x,k}(u, v) := (c(x)\psi(v) + d(x))K^{(k)}\left(\frac{x-u}{h_n}\right), \quad \text{pour } u, v \in \mathbb{R}. \quad (2.8)$$

Ainsi, d'après les précédentes définitions en (2.3), (2.7) et (2.8), nous pouvons écrire

$$W_{n,k}(x, \psi) = n^{1/2}\alpha_n(\eta_{n,x,k}). \quad (2.9)$$

L'étude du comportement limite de $W_{n,k}(x, \psi)$ sur l'intervalle I se réduit donc à l'étude du comportement limite du processus empirique $\alpha_n(\cdot)$ indexé par la classe de fonctions $\{\eta_{n,x,k} : x \in I\}$.

Pour n'importe quelle fonction ϕ à valeurs réelles définie sur un ensemble $B \subseteq \mathbb{R}$, on pose $\|\phi\|_B := \sup_{z \in B} |\phi(z)|$, et lorsque $B = \mathbb{R}$, on écrit simplement $\|\phi\|_B = \|\phi\|$. On désigne par $[u]$ la partie entière de u , telle que $[u] \leq u < [u] + 1$.

Après ces quelques notations préliminaires, nous pouvons commencer la discrétisation. Soit $0 < \delta < 1$ fixé, nous allons diviser l'intervalle $I = [a, b]$ en segments de longueur δh_n (c'est à dire δh_n dénote le pas de discrétisation). Nous posons, pour chaque $n \geq 1$,

$$z_{n,i} = a + i\delta h_n, \quad 0 \leq i \leq l_n := \lfloor (b-a)/(\delta h_n) \rfloor. \quad (2.10)$$

Pour simplifier notre écriture, on définit

$$g_{n,i}(u, v) := \eta_{n,z_{n,i},k}(u, v), \quad 0 \leq i \leq l_n. \quad (2.11)$$

Nous considérons alors le processus empirique indexé par la classe de fonctions suivante, pour $n \geq 1$,

$$\mathcal{G}_n := \left\{ g_{n,i} : 0 \leq i \leq l_n \right\}.$$

Nous pouvons maintenant travailler avec le processus empirique $\alpha_n(g)$ indexé par $g \in \mathcal{G}_n$ (c'est à dire la version discrétisée de notre processus) et montrer le résultat suivant.

Proposition 2.4.1 *Supposons que (2.10) soit vérifiée pour un certain $0 < \delta < 1/2$, alors pour tout $\tau > 0$, nous avons*

$$\mathbb{P}\left\{\max_{0 \leq i \leq l_n} \frac{|\alpha_n(g_{n,i})|}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I)(1 + \tau)\right\} = O(h_n^{\tau/2}) = o(1).$$

Cette proposition s'appuie sur une inégalité exponentielle de type Bernstein que nous rappelons ci-dessous. Notons que c'est l'inégalité de Bernstein qui nous permet de déterminer la constante ou borne limite $\sigma_W(I)$. La deuxième inégalité exponentielle en dimension infinie n'est pas assez précise mais nous permet de contrôler les incréments du processus. Par contre, l'inégalité de Bernstein à la bonne constante multiplicative gaussienne sur la droite (cf. (2.12) ci-dessous).

-Aparté : Inégalité de Bernstein et borne Gaussienne

Soit ξ une variable aléatoire gaussienne réelle $\mathcal{N}(0, 1)$ centrée réduite. Nous avons l'encadrement suivant :

$$\left\{\frac{1}{t} - \frac{1}{t^3}\right\} \frac{\exp\{-t^2/2\}}{\sqrt{2\pi}} \leq \mathbb{P}\{\xi \geq t\} < \frac{1}{t} \frac{\exp\{-t^2/2\}}{\sqrt{2\pi}}.$$

Le facteur important est $\exp\{-t^2/2\}$, car il implique une décroissance rapide de la queue de la distribution considérée. Nous remarquons également que l'inégalité de Bernstein est un cas particulier de l'inégalité de Bennett (cf. Pollard (1984) [112], p. 191–193). Ces inégalités exponentielles sont applicables lorsque les variables aléatoires considérées sont bornées, ce qui justifie notre emploi de (F.3).

Nous présentons la version maximale de l'inégalité de Bernstein, conséquence directe du lemme 2.2, p. 1393, [39].

Résultat 2.4.1 *Soient Z_1, \dots, Z_n des variables aléatoires centrées de variance identique $0 < \sigma^2 < \infty$. De plus, nous supposons qu'il existe un certain $M > 0$ tel que $|Z_r| < M$, $r = 1, \dots, n$. Alors, pour tout réel $t > 0$, nous avons*

$$\mathbb{P}\left\{Z_1 + \dots + Z_n > t\sqrt{n}\right\} \leq \exp\left\{-\frac{3t^2}{6\sigma^2 + 2Mn^{-1/2}t}\right\}. \quad (2.12)$$

Le choix naturel pour les Z_r est

$$Z_r = g_{n,i}(X_r, Y_r) - \mathbb{E}[g_{n,i}(X_r, Y_r)], \quad r = 1, \dots, n.$$

Ces variables sont bien centrées et de même loi, vérifions qu'elles soient bornées. Pour tout $1 \leq r \leq n$, nous avons, via (F.3) et (K.4),

$$\begin{aligned} |Z_r| &= |g_{n,i}(X_r, Y_r) - \mathbb{E}[g_{n,i}(X_r, Y_r)]| \\ &\leq |g_{n,i}(X_r, Y_r)| + |\mathbb{E}[g_{n,i}(X_r, Y_r)]| \\ &\leq 2 \times \{\|c\| \times \|\psi\| + \|d\|\} \|K^{(k)}\| := M. \end{aligned} \quad (2.13)$$

Remarque 2.4.1 La fonction $K^{(k)}$ est à variation bornée, ce qui implique $\|K^{(k)}\| < \infty$ clairement (cf. Natanson (1955) [110] ou Schuster (1969), p. 1188, [119]).

Il reste à contrôler la variance des Z_r , $\forall 1 \leq r \leq n$,

$$\begin{aligned} \text{Var} [Z_r] &= \text{Var} [g_{n,i}(X, Y) - \mathbb{E}[g_{n,i}(X, Y)]] \\ &= \text{Var}[g_{n,i}(X, Y)] \\ &\leq \mathbb{E}[g_{n,i}^2(X, Y)]. \end{aligned}$$

En utilisant un argument de conditionnement, nous pouvons majorer ce terme en faisant apparaître la variance conditionnelle, terme crucial lors de démonstrations de lois limites uniformes du logarithme (et de lois du logarithme itéré),

$$\sigma_W^2(I) = \sup_{x \in I} \mathbb{E} \left[\left(c(x)\psi(Y) + d(x) \right)^2 \middle| X = x \right] f_X(x) \int_{\mathbf{R}} [K^{(k)}(v)]^2 dv.$$

Nous avons, via (F.1) et (K.2),

$$\begin{aligned} \mathbb{E}[g_{n,i}^2(X, Y)] &= \int_{\mathbf{R}} \mathbb{E} \left[\left(c(z_{n,i})\psi(Y) + d(z_{n,i}) \right)^2 \middle| X = t \right] f_X(t) \left\{ K^{(k)} \left(\frac{z_{n,i} - t}{h_n} \right) \right\}^2 dt \\ &\stackrel{(F.1)}{\leq} \frac{\sigma_W^2(I)}{\int_{\mathbf{R}} [K^{(k)}(v)]^2 dv} \int_{|z_{n,i} - t| \leq h_n/2} \frac{f_X(t)}{f_X(z_{n,i})} \times \left\{ K^{(k)} \left(\frac{z_{n,i} - t}{h_n} \right) \right\}^2 dt \\ &\leq \frac{\sigma_W^2(I)h_n}{\int_{-1/2}^{1/2} [K^{(k)}(v)]^2 dv} \int_{-1/2}^{1/2} \frac{f_X(z_{n,i} - h_n u)}{f_X(z_{n,i})} \times \left\{ K^{(k)}(u) \right\}^2 du, \end{aligned}$$

ce qui, d'après (F.2) et le développement limité de la fonction $f_X(\cdot)$ autour du point $z_{n,i}$, nous donne finalement

$$\text{Var} [Z_r] \leq \sigma_W^2(I)h_n + o(h_n).$$

Notons que c'est l'hypothèse (F.2) qui est fondamentale pour obtenir la borne sur la variance. Ainsi, quelque soit $\tau > 0$, lorsque n est suffisamment grand, il s'ensuit la borne suivante :

$$\max_{0 \leq i \leq l_n} \text{Var} [g_{n,i}(X, Y)] \leq \sigma_W^2(I) (1 + \tau) h_n. \quad (2.14)$$

Nous pouvons alors appliquer le résultat 2.4.1 avec $t = \sigma_W(I)(1 + \tau)\sqrt{2h_n \log(1/h_n)}$ et conclure que, lorsque n est suffisamment grand, via (2.14),

$$\begin{aligned} &\mathbb{P} \left\{ \max_{0 \leq i \leq l_n} \frac{|\alpha_n(g_{n,i})|}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I)(1 + \tau) \right\} \\ &\leq 2(l_n + 1) \exp \left\{ - \frac{2\sigma_W^2(I)(1 + \tau)h_n \log(1/h_n)}{2\sigma_W^2(I)h_n + \frac{2}{3}M\sigma_W(I) \times \sqrt{\frac{2h_n \log(1/h_n)}{n}}} \right\}. \end{aligned}$$

Nous avons d'après (H.2) ou (H.3) lorsque $k = 0$,

$$\lim_{n \rightarrow \infty} \frac{nh_n}{\log(1/h_n)} = \infty.$$

Il s'ensuit

$$\frac{\log(1/h_n)}{n} = o(h_n),$$

ainsi le deuxième terme du dénominateur est négligeable asymptotiquement. Nous obtenons, en conséquence, pour n suffisamment grand,

$$\begin{aligned} \mathbb{P}\left\{ \max_{0 \leq i \leq l_n} \frac{|\alpha_n(g_{n,i})|}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I)(1 + \tau) \right\} &\leq 2(l_n + 1) \exp\left\{ - (1 + \tau) \log(1/h_n) \right\} \\ &\leq 2(l_n + 1)h_n^{1+\tau/2} \\ &= O(h_n^{\tau/2}), \end{aligned}$$

en utilisant (2.10), i.e. $l_n = O(h_n^{-1})$. □

Nous avons ainsi montré la borne supérieure pour le processus $W_{n,k}(\cdot, \cdot)$ sur le maillage représenté par la classe de fonctions \mathcal{G}_n .

Remarque 2.4.2 Nous présentons brièvement la méthodologie pour passer de la convergence en probabilité à la convergence presque sûre. Comme les arguments sont toujours similaires, ils ne seront pas répétés dans les démonstrations. L'hypothèse additionnelle (H.5) sur la fenêtre nous servira à contrôler des séries en vue d'une application du lemme de Borel-Cantelli et l'hypothèse (H.4) de monotonie des suites h_n et nh_n est utile pour des détails techniques. Le lemme de Borel-Cantelli est un outil classique lors des démonstrations de résultats presque sûrs. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité quelconque. Nous désignons par $\{A_n : n \geq 1\} \subseteq \mathcal{A}$ une suite d'événements mesurables. On pose

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} \left\{ \bigcup_{m \geq n} A_m \right\} \quad \text{et} \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} \left\{ \bigcap_{m \geq n} A_m \right\}.$$

Soit $\bar{A} = \Omega - A$ le complémentaire de A .

Résultat 2.4.2 Lemme de Borel-Cantelli

Pour toute suite $\{A_n : n \geq 1\} \subseteq \mathcal{A}$ d'événements mesurables, nous avons

$$\sum_{i=1}^n \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0 \Leftrightarrow \mathbb{P}(\liminf_{n \rightarrow \infty} \bar{A}_n) = 1.$$

Lorsque les événements A_n sont indépendants, nous avons également

$$\sum_{i=1}^n \mathbb{P}(A_n) = \infty \Rightarrow \mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1 \Leftrightarrow \mathbb{P}(\liminf_{n \rightarrow \infty} \bar{A}_n) = 0.$$

Intuitivement, cela signifie que si la somme des probabilités pour qu'un événement arrive à l'instant n , pour $n \in \mathbb{N}$, tend vers l'infini, alors l'événement a une probabilité 1 d'avoir lieu une infinité de fois. De l'autre côté, si la somme des probabilités converge, l'événement a une probabilité 1 d'avoir lieu un nombre fini de fois.

On choisit une sous-suite géométrique $n_r = 2^r$ ou $n_r = \lfloor \lambda^r \rfloor$ avec $\lambda > 1$. Posons :

$$z_{r,i} = a + i\delta h_{n_r}, \quad 0 \leq i \leq l_r := \lfloor (b-a)/(\delta h_{n_r}) \rfloor,$$

et

$$g_{r,i}(u, v) := \eta_{n_r, z_{r,i}, k}, \quad 0 \leq i \leq l_r.$$

En reprenant les arguments précédents, on obtient

$$\begin{aligned} & \mathbb{P} \left\{ \max_{0 \leq i \leq l_r} \max_{n_{r-1} < n \leq n_r} \frac{|\alpha_n(g_{r,i})|}{\sqrt{2h_{n_r} \log(1/h_{n_r})}} > \sigma_W(I)(1 + \tau) \right\} \\ & \leq 2(l_r + 1) \exp \left(- (1 + \tau) \log(1/h_{n_r}) \right) \\ & \leq 2(l_r + 1) h_{n_r}^{1+\tau/2} \\ & = O(h_{n_r}^{\tau/2}). \end{aligned} \tag{2.15}$$

Nous rappelons la condition (H.5) sur la fenêtre

$$\frac{|\log h_n|}{\log \log n} \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty.$$

Cette hypothèse nous assure que

$$\sum_{k=1}^{\infty} h_{n_r}^{\tau/2} < \infty.$$

La convergence de la série ci-dessus combinée avec la borne (2.15) implique

$$\limsup_{r \rightarrow \infty} \max_{0 \leq i \leq l_r} \max_{n_{r-1} < n \leq n_r} \frac{|\alpha_n(g_{r,i})|}{\sqrt{2h_{n_r} \log(1/h_{n_r})}} \leq \sigma_W(I)(1 + \tau) \quad \text{presque sûrement,}$$

via le lemme de Borel-Cantelli.

On note enfin que pour l'obtention d'inégalités presque sûres, l'inégalité maximale de Montgomery-Smith (A.22), présentée en annexe, s'avère un outil de choix (cf. [56]).

Oscillation

Il reste à étudier le comportement du processus $W_{n,k}(\cdot, \psi)$ entre les points du quadrillage. Nous cherchons à démontrer que les incréments sont négligeables sur les segments délimités par les couples de points $(z_{n,i}, z_{n,i+1})$, pour $0 \leq i \leq l_n$. Par convention, nous posons $z_{n, l_n+1} := b$ afin d'être sûr de couvrir exactement l'intervalle I .

Nous considérons le processus empirique indexé par la classe de fonctions suivante :

$$\mathcal{G}'_{n,i} := \left\{ g_{n,i} - \eta_{n,z,k} : z_{n,i} \leq z \leq z_{n,i+1} \right\}.$$

Cette classe permet l'étude de l'oscillation maximale sur un des intervalles engendrés par la discrétisation. Par la suite, nous constaterons que la partie due à l'oscillation est négligeable. Le principal outil de la démonstration est une remarquable inégalité exponentielle pour le processus empirique indexé par des classes de fonctions démontrée par Talagrand (1994) [137], combinée à une borne pour la norme L_1 du processus empirique symétrisé et indexé par des classes de fonctions de type Vapnik-Chervonenkis (cf. [42] et les résultats A.3.2 et A.3.4 en annexe).

Proposition 2.4.2 *Il existe une constante $A > 0$ telle que, quelque soit $\epsilon > 0$, nous pouvons trouver un δ_ϵ vérifiant (2.10) avec $0 < \delta < \delta_\epsilon$ de sorte que*

$$\mathbb{P} \left\{ \max_{0 \leq i \leq l_n} \frac{\|\alpha_n\|_{\mathcal{G}'_{n,i}}}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I) A \sqrt{\epsilon} \right\} = o(1).$$

La démonstration de la proposition 2.4.2 est basée sur quatre lemmes.

Comme $K^{(k)}$ est continue et à variation bornée sur \mathbb{R} , nous avons la décomposition suivante $K^{(k)} =: K_1^{(k)} - K_2^{(k)}$, avec $K_1^{(k)}$ et $K_2^{(k)}$ deux fonctions croissantes, continues et à variations bornées sur \mathbb{R} telles que $|K^{(k)}|_v = |K_1^{(k)}|_v + |K_2^{(k)}|_v$. On note que cette décomposition est directement liée à la définition de la notion de variation totale d'une fonction, elle reste notamment valable pour une fonction multivariée.

Lemme 2.4.1 *Nous supposons les hypothèses (K.1-2) vérifiées. Soit $0 < \delta < 1/2$ arbitraire. Nous obtenons, uniformément en $z_1, z_2 \in I$ vérifiant $|z_1 - z_2| \leq \delta h_n$,*

$$\mathbb{E} \left| K^{(k)} \left(\frac{z_2 - X}{h_n} \right) - K^{(k)} \left(\frac{z_1 - X}{h_n} \right) \right|^2 \leq C_{1,\delta} h_n,$$

où

$$C_{1,\delta} := |K^{(k)}|_v^2 \|f_X\| \delta.$$

De plus, pour n suffisamment grand, nous pouvons remplacer $\|f_X\|$ par $\|f_X\|_J$ ci-dessus.

Remarquons que

$$\begin{aligned} & \left| K^{(k)} \left(\frac{z_2 - X}{h_n} \right) - K^{(k)} \left(\frac{z_1 - X}{h_n} \right) \right| = \left| \int_{(z_1 - X)/h_n}^{(z_2 - X)/h_n} dK^{(k)}(y) \right| \\ & \leq \int_{\mathbb{R}} \left| \mathbb{I} \left\{ \frac{z_2 - X}{h_n} > y \right\} - \mathbb{I} \left\{ \frac{z_1 - X}{h_n} > y \right\} \right| d\{K_1^{(k)}(y) + K_2^{(k)}(y)\}. \end{aligned}$$

Ainsi, via l'inégalité de Hölder, nous avons la borne suivante

$$\mathbb{E} \left[\left| K^{(k)} \left(\frac{z_2 - X}{h_n} \right) - K^{(k)} \left(\frac{z_1 - X}{h_n} \right) \right|^2 \right]$$

$$\begin{aligned}
&\leq |K^{(k)}|_v \int_{\mathbf{R}} \mathbb{E} \left| \mathbb{I} \left\{ \frac{z_2 - X}{h_n} > y \right\} - \mathbb{I} \left\{ \frac{z_1 - X}{h_n} > y \right\} \right| d\{K_1^{(k)}(y) + K_2^{(k)}(y)\} \\
&\leq |K^{(k)}|_v \int_{\mathbf{R}} \left| \int_{z_2 - h_n y}^{z_1 - h_n y} f_X(x) dx \right| d\{K_1^{(k)}(y) + K_2^{(k)}(y)\} \\
&\leq |K^{(k)}|_v \|f_X\| \times \{|K^{(k)}|_v \delta h_n\}, \tag{2.16}
\end{aligned}$$

ce qui nous donne la première partie du lemme. La fonction $K^{(k)}(\cdot)$ est à support compact. Dans l'avant-dernière inégalité, la variable y est donc bornée. En conséquence, lorsque n est suffisamment grand (i.e. h_n suffisamment petit), on peut remplacer $\|f_X\|$ par $\|f_X\|_J$ dans (2.16). \square

Après ce lemme technique, on cherche une borne pour la variance du processus empirique, indexé par la classe de fonctions $\mathcal{G}'_{n,i}$, qui dépende de δh_n . En d'autres termes, il sera démontré que la variance du processus empirique engendré par les incréments est arbitrairement petite.

Lemme 2.4.2 *Lorsque (2.10) est vérifiée pour $0 < \delta < 1/2$, nous avons, uniformément en $0 \leq i \leq l_n$, $z \in I$ satisfaisant $z_{n,i} \leq z \leq z_{n,i+1}$, et $g_{n,i} - \eta_{n,z,k} \in \mathcal{G}'_{n,i}$,*

$$\mathbb{E} \left[(g_{n,i}(X, Y) - \eta_{n,z,k}(X, Y))^2 \right] \leq C_{2,\delta} h_n, \tag{2.17}$$

où

$$C_{2,\delta} = 4\beta \|f_X\|_J \left\{ \omega_c^2(\delta h_n) \vee \omega_d^2(\delta h_n) \right\} \|K^{(k)}\|^2 + \{ \|c\|_J^2 \vee \|d\|_J^2 \} C_{1,\delta},$$

avec

$$\omega_\phi(\delta) := \sup \{ |\phi(x) - \phi(y)| : |x - y| \leq \delta; \text{ et } x, y \in I \},$$

$$\text{et } \beta := \|\psi\|^2 + 1 < \infty.$$

D'après les définitions (2.8), (2.11) et via l'inégalité $(a + b)^2 \leq 2(a^2 + b^2)$, nous obtenons

$$\begin{aligned}
&\mathbb{E} \left[(g_{n,i}(X, Y) - \eta_{n,z,k}(X, Y))^2 \right] = \mathbb{E} \left[(\eta_{n,z_{n,i},k}(X, Y) - \eta_{n,z,k}(X, Y))^2 \right] \\
&= \mathbb{E} \left[\left\{ (c(z_{n,i})\psi(Y) + d(z_{n,i}))K^{(k)}\left(\frac{z_{n,i} - X}{h_n}\right) - (c(z)\psi(Y) + d(z))K^{(k)}\left(\frac{z - X}{h_n}\right) \right\}^2 \right] \\
&\leq 2\mathbb{E} \left[(\{c(z_{n,i}) - c(z)\}\psi(Y) + \{d(z_{n,i}) - d(z)\})^2 \left\{ K^{(k)}\left(\frac{z_{n,i} - X}{h_n}\right) \right\}^2 \right] \\
&+ 2\mathbb{E} \left[(c(z)\psi(Y) + d(z))^2 \left\{ K^{(k)}\left(\frac{z_{n,i} - X}{h_n}\right) - K^{(k)}\left(\frac{z - X}{h_n}\right) \right\}^2 \right] \\
&\leq 4\beta \{ \omega_c^2(\delta h_n) \vee \omega_d^2(\delta h_n) \} \times \mathbb{E} \left[K^{(k)}\left(\frac{z_{n,i} - X}{h_n}\right)^2 \right] \\
&+ 4\beta \{ \|c\|_J^2 \vee \|d\|_J^2 \} \times \mathbb{E} \left[\left\{ K^{(k)}\left(\frac{z_{n,i} - X}{h_n}\right) - K^{(k)}\left(\frac{z - X}{h_n}\right) \right\}^2 \right] =: (I) + (II).
\end{aligned}$$

Le noyau $K^{(k)}$ étant à support compact d'après (K.2), nous avons aisément

$$K^{(k)}\left(\frac{z_{n,i} - X}{h_n}\right)^2 \leq \|K^{(k)}\|^2 \times \mathbb{I}\{|X - z_{n,i}| \leq h_n/2\}.$$

En passant à l'espérance et en effectuant un changement de variable classique, on obtient alors une borne pour (I), uniformément en $0 \leq i \leq l_n$. Pour n suffisamment grand,

$$(I) \leq 4\beta \{ \omega_c^2(\delta h_n) \vee \omega_d^2(\delta h_n) \} \times \|K^{(k)}\|^2 h_n \|f_X\|_J. \quad (2.18)$$

Afin de borner (II), nous inférons du précédent lemme 2.4.1 que

$$(II) \leq 4\beta \{ \|c\|_J^2 \vee \|d\|_J^2 \} \times C_{1,\delta} h_n. \quad (2.19)$$

En combinant (2.18) et (2.19), il s'ensuit

$$(I) + (II) \leq C_{2,\delta} h_n,$$

ce qui clôt la démonstration. \square

Pour simplifier les notations, nous posons, pour $0 \leq i \leq l_n$,

$$\sigma_i^2(\psi) = \sup \left\{ \text{Var} [g(X, Y)] : g \in \mathcal{G}'_{n,i} \right\}.$$

Une application directe du lemme 2.4.2 implique le lemme ci-dessous.

Lemme 2.4.3 *Pour $\epsilon > 0$ fixé, on peut trouver δ_ϵ tel que, pour n suffisamment grand,*

$$\max_{0 \leq i \leq l_n} \sigma_i^2(\psi) \leq \epsilon h_n \sigma_W^2(I),$$

lorsque (2.10) est vérifiée avec $0 < \delta \leq \delta_\epsilon$.

D'après l'inégalité (2.17), nous avons

$$\begin{aligned} \text{Var} [g_{n,i}(X, Y) - \eta_{n,z,k}(X, Y)] &\leq \mathbb{E} \left[(g_{n,i}(X, Y) - \eta_{n,z,k}(X, Y))^2 \right] \\ &\leq C_{2,\delta} h_n. \end{aligned}$$

Il suffit de choisir δ_ϵ suffisamment petit, tel que

$$C_{2,\delta_\epsilon} \leq \epsilon \sigma_W^2(I),$$

ceci étant possible grâce à la continuité des fonctions $c(\cdot)$ et $d(\cdot)$ notamment. \square

Nous avons donc démontré que sur chaque sous-intervalle symbolisé par la classe de fonctions $\mathcal{G}'_{n,i}$, nous avons une borne pour la variance, uniformément en $0 \leq i \leq l_n$. Nous introduisons alors

$$\mathcal{G}'_n := \bigcup_{0 \leq i \leq l_n} \mathcal{G}'_{n,i},$$

la classe qui recouvre totalement l'intervalle $I = [a, b]$ ou qui considère tous les incréments possibles sur I suivant la discrétisation précédente.

Il faut que la classe de fonctions \mathcal{G}'_n satisfasse une condition d'entropie, afin de pouvoir appliquer efficacement l'inégalité exponentielle de Talagrand (cf. résultat A.3.1 situé en annexe), c'est à dire pour satisfaire les conditions du résultat A.3.2.

Une classe de fonctions \mathcal{G} vérifie la condition $[\mathcal{E}]$ si :

- i) il existe une fonction enveloppe $G(\cdot)$ à valeurs finies satisfaisant $G(x) \geq \sup_{g \in \mathcal{G}} |g(x)|$ pour tout $x \in \mathcal{X}$;
- ii) pour certaines constantes $C_0 > 0$ et $\nu > 0$,

$$N(\epsilon, \mathcal{G}) \leq C_0 \epsilon^{-\nu}, \quad 0 < \epsilon < 1 \quad (\text{condition d'entropie}). \quad (2.20)$$

On se réfère à l'annexe, définition A.3.3, pour une définition précise du nombre de recouvrement $N(\epsilon, \mathcal{G})$. Si la classe \mathcal{G} vérifie la condition $[\mathcal{E}]$, elle est aussi appelée classe de fonctions à nombre de recouvrement uniformément polynomial.

La condition d'entropie (2.20) est vérifiée pour des classes de fonctions particulières, appelées *Vapnik-Chervonenkis graph class* (VCGC) ou classes de graphes VC (CGVC). Comme pour les ensembles, le nombre de recouvrement des classes de fonctions VC a une vitesse de croissance polynomiale.

Définition 2.4.1 Le **graphe** G_f d'une fonction $f : \mathcal{X} \rightarrow \mathbb{R}$ mesurable est le sous-ensemble de $\mathcal{X} \times \mathbb{R}$ défini par

$$G_f := \{(x, t) \in \mathcal{X} \times \mathbb{R} : 0 \leq t \leq f(x) \text{ ou } f(x) \leq t \leq 0\}.$$

Définition 2.4.2 Une collection \mathcal{F} de fonctions mesurables est appelée une **classe de graphes VC**, si la collection de tous les graphes indexée par les fonctions $f \in \mathcal{F}$ forme une classe VC d'ensembles dans $\mathcal{X} \times \mathbb{R}$

On rappelle qu'une classe ou une collection d'ensembles mesurables est appelée une **classe VC** si son index VC est fini.

Remarque 2.4.3 Les auteurs Van der Vaart et Wellner parlent plutôt de "*between*" *graphs* qui forment une classe VC d'ensembles si et seulement si la classe est une *VC-subgraph class*.

Classe VC d'ensembles

Soit \mathcal{C} une collection de sous-ensembles d'un ensemble dénoté \mathcal{X} , i.e. avec $\mathcal{C} \subset 2^{\mathcal{X}}$. Un ensemble arbitraire de n points $\{x_1, \dots, x_n\}$ possède 2^n sous-ensembles. On dénote par $|F|$ le cardinal d'un ensemble F arbitraire. Une collection \mathcal{C} est appelée une **classe VC** (ou CVC), si elle vérifie la condition suivante :

$$\exists s \in \mathbb{N} \text{ tel que } \forall F \subseteq \mathcal{X} \text{ avec } |F| = s, \quad \Delta^{\mathcal{C}}(F) < 2^s, \quad (2.21)$$

où $\Delta^{\mathcal{C}}(F) := |\{F \cap C : C \in \mathcal{C}\}|$. La condition (2.21) signifie que la classe \mathcal{C} , ou plus généralement qu'une CVC, n'est pas trop riche d'un point de vue combinatoire. C'est à dire, pour tout ensemble $F \subseteq \mathcal{X}$ de cardinal s , il existe au moins un sous-ensemble $F' \subset F$ tel que $F' \neq F \cap C$ pour tout $C \in \mathcal{C}$.

Une autre formulation possible est : \mathcal{C} est une CVC lorsque,

$$\exists s \in \mathbb{N} \text{ tel que } m^{\mathcal{C}}(s) < 2^s,$$

avec $m^{\mathcal{C}}(n) := \max \{ \Delta^{\mathcal{C}}(F) : F \subseteq \mathcal{X}, |F| = n \}$, $n \in \mathbb{N}$. Alors le nombre $V(\mathcal{C}) := \min \{ s \in \mathbb{N} : m^{\mathcal{C}}(s) < 2^s \}$ est appelé l'**index de Vapnik-Cervonenkis** de la collection \mathcal{C} .

Dans le cas particulier où $\mathcal{X} = \mathbb{R}$ et $\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}$, (2.21) est vérifiée dès lors que $s = 2$. Donc l'index VC de la collection de fonctions \mathcal{C} est de deux. La justification est immédiate :

$$\forall F = \{x_1, x_2\}, x_1 < x_2 \implies F' = \{x_2\} \neq F \cap (-\infty, t], \forall t \in \mathbb{R}.$$

Comme deuxième exemple dans \mathbb{R} , nous considérons la collection $\mathcal{C} = \{(a, b] : (a, b) \in \mathbb{R}^2\}$ de tous les intervalles de la forme $(a, b]$. L'index VC est alors de trois,

$$\forall F = \{x_1, x_2, x_3\}, x_1 < x_2 < x_3 \implies F' = \{x_1, x_3\} \neq F \cap (a, b], \forall (a, b) \in \mathbb{R}^2.$$

Ces deux exemples se généralisent dans \mathbb{R}^d avec comme index VC, $d + 1$ et $2d + 1$, respectivement.

De l'autre coté, une collection d'ensembles \mathcal{C} n'est pas une CVC si,

$$\forall n \in \mathbb{N} \exists F \subseteq \mathcal{X} \text{ avec } |F| = n, \quad \text{tel que } \Delta^{\mathcal{C}}(F) = 2^n.$$

D'après le théorème 2.6.7, p. 141, [145], énoncé ci-dessous, nous savons qu'une CVC de fonctions \mathcal{F} , munie d'une fonction enveloppe mesurable $F(\cdot)$, vérifie bien la condition d'entropie $[\mathcal{E}]$.

Théorème 2.4.1 *Pour une CVC de fonctions munie d'une fonction enveloppe mesurable F et $r \geq 1$, on a, pour toute mesure de probabilité Q telle que $\|F\|_{Q,r} > 0$,*

$$N(\epsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})} \left\{ \frac{1}{\epsilon} \right\}^{r(V(\mathcal{F})-1)},$$

pour une constante universelle K et $0 < \epsilon < 1$.

Ce théorème est apparemment une version du lemme 2.7 de Alexander (1984) [2], confer également le lemme 25, section II.5, de Pollard (1984) [112]. Notons qu'il existe plusieurs variantes dans la littérature pour la démonstration de ce type de résultat. Pour notre part, l'essentiel est de retenir qu'une classe de graphe VC est à nombre de recouvrement polynomial.

Remarque 2.4.4 Le théorème 2.4.1 sera appliqué dans nos travaux pour le choix particulier $r = 2$ correspondant à la distance L_2 .

Nous énonçons à présent un lemme très utile qui permet de caractériser rapidement une CGVC de fonctions.

Lemme 2.4.4 *Un espace vectoriel \mathcal{F} de dimension finie, composé de fonctions mesurables $f : \mathcal{X} \rightarrow \mathbb{R}$, est un CGVC d'index inférieur ou égal à $\dim(\mathcal{F}) + 2$.*

Voir, par exemple, la démonstration du lemme 2.6.15, p. 146, dans Van der Vaart et Wellner (1996) [145]. Pour des références plus précises, voir p. 271, [145]. \square

Soient les classes de fonctions

$$\mathbb{H} = \{u + z : z \in \mathbb{R}\},$$

$$\mathbb{K}_1 = \{K_1(u + z) : z \in \mathbb{R}\},$$

$$\mathbb{K}_2 = \{K_2(u + z) : z \in \mathbb{R}\}.$$

En appliquant le lemme 2.4.4 ci-dessus pour la classe de graphes \mathbb{H} puis le lemme 2.6.18, partie (viii), p. 147 de ([145]) pour \mathbb{K}_1 et \mathbb{K}_2 fonctions monotones, nous obtenons que \mathbb{H} , \mathbb{K}_1 et \mathbb{K}_2 sont des CGVC. Ce dernier point confirme que la condition (K.1), c'est à dire $K(\cdot)$ à variation bornée sur \mathbb{R} , est probante pour obtenir la condition d'entropie. Par la suite une application directe du théorème 2.4.1 implique que les classes \mathbb{H} , \mathbb{K}_1 et \mathbb{K}_2 vérifient la condition $[\mathcal{E}]$. En conséquence, comme $K = K_1 - K_2$, il s'ensuit

$$\mathbb{K} = \{K(u + z) : z \in \mathbb{R}\} \text{ satisfait } [\mathcal{E}].$$

A présent, considérons la classe de fonctions suivante

$$\mathcal{F} = \{a\psi(v) + b : |a| \leq C, |b| \leq C\},$$

où $C > 0$ borne les deux fonctions $c(\cdot)$ et $d(\cdot)$ dans (2.3). En utilisant les même arguments que précédemment nous concluons que \mathcal{F} vérifie $[\mathcal{E}]$. Alors, une simple application du lemme A.3.4, situé en annexe, entraîne que, pour une certaine constante $0 < C < \infty$, la classe produit

$$\{(a\psi(v) + b)K(u + z) : z \in \mathbb{R}, |a| \leq C, |b| \leq C\} \text{ satisfait } [\mathcal{E}].$$

Il en découle aisément que la classe de fonctions définie pour $u, v \in \mathbb{R}$ par

$$\begin{aligned} \mathcal{G}' &= \{(a\psi(v) + b)K(u + z) - (a'\psi(v) + b')K(u + z') : \\ & z \in \mathbb{R}, z' \in \mathbb{R}, |a| \leq C, |a'| \leq C, |b| \leq C, |b'| \leq C\} \text{ satisfait } [\mathcal{E}]. \end{aligned}$$

Comme $\mathcal{G}'_n \subseteq \mathcal{G}'$, la classe \mathcal{G}'_n , qui sert à contrôler les incréments du processus empirique, appartient bien à la classe des fonctions mesurables avec nombre de recouvrement uniformément polynomial. Cet argument va nous servir à démontrer le lemme central de cette section concernant les oscillations du processus empirique.

Lemme 2.4.5 *Sous les conditions du lemme 2.4.3, il existe une constante $B > 0$ telle que, lorsque (2.10) est vérifiée pour $0 < \delta \leq \delta_\epsilon$,*

$$\mathbb{P}\left\{\frac{\|\alpha_n\|_{\mathcal{G}'_n}}{\sqrt{h_n \log(1/h_n)}} > B\sqrt{\epsilon}\right\} = o(1). \quad (2.22)$$

On rappelle que $0 < M = 2\{\|c\| \times \|\psi\| + \|d\|\} \|K^{(k)}\| < \infty$ d'après (2.13). Nous remarquons alors que, uniformément en $g \in \mathcal{G}'_n$,

$$\|g\| \leq \|g_{n,i}\| + \|\eta_{n,z,k}\| \leq M.$$

De plus, le lemme 2.4.3 implique l'existence d'un $\delta_\epsilon > 0$ tel que, lorsque (2.10) est vérifiée pour $0 < \delta \leq \delta_\epsilon$, nous avons

$$\sigma_{\mathcal{G}'_n}^2 = \sup_{g \in \mathcal{G}'_n} \text{Var}[g(X, Y)] \leq \epsilon h_n \sigma_W^2(I). \quad (2.23)$$

Ainsi d'après le résultat A.3.1, $\forall t > 0$ et pour A_1, A_2 des constantes convenablement choisies, il s'ensuit

$$\begin{aligned} & \mathbb{P}\left\{\|n^{1/2}\alpha_n\|_{\mathcal{G}'_n} \geq A_1 \left(\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(X_i, Y_i)\right\|_{\mathcal{G}'_n} + t\right)\right\} \\ & \leq 2 \left\{ \exp\left\{\frac{-A_2 t^2}{n\sigma_{\mathcal{G}'_n}^2}\right\} + \exp\left\{\frac{-A_2 t}{M}\right\} \right\}. \end{aligned} \quad (2.24)$$

Ensuite, en utilisant le résultat A.3.2 (ou sa version améliorée, le résultat A.3.4) pour la classe de fonctions $\mathcal{G} = \mathcal{G}'_n$, la fonction enveloppe $G(\cdot) = \sup_{g \in \mathcal{G}'_n} |g(\cdot)|$ et la variance $\sigma^2 = \sigma_{\mathcal{G}'_n}^2$, nous obtenons la borne suivante

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(X_i, Y_i)\right\|_{\mathcal{G}'_n} \leq A_3 \sqrt{\nu \epsilon n h_n \log(1/h_n)}, \quad (2.25)$$

où $A_3 > 0$ désigne une constante. Ainsi, en combinant l'inégalité (2.24) avec (2.23) et (2.25), lorsque $t = A_3 \sqrt{\nu \epsilon n h_n \log(1/h_n)}$, il s'ensuit

$$\begin{aligned} \mathbb{P}\left\{\|n^{1/2}\alpha_n\|_{\mathcal{G}'_n} \geq 2A_1 A_3 \sqrt{\nu \epsilon n h_n \log(1/h_n)}\right\} & \leq 2 \left\{ \exp\left\{\frac{-A_2 A_3^2 \nu \log(1/h_n)}{\sigma_W^2(I)}\right\} \right. \\ & \quad \left. + \exp\left\{\frac{-A_2 A_3 \sqrt{\nu \epsilon n h_n \log(1/h_n)}}{M}\right\} \right\} \\ & = o(1), \quad \text{via (H.1) et (H.2)}. \end{aligned}$$

En posant $B = 2A_1 A_3 \sqrt{\nu}$ ci-dessus, nous concluons à (2.22), ce qui clôt la démonstration. \square

Démonstration de la proposition 2.4.2

Comme $\mathcal{G}'_{n,i} \subseteq \mathcal{G}'_n$, pour tout $0 \leq i \leq l_n$, nous avons

$$\max_{0 \leq i \leq l_n} \frac{\|\alpha_n\|_{\mathcal{G}'_{n,i}}}{\sqrt{2h_n \log(1/h_n)}} \leq \frac{\|\alpha_n\|_{\mathcal{G}'_n}}{\sqrt{2h_n \log(1/h_n)}}.$$

D'après (2.22),

$$\mathbb{P}\left\{\frac{\|\alpha_n\|g'_{n,i}}{\sqrt{h_n \log(1/h_n)}} > B\sqrt{\epsilon}\right\} = o(1).$$

Ainsi, en fixant $A = \frac{B}{\sigma_W(I)\sqrt{2}}$, nous obtenons

$$\mathbb{P}\left\{\max_{0 \leq i \leq l_n} \frac{\|\alpha_n\|g'_{n,i}}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I)A\sqrt{\epsilon}\right\} = o(1),$$

lorsque (2.10) est vérifiée pour $0 < \delta \leq \delta_\epsilon$.

Conclusion

Finalement, en combinant les propositions 2.4.1 et 2.4.2, nous concluons qu'il existe une constante $A > 0$, tel que pour n'importe quel $\epsilon > 0$, nous puissions trouver un $\delta_\epsilon > 0$ vérifiant la condition (2.10) pour $0 < \delta \leq \delta_\epsilon$, et $\forall \tau > 0$,

$$\begin{aligned} & \mathbb{P}\left\{\sup_{x \in I} \frac{|W_{n,k}(x, \psi)|}{\sqrt{2nh_n \log(1/h_n)}} > \{1 + \tau + A\sqrt{\epsilon}\}\sigma_W(I)\right\} \\ & \leq \mathbb{P}\left\{\max_{0 \leq i \leq l_n} \frac{|\alpha_n(g_{n,i})|}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I)(1 + \tau)\right\} \\ & + \mathbb{P}\left\{\max_{0 \leq i \leq l_n} \frac{\|\alpha_n\|g'_{n,i}}{\sqrt{2h_n \log(1/h_n)}} > \sigma_W(I)A\sqrt{\epsilon}\right\} = o(1). \end{aligned}$$

Pour finir, ϵ et τ étant arbitrairement petits, nous les choisissons tels que $\epsilon \geq \tau + A\epsilon$, ce qui nous donne clairement la borne supérieure (2.6) énoncé en début de section.

2.4.2 Borne inférieure

Le but de cette sous-section est de prouver que,

$$\liminf_{n \rightarrow \infty} \sup_{x \in I} \frac{\pm W_{n,k}(x, \psi)}{\sqrt{2nh_n \log(1/h_n)}} \stackrel{\mathbf{P}}{\geq} \sigma_W(I). \quad (2.26)$$

Résultats nécessaires pour le traitement de la borne inférieure

Soit $Z = Z_1, Z_2, \dots$, une suite de vecteurs aléatoires i.i.d. à valeurs dans \mathbb{R}^2 . Pour chaque $n \geq 1$, on considère la fonction de répartition empirique, basée sur les n premiers vecteurs aléatoires, définie par

$$G_n(s) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{Z_i \leq s\}, \quad s \in \mathbb{R}^2,$$

où $z \leq s$ signifie que chaque composante de z est inférieure ou égale à la composante de s correspondante. Ensuite, pour n'importe quelle fonction mesurable $g(\cdot)$ à valeurs réelles et définie sur \mathbb{R}^2 ($g : \mathbb{R}^2 \rightarrow \mathbb{R}$), on pose

$$G_n(g) = \int_{\mathbb{R}^2} g(s) dG_n(s), \quad \mu(g) = \mathbb{E}[g(Z)] \quad \text{et} \quad \bar{\sigma}(g) = \sqrt{\text{Var}[g(Z)]}.$$

Soit $\{a_n : n \geq 1\}$ une suite de constantes positives convergeant vers zéro. On considère une suite $\mathcal{G}_n = \{g_i^{(n)} : i = 1, \dots, k_n\}$ d'ensembles de fonctions mesurables à valeurs réelles sur \mathbb{R}^2 , i.e. pour chaque $n \geq 1$ la classe \mathcal{G}_n contient k_n fonctions de la forme $g_i^{(n)}(\cdot)$. Ici, le (n) exprime la dépendance en n et non une puissance ou un degré de dérivation. Pour chaque fonction $g_i^{(n)} \in \mathcal{G}_n$, les conditions suivantes sont vérifiées,

$$\mathbb{P}\left\{g_i^{(n)}(Z) \neq 0, g_j^{(n)}(Z) \neq 0\right\} = 0, \quad \forall 1 \leq i \neq j \leq k_n; \quad \sum_{i=1}^{k_n} \mathbb{P}\left\{g_i^{(n)}(Z) \neq 0\right\} \leq 1/2. \quad (2.27)$$

De plus, on suppose les hypothèses suivantes,

(R.1) pour un certain $0 < r < \infty$, la suite $a_n k_n \rightarrow r$ lorsque $n \rightarrow \infty$;

(R.2) pour certains $-\infty < \mu_1, \mu_2 < \infty$, uniformément en $i = 1, \dots, k_n$, et pour n suffisamment grand,

$$a_n \mu_1 \leq \mu(g_i^{(n)}) \leq a_n \mu_2; \quad (2.28)$$

(R.3) pour certains $0 < \sigma_1 < \sigma_2 < \infty$, uniformément en $i = 1, \dots, k_n$, et pour n suffisamment grand,

$$\sqrt{a_n} \sigma_1 \leq \bar{\sigma}(g_i^{(n)}) \leq \sqrt{a_n} \sigma_2; \quad (2.29)$$

(R.4) pour un certain $0 < M < \infty$, uniformément en $i = 1, \dots, k_n$, et pour n suffisamment grand,

$$|g_i^{(n)}| \leq M.$$

Lemme 2.4.6 *Sous les conditions (R.1–4), pour chaque $0 < \epsilon < 1$*

$$\mathbb{P}\left\{\max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)}) \sqrt{2 \log(1/a_n)}} \geq 1 - \epsilon\right\} \rightarrow 1. \quad (2.30)$$

Remarque 2.4.5 Ce lemme est la clef de notre démonstration pour la borne inférieure. Il repose sur une approximation poissonnienne du processus empirique (voir proposition 2.2 dans Einmahl et Mason (2000)).

Soit $n\Pi_n$ un processus de Poisson sur \mathbb{R}^2 tel que, pour tout borélien A de \mathbb{R}^2 ,

$$n\mathbb{E}[\Pi_n(A)] = n\mathbb{P}\{Z \in A\}.$$

Lemme 2.4.7 *Soit $\{g_i^{(n)} : 1 \leq i \leq k_n\}$, un ensemble de fonctions mesurables telles que les conditions (2.27) soient vérifiées. Alors, pour tout boréliens B_1, \dots, B_{k_n} de \mathbb{R} , on a*

$$\mathbb{P}\left\{G_n(g_i^{(n)}) \in B_i, i = 1, \dots, k_n\right\} \leq 2 \prod_{i=1}^{k_n} \mathbb{P}\left\{\Pi_n(g_i^{(n)}) \in B_i\right\}.$$

La démonstration est similaire à celle du lemme 2.1, p. 1253-1254, [26]. \square

Pour chaque $i = 1, \dots, k_n$, on dénote par $A_i^{(n)}$ l'événement :

$$\frac{n^{1/2}\{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)})\sqrt{2\log(1/a_n)}} < 1 - \epsilon.$$

D'après le lemme 2.4.7, il s'ensuit

$$\mathbb{P}\left\{\bigcap_{i=1}^{k_n} A_i^{(n)}\right\} \leq 2 \prod_{i=1}^{k_n} \mathbb{P}\{B_i^n\},$$

lorsque B_i^n dénote l'événement :

$$\frac{n^{1/2}\{\Pi_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)})\sqrt{2\log(1/a_n)}} < 1 - \epsilon.$$

On rappelle que

$$n\Pi_n(s) \stackrel{\mathcal{L}}{=} \sum_{i \leq \pi_n} \mathbb{I}\{Z_i \leq s\},$$

où π_n désigne une variable aléatoire $\mathcal{P}(n)$ (i.e., une v.a. de Poisson de moyenne n) indépendante des $\{Z_i : i \geq 1\}$. A présent, choisissons un $\delta > 0$ tel que $\delta\mu_2/\sigma_1 < (\epsilon/2)^2$, en vue de (2.28) et (2.29). Nous avons clairement,

$$\begin{aligned} \mathbb{P}\left\{(B_i^n)^c\right\} &= \sum_{m \geq 1} \mathbb{P}\left\{(B_i^n)^c \mid \pi_n = m\right\} \mathbb{P}\{\pi_n = m\} \\ &\geq \sum_{|n-m| \leq \delta\sqrt{n}} \mathbb{P}\left\{(B_i^n)^c \mid \pi_n = m\right\} \mathbb{P}\{\pi_n = m\}. \end{aligned}$$

Cette dernière quantité est supérieure ou égale à, d'après l'égalité en distribution ci-dessus,

$$\begin{aligned} &\sum_{|n-m| \leq \delta\sqrt{n}} \mathbb{P}\left\{\frac{n^{1/2}\{(m/n)G_m(g_i^{(n)}) - (m/n)\mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)})\sqrt{2\log(1/a_n)}} \geq 1 - \epsilon + \frac{\delta\mu_2}{\sigma_1} \mid \pi_n = m\right\} \\ &\quad \times \mathbb{P}\{\pi_n = m\} \\ &\geq \sum_{|n-m| \leq \delta\sqrt{n}} \mathbb{P}\left\{\frac{n^{1/2}\{(m/n)G_m(g_i^{(n)}) - (m/n)\mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)})\sqrt{2\log(1/a_n)}} \geq (1 - \epsilon)^2 \mid \pi_n = m\right\} \\ &\quad \times \mathbb{P}\{\pi_n = m\} \end{aligned}$$

Nous obtenons (2.30) ou sa version presque sûre en suivant les arguments p. 26-27 [42].

La démonstration de la borne inférieure (2.26) est équivalente à la preuve de la proposition ci-dessous.

Proposition 2.4.3 *Sous les hypothèses du théorème 2.3.1, pour tout $0 < \varepsilon < 1/2$, nous avons,*

$$\mathbb{P}\left\{\sup_{x \in I} \frac{\pm W_{n,k}(x, \psi)}{\sqrt{2nh_n \log(1/h_n)}} < (1 - \varepsilon)\sigma(\psi)\right\} = o(1). \quad (2.31)$$

Cette proposition est une conséquence du lemme 2.4.6. Il faut donc vérifier les conditions du lemme 2.4.6. La première étape consiste à choisir un sous-intervalle $I_1 = [a_1, b_1]$ de $I = [a, b]$ tel que, pour $\varepsilon > 0$,

$$\inf_{x \in I_1} \mathbb{E}\left[(c(x)\psi(Y) + d(x))^2 | X = x\right] f_X(x) \int_{\mathbb{R}} \left[K^{(k)}(t)\right]^2 dt > \sigma_W^2(I)(1 - \varepsilon/2), \quad (2.32)$$

où $\sigma_W^2(I)$ est définie en (2.4) et

$$\mathbb{P}\{X \in I_1\} \leq 1/2. \quad (2.33)$$

Ceci est possible d'après (F.1-2) qui impliquent la continuité sur I de la fonction

$$x \rightarrow \mathbb{E}\left[(c(x) + \psi(Y)d(x))^2 | X = x\right] f_X(x) \int_{\mathbb{R}} \left[K^{(k)}(t)\right]^2 dt.$$

Afin de satisfaire les conditions du précédent Lemme, on discrétise l'intervalle I_1 en k_n points :

$$x_{i,n} = a_1 + 2ih_n, \quad \text{pour } i = 1, \dots, \lfloor (b_1 - a_1)/2h_n \rfloor - 1 := k_n.$$

D'après cette définition de k_n , l'hypothèse (R.1) est bien vérifiée avec $a_n = h_n$, i.e. $\lim_{n \rightarrow \infty} h_n k_n \approx \lfloor (b_1 - a_1)/2 \rfloor$.

Pour chaque $x_{i,n}$, $1 \leq i \leq k_n$, on associe la fonction

$$g_i^{(n)}(x, y) := (c(x_{i,n})\psi(y) + d(x_{i,n})) K^{(k)}\left(\frac{x_{i,n} - x}{h_n}\right).$$

Ainsi, la condition (R.4) est bien vérifiée, uniformément en $1 \leq i \leq k_n$ nous avons

$$\|g_i^{(n)}\| \leq \{\|c\|_J \|\psi\| + \|d\|_J\} \|K^{(k)}\|.$$

Maintenant, rappelons que le noyau $K(\cdot)$ vérifie

$$\begin{aligned} K(u) &= 0 \text{ pour } u \notin [-1/2, 1/2] \text{ donc} \\ K^{(k)}(u) &= 0 \text{ pour } u \notin [-1/2, 1/2] \text{ et alors} \\ g_i^{(n)}(X, Y) &\neq 0 \iff |x_{i,n} - X| \leq h_n/2 \\ |x_{j,n} - X| &= |x_{j,n} - x_{i,n} + x_{i,n} - X| \geq 2h_n - h_n/2 \text{ pour } i \neq j. \end{aligned}$$

En conséquence, pour $1 \leq i \neq j \leq k_n$

$$\mathbb{P}\left\{g_i^{(n)}(X, Y) \neq 0 \text{ et } g_j^{(n)}(X, Y) \neq 0\right\} = 0.$$

Par la suite, on remarque que

$$\text{Var}(g_i^{(n)}(X, Y)) \leq \mathbb{E}[(g_i^{(n)}(X, Y))^2]$$

$$= \int_{\mathbf{R}} \mathbb{E} \left[(c(x_{i,n})\psi(Y) + d(x_{i,n}))^2 | X = x \right] \left[K^{(k)} \left(\frac{x_{i,n} - x}{h_n} \right) \right]^2 f_X(x) dx$$

Une application du lemme de Bochner (résultat A.2.1) nous indique que

$$\begin{aligned} & \int_{\mathbf{R}} \mathbb{E} \left[(c(x_{i,n})\psi(Y) + d(x_{i,n}))^2 | X = x \right] \left[K^{(k)} \left(\frac{x_{i,n} - x}{h_n} \right) \right]^2 f_X(x) dx \rightarrow \\ & h_n \mathbb{E} \left[(c(x_{i,n})\psi(Y) + d(x_{i,n}))^2 | X = x_{i,n} \right] f_X(x_{i,n}) \int_{\mathbf{R}} \left[K^{(k)}(u) \right]^2 du \\ & \leq h_n \sigma_W^2(I). \end{aligned}$$

Nous obtenons alors la borne supérieure de (R.3); pour tout $\epsilon > 0$, si n suffisamment grand, uniformément en $1 \leq i \leq k_n$,

$$\text{Var}(g_i^{(n)}(X, Y)) \leq h_n \sigma_W^2(I)(1 + \epsilon).$$

En outre, une application du lemme de Bochner combiné avec (2.32) nous donne, pour n suffisamment grand, uniformément en $1 \leq i \leq k_n$,

$$\text{Var}(g_i^{(n)}(X, Y)) \geq h_n \sigma_W^2(I)(1 - \epsilon).$$

Au final, $\sigma_W(I)$ étant positif,

$$\sqrt{h_n} \sigma_W(I) \sqrt{(1 - \epsilon)} \leq \left[\text{Var}(g_i^{(n)}(X, Y)) \right]^{1/2} =: \bar{\sigma}(g_i^{(n)}) \leq \sqrt{h_n} \sigma_W(I) \sqrt{(1 + \epsilon)},$$

toujours uniformément en $1 \leq i \leq k_n$ et pour n assez grand.

Nous procédons identiquement pour montrer l'équivalent de (R.2) avec $a_n = h_n$. On note que

$$\bar{\sigma}(g_i^{(n)})(1 - \epsilon) \geq (1 - \epsilon) \sqrt{(1 - \epsilon)} \sqrt{h_n} \sigma_W(I) = (1 - \epsilon)^{3/2} \sigma_W(I) \sqrt{h_n}.$$

En conséquence,

$$\begin{aligned} & \mathbb{P} \left\{ \max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)}) \sqrt{2 \log(1/h_n)}} \geq 1 - \epsilon \right\} \\ & \leq \mathbb{P} \left\{ \max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\sqrt{2 h_n \log(1/h_n)}} \geq (1 - \epsilon)^{3/2} \sigma_W(I) \right\} \end{aligned}$$

En appliquant le lemme 2.4.6 avec $a_n = h_n$, il s'ensuit

$$\mathbb{P} \left\{ \max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\sqrt{2 h_n \log(1/h_n)}} \geq (1 - \epsilon)^{3/2} \sigma_W(I) \right\} \rightarrow 1,$$

ou

$$\mathbb{P} \left\{ \max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\sqrt{2 h_n \log(1/h_n)}} < (1 - \epsilon) \sigma_W(I) \right\} = o(1),$$

avec $(1 - \epsilon)^{3/2} = 1 - \epsilon$. Finalement, l'inégalité

$$\sup_{x \in I} \frac{W_{n,k}(x, \psi)}{\sqrt{2nh_n \log(1/h_n)}} \geq \max_{1 \leq i \leq k_n} \frac{n^{1/2} \{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\sqrt{2h_n \log(1/h_n)}},$$

entraîne (2.31) et la validité de la proposition 2.4.3.

Le cas "–" est similaire et ne sera pas présenté ici par soucis de concision. \square

2.4.3 Démonstration du théorème 2.3.1

Premièrement, nous remarquons que

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{x \in I} \frac{\pm W_{n,k}(x, \psi)}{\sqrt{2nh_n \log(1/h_n)}} > (1 + \epsilon) \sigma_W(I) \right\} \\ & \leq \mathbb{P} \left\{ \sup_{x \in I} \frac{|W_{n,k}(x, \psi)|}{\sqrt{2nh_n \log(1/h_n)}} > (1 + \epsilon) \sigma_W(I) \right\} = o(1), \end{aligned}$$

en se référant à (2.6). Ceci, combiné à (2.26), entraîne

$$\left| \sup_{x \in I} \frac{\pm W_{n,k}(x, \psi)}{\sqrt{2nh_n \log(1/h_n)}} - \sigma_W(I) \right| = o_{\mathbf{P}}(1),$$

ce qui complète la démonstration du théorème 2.3.1.

2.4.4 Démonstration des corollaires 2.3.1 et 2.3.2

En appliquant le théorème 2.3.1 pour le choix particulier de fonctions $c(x) = 0$ et $d(x) = 1$ puis, pour le choix $c(x) = 1$ et $d(x) = 0$ nous obtenons les corollaires 2.3.1 et 2.3.2, respectivement. Plus précisément, lorsque $c(x) = 0$ et $d(x) = 1$,

$$W_{n,k}(x, \psi) = nh_n^{1+k} \{ \hat{f}_{X;n}^{(k)}(x) - f_{X;n}^{(k)}(x) \}.$$

De même, lorsque $c(x) = 1$ et $d(x) = 0$,

$$W_{n,k}(x, \psi) = nh_n^{1+k} \{ \hat{r}_{\psi;n}^{(k)}(x) - r_{\psi;n}^{(k)}(x) \}.$$

2.4.5 Démonstration du théorème 2.3.2

La démonstration du théorème 2.3.2 est une conséquence du lemme 2.4.8 ci-dessous. En posant $c(x) = 1/f_X(x)$ et $d(x) = -m_\psi(x)/f_X(x)$ dans la définition (2.3.1) de $W_{n,k}(x)$, nous obtenons

$$W_{n,k}(x) = nh_n^{k+1} \left\{ \frac{\hat{r}_{\psi;n}^{(k)}(x)}{f_X(x)} - \frac{r_{\psi;n}^{(k)}(x)}{f_X(x)} - \frac{r_{\psi}(x)}{f_X(x)} \times \frac{\{ \hat{f}_{X;n}^{(k)}(x) - f_{X;n}^{(k)}(x) \}}{f_X(x)} \right\}.$$

Remarque 2.4.6 Dans les précédentes démonstrations nous avons supposé que les fonctions $c(\cdot)$ et $d(\cdot)$ étaient continues et bornées sur l'intervalle I . Sous les hypothèses (F.1–3) cette condition est bien vérifiée lorsque $c(x) = 1/f_X(x)$ et $d(x) = -m_\psi(x)/f_X(x)$, d'après la remarque 2.2.1 combinée à la compacité de l'intervalle I .

Lemme 2.4.8 *Sous les hypothèses du théorème 2.3.1, nous obtenons l'approximation suivante, lorsque $n \rightarrow \infty$,*

$$\epsilon_n := \{\theta_n\}^{-1/2} \sup_{x \in I} \left| W_{n,k}(x) - nh_n^{k+1} \left\{ \hat{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)] \right\} \right| = o_{\mathbf{P}}(1),$$

où

$$\theta_n = nh_n \log(1/h_n).$$

On commence par $k = 0$. Nous remarquons que,

$$\begin{aligned} \hat{r}_{\psi;n}(x) - r_{\psi;n}(x) - m(x) \{ \hat{f}_{X;n}(x) - f_{X;n}(x) \} = \\ \{ \hat{f}_{X;n}(x) - f_{X;n}(x) \} \{ \hat{m}_{\psi;n}(x) - m(x) \} + f_{X;n}(x) \{ \hat{m}_{\psi;n}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}(x)] \}. \end{aligned}$$

Il s'ensuit

$$\begin{aligned} \hat{m}_{\psi;n}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}(x)] &= \frac{1}{f_{X;n}(x)} \left\{ \hat{r}_{\psi;n}(x) - r_{\psi;n}(x) - m(x) \{ \hat{f}_{X;n}(x) - f_{X;n}(x) \} \right\} \\ &\quad - \{ \hat{m}_{\psi;n}(x) - m_{\psi;n}(x) \} \times \{ \hat{f}_{X;n}(x) - f_{X;n}(x) \} \times \{ f_{X;n}(x) \}^{-1}. \end{aligned}$$

Comme, sous nos hypothèses, l'estimateurs $\hat{m}_{\psi;n}(x)$ est consistant, uniformément sur I , la deuxième partie du membre de droite de l'égalité ci-dessus est négligeable, via le corollaire 2.3.1.

On traite à présent le cas où $k = 1$. La généralisation découlera d'un simple argument de récurrence.

$$\begin{aligned} \epsilon_n &= \left\{ \frac{nh_n^3}{\log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \left| \frac{\hat{r}'_{\psi;n}(x)}{f_X(x)} - \frac{r'_{\psi;n}(x)}{f_X(x)} - \frac{r_\psi(x)}{f_X(x)} \times \frac{\{ \hat{f}'_{X;n}(x) - f'_{X;n}(x) \}}{f_X(x)} \right. \\ &\quad \left. - \frac{\hat{r}'_{\psi;n}(x)}{\hat{f}_{X;n}(x)} + \frac{\hat{r}_{\psi;n}(x) \hat{f}'_{X;n}(x)}{\hat{f}_{X;n}^2(x)} + \frac{r'_{\psi;n}(x)}{f_{X;n}(x)} - \frac{r_{\psi;n}(x) \hat{f}'_{X;n}(x)}{f_{X;n}^2(x)} \right| \\ &= \left\{ \frac{nh_n^3}{\log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \left| \frac{1}{f_{X;n}(x) f_X(x)} \left\{ \hat{r}'_{\psi;n}(x) - r'_{\psi;n}(x) \right\} \left\{ f_{X;n}(x) - f_X(x) \right\} \right. \\ &\quad \left. - \frac{\hat{r}'_{\psi;n}(x)}{f_{X;n}(x) \hat{f}_{X;n}(x)} \left\{ f_{X;n}(x) - \hat{f}_{X;n}(x) \right\} + \left\{ f_{X;n}(x) f_X(x) \right\}^{-2} \times \right. \end{aligned}$$

$$\left\{ \hat{f}'_{X;n}(x) - f'_{X;n}(x) \right\} \left\{ r_{\psi;n}(x) f_X^2(x) - r_{\psi}(x) f_{X;n}^2(x) \right\} - \frac{\hat{f}'_{X;n}(x)}{\{f_{X;n}(x) \hat{f}_{X;n}(x)\}^2} \times$$

$$\left| \left\{ \hat{f}_{X;n}^2(x) r_{\psi;n}(x) - \hat{r}_{\psi;n}(x) f_{X;n}^2(x) \right\} \right|.$$

Maintenant, en appliquant le théorème 2.3.1 avec $c(x) = 0$ et $d(x) = 1$ puis $c(x) = 1$ et $d(x) = 0$ pour $k = 0, 1$, nous obtenons

$$\left\{ \frac{nh_n^3}{\log(1/h_n)} \right\}^{1/2} \sup_{x \in I} |\hat{f}_{X;n}(x) - f_{X;n}(x)| \stackrel{\mathbf{P}}{=} O(h_n) = o(1),$$

$$\left\{ \frac{nh_n^3}{\log(1/h_n)} \right\}^{1/2} \sup_{x \in I} |\hat{r}_{\psi;n}(x) - r_{\psi;n}(x)| \stackrel{\mathbf{P}}{=} O(h_n) = o(1),$$

$$\left\{ \frac{nh_n^3}{\log(1/h_n)} \right\}^{1/2} \sup_{x \in I} |\hat{f}'_{X;n}(x) - f'_{X;n}(x)| \stackrel{\mathbf{P}}{=} O(1),$$

$$\left\{ \frac{nh_n^3}{\log(1/h_n)} \right\}^{1/2} \sup_{x \in I} |\hat{r}'_{\psi;n}(x) - r'_{\psi;n}(x)| \stackrel{\mathbf{P}}{=} O(1).$$

De plus, une application directe du lemme de Bochner implique les égalités suivantes,

$$\sup_{x \in I} |f_{X;n}(x) - f_X(x)| = o(1) \quad \text{et} \quad \sup_{x \in I} |r_{\psi;n}(x) - r_{\psi}(x)| = o(1).$$

En utilisant les dernières égalités ci-dessus, il s'ensuit aisément

$$\epsilon_n = o_{\mathbf{P}}(1),$$

ce qui clôt la démonstration. Notons qu'il faut décomposer, via des linéarisations successives, chacune des expressions de ϵ_n de manière à faire apparaître les déviations ci-dessus. \square

En combinant le lemme 2.4.8 avec le théorème 2.3.1, nous obtenons directement le théorème 2.3.2.

Remarque 2.4.7 Nous pouvons remarquer que dans l'étude asymptotique des estimateurs des dérivées de la régression $\hat{m}_{\psi;n}^{(k)}(x)$, les termes $\hat{r}_{\psi;n}^{(k)}(x)$ et $\hat{f}_{X;n}^{(k)}(x)$ sont prépondérants, ce sont eux qui déterminent la vitesse exacte de convergence.

2.4.6 Le cas non-borné

Dans cette section, nous ne supposons plus les variables $\{Y_i : 1 \leq i \leq n\}$ bornées (cf. la condition (F.3)). Nous travaillons désormais sous l'hypothèse (F.4), que nous rappelons ici, par convenance,

$$(F.4) \sup_{x \in J} \mathbb{E} \left[|\psi(Y)|^s | X = x \right] < \infty, \text{ pour un certain } s > 2.$$

Cette condition de moment est nécessaire pour traiter le cas non-borné, elle nous servira notamment à traiter la partie aléatoire du reste, via l'inégalité de Markov.

Par la suite, en outre des hypothèses classiques sur la fenêtre h_n , nous nécessitons une hypothèse sur la fenêtre plus forte que (H.2), liée à l'hypothèse (F.4),

$$(H.2)^* \quad n^{1-2/s} h_n \log n \rightarrow \infty \Leftrightarrow n^{1-2/s} h_n \log(1/h_n) \rightarrow \infty.$$

A présent, nous pouvons énoncer, ci-dessous, le théorème 2.3.1 principal sous sa forme la plus générale.

Théorème 2.4.2 *Supposons les hypothèses (F.1–2–4), (H.1), (H.2)* et (K.1–4) vérifiées. Alors nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \{2nh_n \log(1/h_n)\}^{-1/2} \sup_{x \in J} \{ \pm W_{n,k}(x, \psi) \} - \sigma_W(I) \right| \stackrel{\mathbf{P}}{=} o(1).$$

La démonstration du théorème est classique, elle procède d'un argument de troncation essentiellement. L'idée est de tronquer la partie de $W_{n,k}(x, \psi)$ qui dépend des variables Y_i et de montrer que le reste est négligeable pour notre vitesse de convergence (en s'appuyant sur l'hypothèse de moment (F.4) et la condition sur la fenêtre (H.2)* ci-dessus).

Pour cela, nous introduisons donc un nouveau processus où nous pouvons étudier de plus près le comportement du processus général $W_{n,k}(x, \psi)$, lorsque la variable Y n'est pas bornée,

$$V_{n,k}(x, \psi) := c(x) \sum_{j=1}^n \psi(Y_j) K^{(k)}\left(\frac{x - X_j}{h_n}\right) - c(x) n \mathbb{E} \left[\psi(Y) K^{(k)}\left(\frac{x - X_j}{h_n}\right) \right].$$

D'après (F.4), nous remarquons que $\sup_{x \in J} \mathbb{E} \left[|\psi(Y)|^2 | X = x \right] < \infty$.

Proposition 2.4.4 *Sous les hypothèses du précédent théorème, il existe une constante absolue $B > 0$ telle que, nous avons,*

$$\mathbb{P} \left\{ \sup_{x \in I} \frac{|V_{n,k}(x, \psi)|}{\sqrt{2nh_n \log(1/h_n)}} > B \sqrt{\beta_2(\psi)} \right\} = o(1),$$

où

$$\beta_2(\psi) := \sup_{x \in J} \mathbb{E} \left[\{\psi(Y)\}^2 | X = x \right].$$

-Démonstration de la proposition 2.4.4

La démonstration de la proposition 2.4.4 sera une conséquence de deux lemmes, présentés ci-dessous. Dans un premier temps, nous introduisons quelques notations et définitions.

Nous définissons $V_{n,k}(x, \psi_n)$, la partie tronquée de $V_{n,k}(x, \psi)$, telle que

$$\psi_n(y) := \psi(y) \mathbb{I} \{ |\psi(y)| < n^{1/s} \}.$$

Pour $x \in J$, nous posons

$$v_{n,x,k}(u, v) = v_{n,x}(u, v) := c(x) \psi_n(v) K^{(k)}\left(\frac{x - u}{h_n}\right), \quad \text{pour } (u, v) \in \mathbb{R}^2. \quad (2.34)$$

Ainsi, nous pouvons écrire

$$V_{n,k}(x, \psi_n) = n^{1/2} \alpha_n(v_{n,x}).$$

Pour $n \geq 1$, soit la classe de fonctions

$$\mathcal{H}_n := \left\{ v_{n,x} : x \in I \right\}.$$

Nous avons clairement,

$$\frac{\|n^{1/2} \alpha_n\|_{\mathcal{H}_n}}{\sqrt{2nh_n \log(1/h_n)}} = \sup_{x \in I} \frac{|V_{n,k}(x, \psi_n)|}{\sqrt{2nh_n \log(1/h_n)}}.$$

La proposition 2.4.4 est liée au lemme suivant.

Lemme 2.4.9 *Sous les hypothèses de la proposition ci-dessus, nous avons, pour $C > 0$ constante convenablement choisie,*

$$\mathbb{P} \left\{ \|\alpha_n\|_{\mathcal{H}_n} > C \sqrt{\beta_2(\psi) h_n \log(1/h_n)} \right\} = o(1).$$

La démonstration du lemme 2.4.9 est comparable à celle de la borne supérieure (cf. la sous-section 2.4.1) mais en plus simple. Ceci est dû au fait que nous ne cherchons pas à déterminer une borne exacte mais juste la vitesse de convergence uniforme. Plus précisément, l'argumentation sera similaire à la partie oscillation de la sous-section 2.4.1. Dans un premier temps il faut déterminer une borne pour le composant tronqué $v_{n,x}$. D'après la définition en (2.34), il s'ensuit

$$\|v_{n,x}\| \leq \|c\| \|K^{(k)}\| \times n^{1/s} =: M n^{1/s}.$$

En suivant le schéma classique développée dans la précédente démonstration de la borne supérieure, pour pouvoir appliquer la fameuse inégalité exponentielle de Talagrand, il nous reste à borner un terme de variance :

$$\text{Var}[v_{n,x}(X, Y)] \leq \mathbb{E}[v_{n,x}^2(X, Y)] \leq \mathbb{E} \left[(c(x)\psi(Y))^2 \left\{ K^{(k)} \left(\frac{x-X}{h_n} \right) \right\}^2 \right].$$

En utilisant un argument de conditionnement combiné avec (K.2) et (F.2), nous obtenons, lorsque n est assez grand,

$$\begin{aligned} & \mathbb{E} \left[(c(x)\psi(Y))^2 \left\{ K^{(k)} \left(\frac{x-X}{h_n} \right) \right\}^2 \right] \\ & \leq \|c\|^2 \int_{|x-t| \leq h_n/2} \mathbb{E}[\psi^2(Y) | X=t] f_X(t) \left[K^{(k)} \left(\frac{x-t}{h_n} \right) \right]^2 dt \\ & \leq \|c\|^2 \beta_2(\psi) \int_{|x-t| \leq h_n/2} f_X(t) \left[K^{(k)} \left(\frac{x-t}{h_n} \right) \right]^2 dt \\ & \leq h_n \|c\|^2 \beta_2(\psi) \int_{-1}^1 f_X(x - h_n u) [K^{(k)}(u)]^2 du \end{aligned}$$

$$\leq h_n \|c\|^2 \beta_2(\psi) \|f_X\|_J \|K^{(k)}\|_2^2 \quad \text{cf. [lemme de Bochner].}$$

Cette inégalité est vraie uniformément en $x \in I$. Pour n suffisamment grand, il s'ensuit,

$$\sigma_{\mathcal{H}_n}^2 := \sup_{\mathcal{H}_n} \text{Var} \left[v_{n,x}(X, Y) \right] \leq h_n \|c\|^2 \beta_2(\psi) \|f_X\|_J \|K^{(k)}\|_2^2. \quad (2.35)$$

A présent, remarquons que \mathcal{H}_n satisfait la condition d'entropie $[\mathcal{E}]$, c'est à dire

$$N(\epsilon, \mathcal{H}_n) \leq C \epsilon^{-\nu}, \quad 0 < \epsilon < 1.$$

Ainsi d'après le résultat A.3.1, $\forall t > 0$ et pour certaines constantes A_1, A_2 convenablement choisies, nous obtenons

$$\begin{aligned} & \mathbb{P} \left\{ \|n^{1/2} \alpha_n\|_{\mathcal{H}_n} \geq A_1 \left(\mathbb{E} \left\| \sum_{i=1}^n \epsilon_i g(X_i, Y_i) \right\|_{\mathcal{H}_n} + t \right) \right\} \\ & \leq 2 \left\{ \exp \left(\frac{-A_2 t^2}{n \sigma_{\mathcal{H}_n}^2} \right) + \exp \left(\frac{-A_2 t}{M n^{1/s}} \right) \right\}. \end{aligned} \quad (2.36)$$

Ensuite, en appliquant le résultat A.3.2 pour $\mathcal{G} = \mathcal{H}_n$, il s'ensuit la borne suivante :

$$\mathbb{E} \left\| \sum_{i=1}^n \epsilon_i g(X_i, Y_i) \right\|_{\mathcal{H}_n} \leq A_3 \sqrt{\beta_2(\psi) n h_n \log(1/h_n)}, \quad (2.37)$$

où A_3 désigne une constante strictement positive. Ainsi, d'après (2.36) combinée à (2.35) et (2.37), nous avons, pour $t = A_3 \sqrt{\beta_2(\psi) n h_n \log(1/h_n)}$,

$$\begin{aligned} & \mathbb{P} \left\{ \|n^{1/2} \alpha_n\|_{\mathcal{H}_n} \geq 2A_1 A_3 \sqrt{\beta_2(\psi) n h_n \log(1/h_n)} \right\} \leq \\ & 2 \left\{ \exp \left(\frac{-A_2 A_3^2 \log(1/h_n)}{\|c\|^2 \|f_X\|_J \|K^{(k)}\|_2^2} \right) + \exp \left(\frac{-A_2 A_3 \sqrt{\beta_2(\psi) n h_n \log(1/h_n)}}{M n^{1/s}} \right) \right\} = o(1), \end{aligned}$$

où nous justifions l'usage de $(H.2)^* n^{1-2/s} h_n \log(1/h_n) \rightarrow \infty$ afin d'obtenir la dernière égalité en $o(1)$.

En résumé, pour le choix de $C = 2A_1 A_3$, nous validons la démonstration du premier lemme. \square

A présent, nous étudions le reste de la troncature. Notre but est de démontrer que le reste est négligeable asymptotiquement. Pour cela, posons

$$\mu_n(x) := c(x) n \mathbb{E} \left[\bar{\psi}_n(Y) K^{(k)} \left(\frac{x - X}{\lambda h_n} \right) \right], \quad (2.38)$$

où

$$\bar{\psi}_n(y) := \psi(y) - \psi_n(y) = \psi(y) \mathbb{I} \{ |\psi(y)| \geq n^{1/s} \}.$$

-La partie stochastique

Lemme 2.4.10 *Sous les hypothèses de la proposition 2.4.4, nous obtenons*

$$\sup_{x \in I} \frac{|\mu_n(x)|}{\sqrt{nh_n \log(1/h_n)}} \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty.$$

Soit

$$\beta_s(\psi) := \sup_{x \in J} \mathbb{E} \left[|\psi(Y)|^s \mid X = x \right].$$

D'après (2.38), il s'ensuit

$$\begin{aligned} |\mu_n(x)| &\leq \|c\| n \mathbb{E} \left[\psi(Y) \mathbb{I} \{ \psi(Y) > n^{1/s} \} K^{(k)} \left(\frac{x - X}{h_n} \right) \right] \\ &\leq n \|c\| n^{-(s-1)/s} \beta_s(\psi) \int_{-1}^1 f_X(x - h_n u) |K^{(k)}(u)| du \\ &\leq h_n n^{1/s} \|c\|_J \beta_s(\psi) \|f_X\|_J \|K^{(k)}\| =: h_n n^{1/s} C_s. \end{aligned}$$

Pour n suffisamment grand, il s'ensuit

$$\sup_{x \in I} \frac{|\mu_n(x)|}{\sqrt{nh_n \log(1/h_n)}} \leq \sqrt{\frac{h_n^2 n^{2/s}}{nh_n \log(1/h_n)}} \times C_s = \sqrt{\frac{h_n n^{2/s-1}}{\log(1/h_n)}} \times C_s,$$

ce dernier terme convergeant vers 0 d'après les hypothèses (H.1) et (H.2)*. \square

-La partie aléatoire

Nous pouvons à présent terminer la démonstration de la proposition 2.4.4. La fin de la démonstration repose sur l'hypothèse (F.4) $\beta_s(\psi) < \infty$, combinée à l'inégalité de Markov. On remarque que d'après (F.4) il existe un $r \in \mathbb{R}$ tel que $2 < r < s$ et

$$\sup_{x \in J} \mathbb{E} \left[|\psi(Y)|^r \mid X = x \right] < \infty. \quad (2.39)$$

Il s'ensuit également, via (2.39),

$$\mathbb{E} \left[|\psi(Y)|^r \mathbb{I} \{ X \in J \} \right] < \infty.$$

D'après l'inégalité de Markov à l'ordre 1, nous obtenons donc,

$$\mathbb{P} \left\{ \max_{1 \leq i \leq n} |\psi(Y_i)| \mathbb{I} \{ X_i \in J \} \geq n^{1/s} \right\} = n \mathbb{P} \left\{ |\psi(Y)|^r \mathbb{I} \{ X \in J \} \geq n^{r/s} \right\} \leq O(n^{1-r/s}) = o(1).$$

On peut résumer l'inégalité ci-dessus par

$$\left\{ \max_{1 \leq i \leq n} |\psi(Y_i)| \mathbb{I} \{ X_i \in J \} \geq n^{1/s} \right\} \xrightarrow{\mathbf{P}} 0. \quad (2.40)$$

Par la suite, en décomposant suivant la troncation,

$$V_{n,k}(x, \psi) = c(x) \sum_{j=1}^n \psi(Y_j) K^{(k)} \left(\frac{x - X_j}{h_n} \right) - c(x) n \mathbb{E} \left[\psi(Y) K^{(k)} \left(\frac{x - X}{h_n} \right) \right]$$

$$= V_{n,k}(x, \psi_n) + c(x) \sum_{j=1}^n \psi(Y_j) \mathbf{I}\{|\psi(Y_j)| \geq n^{1/s}\} K^{(k)}\left(\frac{x - X_j}{h_n}\right) - \mu_n(x),$$

où le deuxième terme est asymptotiquement nul en probabilité, d'après (2.40). Ainsi, avec probabilité convergeant vers 1 lorsque $n \rightarrow \infty$, nous obtenons la borne supérieure désirée, uniformément en $x \in I$,

$$\begin{aligned} \sup_{x \in I} \frac{|V_{n,k}(x, \psi)|}{\sqrt{2nh_n \log(1/h_n)}} &\leq \sup_{x \in I} \frac{|V_{n,k}(x, \psi_n)|}{\sqrt{2nh_n \log(1/h_n)}} + \sup_{x \in I} \frac{|\mu_n(x)|}{\sqrt{2nh_n \log(1/h_n)}} \\ &\leq C\sqrt{2\beta_2(\psi)}, \end{aligned}$$

ce qui clôt la démonstration de la proposition 2.4.4, avec $B = C\sqrt{2}$.

Démonstration du théorème 2.4.2

Nous pouvons maintenant achever la démonstration du théorème 2.4.2. Il suffit de combiner la proposition 2.4.4 avec un peu d'analyse. Soit $\gamma > 0$ un nombre réel arbitraire. Nous posons,

$$\psi_\gamma(y) := \psi(y) \mathbf{I}\{|\psi(y)| \leq \gamma\} \quad \text{et} \quad \bar{\psi}_\gamma(y) := \psi(y) \mathbf{I}\{|\psi(y)| > \gamma\}.$$

Lorsque nous considérons le processus $W_{n,k}(\cdot, \psi_\gamma)$, on remarque que l'hypothèse (F.3) est bien vérifiée. Ainsi, d'après le théorème 2.3.1, il s'ensuit que, pour n'importe quel $\gamma > 0$ fixé,

$$\left| \{2nh_n \log(1/h_n)\}^{-1/2} \sup_{x \in I} \{ \pm W_{n,k}(x, \psi_\gamma) \} - \sigma_W(I) \right| = o_{\mathbf{P}}(1).$$

De l'autre côté, en appliquant la proposition 2.4.4 à $\bar{\psi}_\gamma$, nous obtenons,

$$\mathbb{P} \left\{ \sup_{x \in I} \frac{|V_{n,k}(x, \bar{\psi}_\gamma)|}{\sqrt{2nh_n \log(1/h_n)}} > B\beta_2(\bar{\psi}_\gamma) \right\} = o(1).$$

Enfin, comme $s > 2$,

$$\begin{aligned} \beta_2(\bar{\psi}_\gamma) &= \sup_{x \in J} \mathbb{E} \left[\psi(Y)^2 \mathbf{I}\{|\psi(Y)| > \gamma\} \mid X = x \right] \\ &\leq \sup_{x \in J} \mathbb{E} \left[|\psi(Y)|^s \mathbf{I}\{|\psi(Y)| > \gamma\} \mid X = x \right] \times \gamma^{2-s} \\ &\leq \beta_s(\psi) \gamma^{2-s}. \end{aligned}$$

Ceci implique, comme $\beta_s(\psi) < \infty$,

$$\lim_{\gamma \rightarrow \infty} \beta_2(\bar{\psi}_\gamma) = 0.$$

Ainsi, il suffit de choisir un γ suffisamment grand, de telle sorte que la partie non-tronquée devienne négligeable asymptotiquement.

2.5 Généralisation multidimensionnelle du théorème 2.3.2

2.5.1 Le cas où $X \in \mathbb{R}^p$

Le passage au cadre où la variable explicative (ou prédictrice) $X \in \mathbb{R}^p$ ne présente pas de difficultés particulières. Il suffit d'adapter les hypothèses portant sur le noyau et la fenêtre au contexte multivarié. Notons que l'argument de poissonisation utilisé pour la démonstration de la borne inférieure reste valide quelle que soit la dimension des variables aléatoires considérées. Avant de présenter nos théorèmes, nous rappelons quelques uns des résultats centraux de la littérature. Les premiers résultats de la forme loi limite uniforme du logarithme dans un modèle multivarié sont dus à Stute (1984) [134]. Sous des hypothèses classiques, proches de celles énoncées ci-après (cf. corollaire 2.5.1), il obtient une loi limite uniforme pour l'estimateur à noyau [PR] de la densité multivariée. Soient \mathbf{x} et \mathbf{u} des vecteurs de \mathbb{R}^p .

Théorème 2.5.1 Stute (1984)

$$\lim_{n \rightarrow \infty} \left\{ \frac{nh_n^p}{2 \log h_n^{-p}} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{V}} \frac{|\hat{f}_{X;n}(\mathbf{x}) - f_{X;n}(\mathbf{x})|}{\sqrt{f_X(\mathbf{x})}} = \left\{ \int_{\mathbb{R}^p} K^2(\mathbf{u}) d\mathbf{u} \right\}^{1/2}. \quad (2.41)$$

Ce théorème est valable uniformément sur les parallélépipèdes (ou hyper-rectangles) \mathcal{V} compacts tels que la fonction de densité $f_X(\cdot)$ soit différente de zéro. En ce qui concerne la convergence ponctuelle, nous citons le résultat de Deheuvels and Mason (1994) [27] (cf. la remarque 3.4, p. 1657), fondé sur une belle loi du logarithme itéré fonctionnelle concernant une version du processus empirique local (ou processus empirique indexé par un certain ensemble de \mathbb{R}^p du type voisinage d'un point).

Théorème 2.5.2 Deheuvels and Mason (1994)

Soient $\mathbf{x}_i, i = 1, \dots, N$, des points distincts de \mathbb{R}^p . L'ensemble limite de la suite de vecteurs aléatoires dans \mathbb{R}^N définie par

$$\left\{ \left\{ \frac{nh_n^p}{2 \log_2 n} \right\}^{1/2} \left\{ \frac{\hat{f}_{X;n}(\mathbf{x}_i) - f_{X;n}(\mathbf{x}_i)}{\sqrt{f_X(\mathbf{x}_i)}} \right\} \left\{ \int_{\mathbb{R}^p} K^2(\mathbf{u}) d\mathbf{u} \right\}^{-1/2}, i = 1, \dots, N \right\},$$

est presque sûrement égal à la boule unité de \mathbb{R}^N .

Signalons également l'article de Hall (1991) [62] qui utilise une méthodologie différente pour prouver une loi du logarithme itéré pour l'estimateur [NW] dans le cadre du plan fixe multidimensionnel, avec variance conditionnelle supposée unitaire.

Théorème 2.5.3 Hall (1991)

$$\limsup_{n \rightarrow \infty} \left\{ \frac{nh_n^p}{\log \log nh_n^p} \right\}^{1/2} \{ \hat{m}_n(\mathbf{x}) - \mathbb{E}[\hat{m}_n(\mathbf{x})] \} \stackrel{p.s.}{=} \left\{ 2 \frac{\int K^2}{f_X(\mathbf{x})} \right\}^{1/2}$$

Parmi les travaux actuels relatifs aux vitesses de convergence uniforme d'estimateurs à noyaux dans le cadre multivarié, Giné et Guillou (2002) [56] ont complété les résultats de Silverman (1978) [125] et Stute (1984) à propos de l'estimateur à noyau de la densité [PR] multivarié. Leur méthode de démonstration est également fondée sur les travaux de Talagrand ([137] et [138]) et notamment la fameuse inégalité exponentielle de type Borell-Bernstein pour la déviation par rapport à l'espérance de la norme supremum du processus empirique indexé par une classe de fonctions bornée. Comme nous l'avons remarqué précédemment, dans la démonstration de la section 2.4, cette inégalité exponentielle générale est particulièrement efficace lorsque la classe de fonctions considérée satisfait une condition d'entropie (cf. (2.21)) associée à certaines hypothèses de mesurabilité nous permettant d'éviter les mesures de probabilités extérieures. Leur théorème principal, énoncé ci-dessous, établit une loi limite uniforme du logarithme concernant la norme uniforme sur \mathbb{R}^p (et non plus sur un pavé compact) de la déviation par rapport à l'espérance de l'estimateur à noyau de la densité [PR] multivarié. Fait remarquable, ce résultat ne requiert pas la stricte positivité de la densité f_X .

Théorème 2.5.4 Giné and Guillou (2002)

On suppose les hypothèses (K.1–3) sur le noyau (cf. ci-dessous) et la densité $f_X(\cdot)$ bornée et uniformément continue sur \mathbb{R}^p . La fenêtre h_n satisfait

$$h_n \searrow 0, \quad nh_n^p / |\log h_n| \rightarrow \infty, \quad |\log h_n| / \log \log n \rightarrow \infty \quad \text{et} \quad nh_n^p \nearrow \infty,$$

pour un certain $c > 0$. Il s'ensuit

$$\lim_{n \rightarrow \infty} \left\{ \frac{nh_n^p}{2 \log h_n^{-p}} \right\}^{1/2} \sup_{\mathbf{x} \in \mathbb{R}^p} |\hat{f}_{X;n}(\mathbf{x}) - f_{X;n}(\mathbf{x})| \stackrel{p.s.}{=} \sup_{\mathbf{x} \in \mathbb{R}^p} \left\{ f_X(\mathbf{x}) \int_{\mathbb{R}^p} K^2(\mathbf{u}) d\mathbf{u} \right\}^{1/2}.$$

Ce résultat a été également démontré par Deheuvels [24] lorsque la variable X est à valeurs réelles. Récemment, Mason [99] a établi une loi fonctionnelle uniforme du logarithme concernant le processus empirique local au point $z \in \mathbb{R}^d$ indexé par $g \in \mathcal{G}$, défini par,

$$E_n(z, g) := \frac{1}{(nh_n)^{1/2}} \sum_{i=1}^n \left\{ g(h_n^{-1/d}(z - Z_i)) - \mathbb{E}g(h_n^{-1/d}(z - Z_i)) \right\},$$

avec $\{Z_i : 1 \leq i \leq n\}$ des vecteurs aléatoires i.i.d. à valeurs dans \mathbb{R}^d . La méthodologie employée reprend des arguments similaires à [26], mais en utilisant les travaux récents sur le processus empirique indexé par des classes de fonctions, parmi lesquels les principes de grandes déviations fonctionnelles démontrés par Arcones [3] et [4]. Notons que la loi fonctionnelle uniforme du logarithme de Mason permet de nombreuses applications dans le cadre de l'étude de la consistance presque sûre d'estimateurs à noyaux multivariés et règle également le problème de l'uniformité par rapport au noyau.

Cadre de travail et hypothèses

Nous disposons d'un n -échantillon de couples aléatoires (X, Y) à valeurs dans $\mathbb{R}^p \times \mathbb{R}$. La convergence uniforme sera établie sur des hyper-rectangles contenus dans le support

de la densité. Soient $\mathcal{I} = \prod_{i=1}^p [a_i, b_i]$ et $\mathcal{J} = \prod_{i=1}^p [a'_i, b'_i] \supset \mathcal{I}$ deux hyper-rectangles tels que,

$$a'_i < a_i < b_i < b'_i, \quad 1 \leq i \leq p.$$

Les hypothèses sur la distribution du couple (X, Y) restent inchangées et ne suscitent pas de remarques complémentaires.

(F.1) $f_{X,Y}(\cdot, \cdot)$ est continue sur $\mathcal{J} \times \mathbb{R}$;

(F.2) $f_X(\cdot)$ est continue et strictement positive sur \mathcal{J} ;

(F.3) $Y \times \mathbb{1}\{X \in \mathcal{J}\}$ est bornée.

Par contre, le noyau et la fenêtre doivent être adaptés au cadre multidimensionnel avec quelques modifications. Le noyau $K(\cdot)$ est à présent une fonction supposée mesurable définie sur \mathbb{R}^p et à valeurs réelles, c'est à dire $K : \mathbb{R}^p \rightarrow \mathbb{R}$. Il faut notamment ajuster les hypothèses sur K afin que la classe de fonctions

$$\mathbb{K} = \left\{ K\left(\frac{\mathbf{x} - \cdot}{h}\right) : \mathbf{x} \in \mathbb{R}^p, h > 0 \right\}$$

soit une classe de fonctions mesurable ponctuellement (cf. définition A.3.7) et à nombre de recouvrement uniformément polynomial. Dans le modèle univarié, le noyau K était une fonction continue, supposée à variation totale bornée sur \mathbb{R} . Dorénavant, nous supposons les hypothèses suivantes, avec $\mathbf{u} \in \mathbb{R}^p$,

(K.1) $K(\cdot)$ est une fonction bornée, de carré intégrable et de la forme $K(\mathbf{u}) = \zeta(P(\mathbf{u}))$, $P(\cdot)$ désignant un polynôme en p variables et $\zeta(\cdot)$ une fonction mesurable à valeurs réelles et à variation bornée sur \mathbb{R} ;

(K.2) $K(\mathbf{u}) = 0$ pour $\mathbf{u} \notin [-\xi/2; \xi/2]^p$, pour un certain $0 < \xi < \infty$;

(K.3) $\int_{\mathbb{R}^p} K(\mathbf{u}) d\mathbf{u} = 1$;

(K.4) K est k -fois différentiable, avec des dérivées partielles vérifiant (K.1).

Lorsque, par exemple, nous choisissons $K(\cdot)$ comme le produit de noyaux univarié $K_i(\cdot)$, $i = 1, \dots, p$:

$$K(\mathbf{u}) = \prod_{i=1}^p K_i(u_i),$$

la condition (K.1) est bien vérifiée dès lors que chaque K_i est à variation bornée sur \mathbb{R} . Cette construction rejoint la notion de fonctions multivariées à variation bornée au sens de Hardy et Krause (cf. [76] et [91] pour des travaux plus récents concernant la variation totale d'une fonction multivariée).

Les fonctions $f_X(\cdot)$, $r_\psi(\cdot)$ et $m_\psi(\cdot)$ sont définies sur \mathbb{R}^p . Afin d'estimer leurs dérivées partielles, nous introduisons un opérateur de différentiation adéquat. Pour chaque vecteur $\mathbf{x} = (x_1, \dots, x_p) \in \mathcal{J}$ fixé et chaque p -uplet $\mathbf{k} = (k_1, \dots, k_p) \in \mathbb{N}^p$, nous désignons par $D^{(\mathbf{k})}$ l'opérateur défini par

$$D^{(\mathbf{k})} = \left(\frac{\partial}{\partial x_1}\right)^{k_1} \dots \left(\frac{\partial}{\partial x_p}\right)^{k_p} \quad \text{d'ordre } |\mathbf{k}| = k = k_1 + \dots + k_p.$$

Par la suite, pour toute fonction $h : \mathbb{R}^p \rightarrow \mathbb{R}$, on dénote par $h^{(\mathbf{k})} =: D^{(\mathbf{k})}h$ sa dérivée partielle d'ordre k , associée au p -uplet \mathbf{k} .

Sous les conditions (K.1–4), pour chaque p -uplet \mathbf{k} dont la somme de ses éléments est égale à k , la classe de fonctions liée à l'estimation de dérivées partielles d'ordre k :

$$\mathbb{K}_{\mathbf{k}} = \left\{ K^{(\mathbf{k})} \left(\frac{\mathbf{x} - \cdot}{h} \right) : \mathbf{x} \in \mathbb{R}^p, h > 0 \right\}$$

satisfait la condition d'entropie suivante, pour n'importe quelle mesure de probabilité (c'est à dire indépendamment de la distribution du couple (X, Y)),

$$\text{pour certains } C > 0, \nu > 0, \quad N(\epsilon, \mathbb{K}_{\mathbf{k}}) \leq C \epsilon^{-\nu}, \quad 0 < \epsilon < 1.$$

De l'autre coté, le fait que $K^{(\mathbf{k})}$ soit continue (ou continue à droite) entraîne que la classe de fonctions $\mathbb{K}_{\mathbf{k}}$ est mesurable ponctuellement (cf. lemme A.3.3 en annexe).

Les modifications sur la fenêtre sont minimales. En résumé, h_n^p sera substituée à h_n dans les hypothèses (H.1–5) de la précédente section.

(H.1) $h_n \rightarrow 0$, lorsque $n \rightarrow \infty$;

(H.2) $nh_n^p / \log n \rightarrow \infty$, lorsque $n \rightarrow \infty$;

(H.3) $nh_n^{2k+p} / \log(h_n^{-p}) \rightarrow \infty$, lorsque $n \rightarrow \infty$;

(H.4) $h_n \searrow 0$ et $nh_n^p \nearrow \infty$, lorsque $n \rightarrow \infty$;

(H.5) $\log h_n^{-p} / \log \log n \rightarrow \infty$, lorsque $n \rightarrow \infty$.

En vue d'une présentation rigoureuse des estimateurs des dérivées partielles, nous présentons quelques notations additionnelles qui serviront ultérieurement lors du passage au cadre strictement multidimensionnel (c'est à dire lorsque Y n'est plus une variable aléatoire réelle et que $m_\psi : \mathbb{R}^p \rightarrow \mathbb{R}^d$).

Soit $g = (g_1, \dots, g_d)$ avec $g_j : \mathbb{R}^p \rightarrow \mathbb{R}$, $j = 1, \dots, d$. Pour tout p -uplet $\mathbf{k} = (k_1, \dots, k_p)$ d'éléments de \mathbb{N} , tel que $k := |\mathbf{k}| = k_1 + \dots + k_p$, on définit

$$D^{(\mathbf{k})}g = g^{(\mathbf{k})} = (D^{(\mathbf{k})}g_1, \dots, D^{(\mathbf{k})}g_d),$$

où pour tout $j = 1, \dots, d$:

$$D^{(\mathbf{k})}g_j = g_j^{(\mathbf{k})} = \left(\frac{\partial}{\partial x_1} \right)^{k_1} \dots \left(\frac{\partial}{\partial x_p} \right)^{k_p} g_j.$$

Remarque 2.5.1 Pour les p -uplets de \mathbb{N} , \mathbf{k}_0 tel que $|\mathbf{k}_0| = 0$, et \mathbf{k}_j , $1 \leq j \leq p$, tels que $|\mathbf{k}_j| = k_j = 1$, nous avons :

$$g^{(\mathbf{k}_0)} = g = (g_1, \dots, g_d) \quad \text{et} \quad g^{(\mathbf{k}_j)} = \left(\frac{\partial g_1}{\partial x_j}, \dots, \frac{\partial g_d}{\partial x_j} \right).$$

On rappelle la définition des estimateurs à noyaux de $f_X(\mathbf{x})$, $r_\psi(\mathbf{x})$ et $m_\psi(\mathbf{x})$:

$$\hat{f}_{X;n}(\mathbf{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n K \left(\frac{\mathbf{x} - X_i}{h_n} \right), \quad \hat{r}_{\psi;n}(\mathbf{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n \psi(Y_i) K \left(\frac{\mathbf{x} - X_i}{h_n} \right),$$

$$\hat{m}_{\psi;n}(\mathbf{x}) = \frac{\hat{r}_{\psi;n}(\mathbf{x})}{\hat{f}_{X;n}(\mathbf{x})} \quad \text{lorsque } \hat{f}_{X;n}(\mathbf{x}) \neq 0.$$

Les estimateurs des dérivées partielles d'ordre k de $f_X(\mathbf{x})$ et $r_\psi(\mathbf{x})$ sont définis par :

$$\begin{aligned} \hat{f}_{X;n}^{(\mathbf{k})}(\mathbf{x}) &= D^{(\mathbf{k})}(\hat{f}_{X;n}(\mathbf{x})) = \frac{1}{nh_n^{k+p}} \sum_{i=1}^n K^{(\mathbf{k})}\left(\frac{\mathbf{x} - X_i}{h_n}\right), \\ \hat{r}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) &= D^{(\mathbf{k})}(\hat{r}_{\psi;n}(\mathbf{x})) = \frac{1}{nh_n^{k+p}} \sum_{i=1}^n \psi(Y_i) K^{(\mathbf{k})}\left(\frac{\mathbf{x} - X_i}{h_n}\right). \end{aligned}$$

Pour la fonction de régression, nous limitons notre exposition au cas où $k = k_j = 1$ (cf. remarque 2.5.1 ci-dessus) par souci de concision. Lorsque $\hat{f}_{X;n}(\mathbf{x}) \neq 0$,

$$\hat{m}_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x}) = D^{(\mathbf{k}_j)}(\hat{m}_{\psi;n}(\mathbf{x})) = \frac{\hat{r}_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x})}{\hat{f}_{X;n}(\mathbf{x})} - \frac{\hat{r}_{\psi;n}(\mathbf{x}) \hat{f}_{X;n}^{(\mathbf{k}_j)}(\mathbf{x})}{\hat{f}_{X;n}^2(\mathbf{x})},$$

et plus généralement $\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) = D^{(\mathbf{k})}(\hat{m}_{\psi;n}(\mathbf{x}))$. Les termes de centrages seront de la même forme que dans le cas univarié.

Théorèmes

Théorème 2.5.5 *Supposons les hypothèses (F.1-3), (H.1-3), (K.1-4) vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ 2nh_n^p \log(h_n^{-p}) \right\}^{-1/2} \sup_{\mathbf{x} \in \mathcal{I}} \{ \pm W_{n,\mathbf{k}}(\mathbf{x}, \psi) \} - \sigma_W(\mathcal{I}) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_W^2(\mathcal{I}) = \sup_{\mathbf{x} \in \mathcal{I}} \mathbb{E} \left[\left(c(\mathbf{x})\psi(Y) + d(\mathbf{x}) \right)^2 \middle| X = \mathbf{x} \right] f_X(\mathbf{x}) \int_{\mathbf{R}^p} [K^{(\mathbf{k})}(\mathbf{u})]^2 d\mathbf{u}. \quad (2.42)$$

Si la fenêtre satisfait (H.3-5), nous obtenons,

$$\left| \left\{ 2nh_n^p \log(h_n^{-p}) \right\}^{-1/2} \sup_{\mathbf{x} \in \mathcal{I}} \{ \pm W_{n,k}(\mathbf{x}, \psi) \} - \sigma_W(\mathcal{I}) \right| = o(1), \quad \text{presque sûrement.}$$

Corollaire 2.5.1 *Supposons les hypothèses (F.2), (H.1-3), (K.1-4) vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{f}_{X;n}^{(\mathbf{k})}(\mathbf{x}) - f_{X;n}^{(\mathbf{k})}(\mathbf{x}) \} - \sigma_f(\mathcal{I}) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_f^2(\mathcal{I}) = \sup_{\mathbf{x} \in \mathcal{I}} f_X(\mathbf{x}) \int_{\mathbf{R}^p} [K^{(\mathbf{k})}(\mathbf{u})]^2 d\mathbf{u}.$$

Si la fenêtre satisfait (H.3-5), nous obtenons,

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{f}_{X;n}^{(\mathbf{k})}(\mathbf{x}) - f_{X;n}^{(\mathbf{k})}(\mathbf{x}) \} - \sigma_f(\mathcal{I}) \right| = o(1), \quad \text{presque sûrement.}$$

Corollaire 2.5.2 *Supposons les hypothèses (F.1–3), (H.1–3), (K.1–4) vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{r}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - r_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) \} - \sigma_r(\mathcal{I}) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_r^2(\mathcal{I}) = \sup_{\mathbf{x} \in \mathcal{I}} \sigma_{\psi}^2(\mathbf{x}) f_X(\mathbf{x}) \int_{\mathbf{R}^p} [K^{(\mathbf{k})}(\mathbf{u})]^2 d\mathbf{u}.$$

Si la fenêtre satisfait (H.3–5), nous obtenons,

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{r}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - r_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) \} - \sigma_r(\mathcal{I}) \right| = o(1), \quad \text{presque sûrement.}$$

Théorème 2.5.6 *Supposons les hypothèses (F.1–3), (H.1–3), (K.1–4) vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \} - \sigma_m(\mathcal{I}) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_m^2(\mathcal{I}) = \sup_{\mathbf{x} \in \mathcal{I}} \left\{ \frac{\sigma_{\psi}^2(\mathbf{x})}{f_X(\mathbf{x})} \right\} \int_{\mathbf{R}^p} [K^{(\mathbf{k})}(\mathbf{u})]^2 d\mathbf{u}.$$

Si la fenêtre satisfait (H.3–5), nous obtenons,

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \} - \sigma_m(\mathcal{I}) \right| = o(1), \quad \text{presque sûrement.}$$

La démonstration des théorèmes et corollaires ci-dessus est similaire au cas réel. Cette fois-ci la discrétisation s'effectue sur un hyper-rectangle p -dimensionnel.

Idée de démonstration

Nous fixons $p = 2$ afin d'éviter des notations trop lourdes. Nous posons, pour $n \geq 1$,

$$\mathbf{z}_{n,i,j} = \begin{pmatrix} a_1 + i\delta h_n \\ a_2 + j\delta h_n \end{pmatrix} \text{ lorsque } \begin{cases} 0 \leq i \leq l_{n,1} := \lfloor (b_1 - a_1)/(\delta h_n) \rfloor, \\ 0 \leq j \leq l_{n,2} := \lfloor (b_2 - a_2)/(\delta h_n) \rfloor. \end{cases}$$

Pour $u \in \mathbb{R}^2$ et $v \in \mathbb{R}$,

$$g_{n,i,j}(u, v) := \eta_{n, \mathbf{z}_{n,i,j}, \mathbf{k}}, \quad 0 \leq i \leq l_{n,1} \text{ et } 0 \leq j \leq l_{n,2},$$

en référence à (2.11). Nous étudions le processus empirique sur la classe de fonctions suivante, pour $n \geq 1$,

$$\mathcal{G}_n := \left\{ g_{n,i,j} : 0 \leq i \leq l_{n,1}, \quad 0 \leq j \leq l_{n,2} \right\}.$$

Puis, nous déterminons une borne pour la variance, uniformément en $0 \leq i \leq l_{n,1}$ et $0 \leq j \leq l_{n,2}$. Lorsque n est suffisamment grand,

$$\text{Var}(g_{n,i,j}(X, Y)) \leq \sigma_W^2(\mathcal{I})h_n^2(1 + \tau),$$

avec $\tau > 0$ arbitraire. Il s'ensuit, via une application de l'inégalité de Bernstein ou résultat 2.4.1, pour n suffisamment grand,

$$\begin{aligned} \mathbb{P}\left\{\max_{0 \leq i \leq l_{n,1}; 0 \leq j \leq l_{n,2}} \frac{|\alpha_n(g_{n,i,j})|}{\sqrt{2h_n^2 \log(h_n^{-2})}} > \sigma_W(\mathcal{I})(1 + \tau)\right\} &\leq 2(l_{n,1} + 1)(l_{n,2} + 1)h_n^{2(1+\tau/2)} \\ &= O(h_n^\tau) = o(1). \end{aligned}$$

Cette dernière inégalité constitue la première étape de la démonstration de la borne supérieure. Par la suite, on cherche à contrôler les incréments du processus empiriques sur des petits pavés p multidimensionnels. On ne rentrera pas dans les détails de la démonstration par souci de clarté, mais les arguments restent identiques au cadre univarié.

2.5.2 Le cas strictement multivarié : $\psi(Y) \in \mathbb{R}^d$

Rappelons que d'après le théorème 2.5.6 (lorsque $\psi(Y) \in \mathbb{R}$), sous les hypothèses (F.1–3), (H.3–5) et (K.1–4), nous avons, lorsque $n \rightarrow \infty$

$$\left|\left\{\frac{nh_n^{2k+p}}{2 \log(h_n^{-p})}\right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{\hat{m}_{\psi;n}^{(k)}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(\mathbf{x})]\} - \sigma_m(\mathcal{I})\right| \stackrel{p.s.}{=} o(1),$$

où

$$\sigma_m^2(\mathcal{I}) = \sup_{\mathbf{x} \in \mathcal{I}} \left\{ \frac{\sigma_\psi^2(\mathbf{x})}{f_X(\mathbf{x})} \right\} \int_{\mathbb{R}^p} [K^{(k)}(\mathbf{u})]^2 d\mathbf{u} =: \sup_{\mathbf{x} \in \mathcal{I}} \sigma_m^2(\mathbf{x}). \quad (2.43)$$

Remarque 2.5.2 D'après les résultats énoncés dans le premier chapitre concernant la normalité asymptotique et l'étude de la variance dans le cadre multidimensionnel, $\sigma_m^2(\mathcal{I})$ correspond également au supremum sur le pavé \mathcal{I} de la variance asymptotique de $\sqrt{nh_n} \times \{\hat{m}_{\psi;n}^{(k)}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(\mathbf{x})]\}$. Ainsi, les lois uniformes du logarithme que nous présentons peuvent être vues comme une version uniforme presque sûre du théorème central limite (cf. [89], chapitre 8, pour plus de détails).

Le but de cette section est de généraliser le théorème 2.5.6 au cas multidimensionnel, c'est à dire lorsque $\psi(Y) \in \mathbb{R}^d$, avec $d > 1$. Il est toutefois plus difficile d'obtenir des théorèmes limites pour la convergence uniforme. La difficulté majeure est liée à la localisation du supremum de suites de variables multivariées à valeurs dans un espace euclidien de dimension strictement supérieure à un. De plus, la méthode développée par Finkelstein (1971) pour étendre au cadre multidimensionnel la loi du logarithme itéré de Hartman-Wintner se prête mal au caractère uniforme de nos résultats. Néanmoins, on peut contourner ce problème, via une normalisation adéquate. Nous citons, en avant-propos, quelques résultats remarquables démontrés par Einmahl (cf. [37]) concernant la loi du logarithme itéré de variables aléatoires à valeurs dans un espace de Banach et, plus particulièrement, dans

un espace euclidien de dimension 2. Ces travaux nous donnent un aperçu des ensembles limites possibles.

Nous désignons par B un espace de Banach séparable, de norme $\|\cdot\|$, et B^* son dual topologique. Nous supposons que X, X_1, \dots , sont des variables aléatoires i.i.d. à valeurs dans B , telles que $0 < \mathbb{E}[\|X\|] < \infty$. Dans un premier temps, nous présentons une notion délicate, introduite par Klass (1976) [84] qui permet de normaliser la somme de variables aléatoires à valeurs dans un espace de Banach et, de facto, de formuler des Lois du Logarithme Itéré dans un cadre très général (cf. [36], [37] et [38]). Pour n'importe quelle variable aléatoire ξ à valeurs réelles, telle que $0 < \mathbb{E}[|\xi|] < \infty$, on associe une fonction $K_\xi(\cdot)$, définie comme la fonction inverse d'une fonction auxiliaire $G_\xi(\cdot)$, donnée par

$$G_\xi(y) := y^2 \left\{ \int_0^y \mathbb{E}[|\xi| \mathbb{I}\{|\xi| > u\}] du \right\}^{-1}, \quad y > 0.$$

Maintenant, pour n'importe quelle fonctionnelle $f \in B^*$ telle que $\mathbb{E}[f(X)] > 0$, soit K_f la K -fonction correspondant à la variable aléatoire $f(X)$ à valeurs réelles, et posons

$$\begin{aligned} \tilde{K}(y) &:= \sup \{K_f(y) : \|f\| \leq 1\}, \quad y > 0, \\ \gamma_n &:= \sqrt{2} \tilde{K}(n/\log_2 n) \log_2 n, \quad n \geq 3, \end{aligned}$$

où $\log_2 n$ désigne le logarithme itéré.

On se concentre, à présent, sur le cas particulier où $B = \mathbb{R}^2$ et $\|\cdot\|$ désigne la norme Euclidienne. Soit $X = (X^{(1)}, X^{(2)})$ un vecteur aléatoire bidimensionnel et soient X_1, \dots, X_n , des copies indépendantes du vecteur aléatoire X . On pose, par convenance,

$$S_n^{(1)} := X_1^{(1)} + \dots, X_n^{(1)}, \quad \text{et} \quad S_n^{(2)} := X_1^{(2)} + \dots, X_n^{(2)}.$$

Comme l'espace \mathbb{R}^2 est clairement un espace de Banach de type 2, en adaptant le corollaire 2, p. 2017, [36], Einmahl présente le résultat suivant.

Théorème 2.5.7 Einmahl (1995)

Soit $X = (X^{(1)}, X^{(2)})$ un vecteur aléatoire centré, tel que $0 < \mathbb{E}[\|X\|] < \infty$. Alors, nous avons,

$$\limsup_{n \rightarrow \infty} \frac{\|S_n\|}{\gamma_n} \stackrel{p.s.}{=} 1, \tag{2.44}$$

si et seulement si,

$$\sum_n \mathbb{P}\{|X^{(i)}| > \gamma_n\} < \infty, \quad i = 1, 2. \tag{2.45}$$

Soit A l'ensemble limite de $\{S_n/\gamma_n\}$, constitué de tous les points ou valeurs d'adhérence. Nous remarquons, d'après (2.44), que A est un sous-ensemble du disque unité et, de plus,

$$\sup \{\|x\| : x \in A\} = 1.$$

Einmahl [37] démontre également que l'ensemble limite A est symétrique et étoilé par rapport à l'origine. Si nous supposons l'assertion suivante vérifiée,

$$\text{les composantes } X^{(1)} \text{ et } X^{(2)} \text{ sont indépendantes,} \tag{2.46}$$

les différents ensembles limites sont alors contenus dans une classe d'ensembles pouvant être représentés comme les fermetures d'unions d'ellipses dénombrables.

Pour finir cet exposé, nous présentons quelques résultats instructifs concernant le cas indépendant, toujours d'après Einmahl (1995). On pose, pour $0 < a, b < \infty$,

$$\mathcal{E}(a, b) := \left\{ \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : \left\{ \frac{x_1}{a} \right\}^2 + \left\{ \frac{x_2}{b} \right\}^2 \leq 1 \right\}$$

et

$$\mathcal{E}(a, 0) := [-a, a] \times \{0\}, \quad \mathcal{E}(0, b) := \{0\} \times [-b, b], \quad \text{pour } a, b \geq 0.$$

On désigne par $\text{cl}(M)$ (en référence à “*closure*”) la fermeture d'un sous-ensemble arbitraire M appartenant à \mathbb{R}^2 . Nous définissons,

$$\sigma_1 := \limsup_{n \rightarrow \infty} \frac{|S_n^{(1)}|}{\gamma_n},$$

et

$$\sigma_2 := \limsup_{n \rightarrow \infty} \frac{|S_n^{(2)}|}{\gamma_n},$$

en observant que $\sigma_1 \vee \sigma_2 \leq 1$, d'après (2.44).

Théorème 2.5.8 Einmahl (1995)

Soit X un vecteur aléatoire satisfaisant les conditions (2.45) et (2.46). Alors, nous avons, pour des suites convenables $0 \leq a_m \leq \sigma_1$, $0 \leq b_m \leq \sigma_2$,

$$A = \mathcal{E}(\sigma_1, 0) \cup \mathcal{E}(0, \sigma_2) \cup \text{cl} \left\{ \bigcup_{m=1}^{\infty} \mathcal{E}(a_m, b_m) \right\}.$$

Théorèmes et démonstrations

L'approche développée par Einmahl et Mason (2000) et Deheuvels et Mason (2004) n'est pas directement applicable dans le cadre strictement multivarié, où $\psi(Y) \in \mathbb{R}^d$. En effet, la majorité des travaux sur les processus empiriques indexés par des classes de fonctions sont concernés par des classes de fonctions à valeurs réelles. Plus particulièrement, l'argumentation principale concernant la démonstration de la borne supérieure et le contrôle des oscillations du processus empirique repose sur une borne exponentielle pour le supremum du processus empirique indexé par une classe de fonctions mesurables et à valeurs réelles. Par contre, si nous examinons la convergence ponctuelle de la déviation par rapport à l'espérance de notre estimateur multivarié de la dérivée partielle d'ordre \mathbf{k} de la fonction de régression, il est aisé de démontrer une loi du logarithme itéré, via les travaux de Einmahl et Mason [40] et [41] combinés au lemme 2 de Finkelstein (1971) [50].

Le cas ponctuel

Nous supposons que la fenêtre satisfait les conditions suivantes :

$$(H.6) \quad h_n \searrow 0, \quad nh_n \nearrow \infty, \quad nh_n / \log_2 n \rightarrow \infty.$$

Pour démontrer notre théorème principal, on s'appuie sur le résultat ponctuel suivant ou loi du logarithme itéré, lorsque $\psi(Y) \in \mathbb{R}$ ($d = 1$) :

Théorème 2.5.9 *Sous les hypothèses (F.1–3), (H.6), (K.1–4), nous avons,*

$$\limsup_{n \rightarrow \infty} \left\{ \frac{nh_n^{2k+p}}{2 \log_2 n} \right\}^{1/2} \pm \left\{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \right\} \stackrel{p.s.}{=} \sigma_m(\mathbf{x}).$$

C'est une légère modification de la démonstration du théorème 3, p. 80, de Einmahl et Mason (1998) [41], qui constitue la première démonstration valide de loi du logarithme itéré concernant l'estimateur de la régression [NW]. \square

Remarque 2.5.3 Ce dernier théorème s'appuie donc sur les travaux de Einmahl et Mason. Dans un premier temps, ils ont établi une approximation forte du processus empirique local, extension d'une notion plus ancienne développée par Deheuvels et Mason (1994) [27]. Cette approximation forte permet alors de formuler une loi du logarithme itéré compacte pour le processus empirique local. Enfin, ils en déduisent diverses lois du logarithme itéré concernant l'estimateur à noyau de la densité ou de la régression (cf. exemples A.1.2 et A.1.3 en annexe).

Dans le cadre multidimensionnel, la fonction $\psi(\cdot)$ est supposée mesurable et bornée sur tout compact dans \mathbb{R}^d . L'hypothèse (F.1) devient

$$(F.1) \text{ Pour chaque } x \in J, \lim_{x' \rightarrow x; x' \in J} f_{X,Y}(x', y) = f_{X,Y}(x, y) \text{ pour presque tout } y \in \mathbb{R}^d.$$

Remarque 2.5.4 Plus généralement, nous pouvons supposer les variables Y_i , $1 \leq i \leq n$, à valeurs dans \mathbb{R}^q et $\psi : \mathbb{R}^q \rightarrow \mathbb{R}^d$ une fonction borélienne bornée.

Les hypothèses sur le noyau et la fenêtre sont inchangées mais nous avons à introduire certaines notations. Par la suite, nous ferons souvent référence à la matrice de variance-covariance asymptotique suivante (sous réserve de son existence), équivalente d -dimensionnelle de $\sigma_m^2(\mathbf{x})$ définie en (2.43),

$$V_{\mathbf{x}} := \left\{ \frac{1}{f_X(\mathbf{x})} \int_{\mathbb{R}^p} [K^{(\mathbf{k})}(\mathbf{u})]^2 d\mathbf{u} \right\} \times \Sigma_{\psi}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{I}, \quad (2.47)$$

où $\Sigma_{\psi}(\mathbf{x})$ désigne la matrice de variance-covariance de $\psi(Y)$ conditionnelle à $X = x$. Sans perte de généralité, la matrice $\Sigma_{\psi}(\mathbf{x})$ sera supposée strictement définie positive afin de garantir son inversibilité. Notre premier résultat constitue une simple extension multidimensionnelle du théorème 2.5.9. Soit

$$\mathcal{M}_{n,d}(\mathbf{x}) := \left\{ \frac{nh_n^{2k+p}}{2 \log_2 n} \right\}^{1/2} \pm \left\{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \right\}.$$

La matrice $V_{\mathbf{x}}$ étant strictement définie positive et inversible pour chaque $\mathbf{x} \in \mathcal{I}$, nous obtenons le lemme suivant.

Théorème 2.5.10 *Sous les hypothèses (F.1–3), (H.6), (K.1–4), nous avons, pour chaque $\mathbf{x} \in \mathcal{I}$,*

$$C\left(\{\mathcal{M}_{n,d}\}(\mathbf{x})\right) \stackrel{p.s.}{=} \{\mathbf{y} \in \mathbb{R}^d : \mathbf{y}^T V_{\mathbf{x}}^{-1} \mathbf{y} \leq 1\} =: \mathcal{E}_{\mathbf{x}},$$

et

$$C\left(V_{\mathbf{x}}^{-1/2} \{\mathcal{M}_{n,d}\}(\mathbf{x})\right) \stackrel{p.s.}{=} \{\mathbf{y} \in \mathbb{R}^d : \mathbf{y}^T \mathbf{y} \leq 1\} =: \mathcal{B}_d. \quad (2.48)$$

La démonstration s'appuie sur une adaptation du lemme 2 de Finkelstein (1971) (cf. annexe, section A.4). Nous posons

$$\theta_{n,k} := \left\{ \frac{nh_n^{2k+p}}{2 \log_2 n} \right\} \quad \text{et} \quad \mathcal{D}_{n,k,d}(\mathbf{x}) := \pm \left\{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \right\},$$

où k désigne le degré de dérivation et d la dimension du vecteur aléatoire Y .

Le théorème 2.5.9, qui traite le cas où Y est à valeurs réelles, est alors équivalent à

$$\limsup_{n \rightarrow \infty} \left\{ \theta_{n,k} \right\}^{1/2} \left\{ \mathcal{D}_{n,k,1}(\mathbf{x}) \right\} \stackrel{p.s.}{=} \sigma_m(\mathbf{x}).$$

Soient \mathbf{y} un vecteur de \mathbb{R}^d et \mathbf{y}^T son transposé. Maintenant, en utilisant le fait que la matrice $V_{\mathbf{x}}$ est inversible, nous obtenons, $\forall \mathbf{y} \in \mathbb{R}^d$,

$$\limsup_{n \rightarrow \infty} \left\{ \theta_{n,k} \right\}^{1/2} \mathbf{y}^T \left\{ V_{\mathbf{x}}^{-1/2} \left\{ \mathcal{D}_{n,k,d}(\mathbf{x}) \right\} \right\} = \|\mathbf{y}\|_d, \quad \text{presque sûrement.} \quad (2.49)$$

Remarque 2.5.5 La matrice de variance-covariance asymptotique associée au vecteur aléatoire $\left\{ V_{\mathbf{x}}^{-1/2} \left\{ \mathcal{D}_{n,k,d}(\mathbf{x}) \right\} \right\}$ est la matrice identité d -dimensionnelle.

D'après (2.49), pour un choix convenable d'une suite de vecteurs $\{\mathbf{y}_n\}_{n \geq 1}$, il s'ensuit

$$\limsup_{n \rightarrow \infty} \left\{ \theta_{n,k} \right\}^{1/2} \times \left\| V_{\mathbf{x}}^{-1/2} \left\{ \mathcal{D}_{n,k,d}(\mathbf{x}) \right\} \right\|_d = 1, \quad \text{presque sûrement.} \quad (2.50)$$

Soit $\mathcal{S}_d := \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\|_d = 1\}$ la sphère unité d -dimensionnelle. Pour chaque $\mathbf{y}_0 \in \mathcal{S}_d$, via (2.49) et (2.50), nous avons

$$\liminf_{n \rightarrow \infty} \left\| \left\{ \theta_{n,k} \right\}^{1/2} \left\{ V_{\mathbf{x}}^{-1/2} \left\{ \mathcal{D}_{n,k,d}(\mathbf{x}) \right\} \right\} - \mathbf{y}_0 \right\|_d^2 = 0, \quad \text{presque sûrement.} \quad (2.51)$$

Les équations (2.50) et (2.51) entraînent,

$$C\left(V_{\mathbf{x}}^{-1/2} \left\{ \mathcal{M}_{n,d} \right\}(\mathbf{x})\right) \subseteq \mathcal{B}_d \quad \text{p.s.}, \quad (2.52)$$

et

$$C\left(V_{\mathbf{x}}^{-1/2} \left\{ \mathcal{M}_{n,d} \right\}(\mathbf{x})\right) \supseteq \mathcal{B}_d \quad \text{p.s.}. \quad (2.53)$$

D'après (2.52) et (2.53), nous obtenons bien (2.48). La deuxième partie du lemme vient en utilisant $V_{\mathbf{x}}^{1/2}$ comme un opérateur linéaire sur des ensembles de \mathbb{R}^d . Il s'ensuit,

$$C(\{\mathcal{M}_{n,d}(x)\}) = V_{\mathbf{x}}^{1/2} \{\mathcal{B}_d\} = \mathcal{E}_{\mathbf{x}}, \quad \text{presque sûrement.}$$

□

Le cadre uniforme

En argumentant comme précédemment, nous obtenons également une extension du théorème 2.5.6. Soit

$$\mathcal{R}_{n,d}(\mathbf{x}) := \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \pm \left\{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \right\}.$$

Théorème 2.5.11 *Sous les hypothèses (F.1–3), (H.3–5), (K.1–4), nous obtenons,*

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{x} \in \mathcal{I}} \left\{ V_{\mathbf{x}}^{-1/2} \mathcal{R}_{n,d}(\mathbf{x}) \right\} \stackrel{p.s.}{=} \{ \mathbf{y} \in \mathbb{R}^d : \mathbf{y}^T \mathbf{y} = 1 \} =: \mathcal{S}_d. \quad (2.54)$$

Remarque 2.5.6 La généralisation multidimensionnelle s’appuie donc essentiellement sur une normalisation appropriée de la déviation $\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})]$. Par contre, il apparaît plus difficile de déterminer l’ensemble limite de

$$\left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \left\{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \right\}.$$

On pourrait conjecturer que l’ensemble limite est contenu dans une union infinie d’ellipsoïdes de la forme :

$$\bigcup_{\mathbf{x} \in \mathcal{I}} \mathcal{E}_{\mathbf{x}}.$$

Une application utile : l’estimation de la fonction de répartition conditionnelle

Nous finissons cette section, en présentant un exemple d’application pour la fonction ψ . Dans le cadre multidimensionnel, on a supposé la fonction $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ mesurable et bornée sur les ensembles compacts de \mathbb{R}^d . On peut proposer une formulation plus générale :

la fonction $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^q$ est borélienne et bornée,

où $q \in \mathbb{N}$ quelconque. Cette hypothèse nous permet de traiter aisément le cas particulier de l’estimation non-paramétrique de la fonction de répartition conditionnelle à partir d’un échantillon de même loi que le couple (X, Y) à valeurs dans $\mathbb{R}^p \times \mathbb{R}^d$. On pose, pour $\mathbf{y} \in \mathbb{R}^d$,

$$\psi(\mathbf{y}) = \mathbb{I}\{\mathbf{y} \leq \mathbf{t}\}, \quad \text{avec } \mathbf{t} \in \mathbb{R}^d \text{ arbitraire mais fixé.} \quad (2.55)$$

En remplaçant la définition (2.55) de $\psi(\cdot)$ ci-dessus dans la définition de la fonction de régression (2.1), nous obtenons la fonction de répartition conditionnelle, définie par,

$$F(\mathbf{t}|\mathbf{x}) := \mathbb{E}\left[\mathbb{I}\{Y \leq \mathbf{t}\} | X = \mathbf{x}\right] = \mathbb{P}\{Y \leq \mathbf{t} | X = \mathbf{x}\}, \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

On rappelle la définition de l’estimateur à noyau de la fonction de répartition conditionnelle, $\forall \mathbf{x} \in \mathbb{R}^p$,

$$\hat{F}_n(\mathbf{t}|\mathbf{x}) := \frac{\sum_{i=1}^n \mathbb{I}\{Y_i \leq \mathbf{t}\} K\left(\frac{\mathbf{x} - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h_n}\right)} \times \mathbb{I}\left\{\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h_n}\right) \neq 0\right\},$$

et son terme de centrage associé,

$$F_n(\mathbf{t}|\mathbf{x}) := \frac{r_{n;\psi}(\mathbf{x})}{f_{n;X}(\mathbf{x})} \times \mathbb{I}\{f_{n;X}(\mathbf{x}) \neq 0\},$$

lorsque ψ vérifie (2.55). Il s’ensuit, d’après le théorème 2.5.6 avec $k = 0$, le corollaire suivant.

Corollaire 2.5.3 *Supposons les hypothèses (F.1–3), (H.1–3), (K.1–4) vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^p}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{F}_n(\mathbf{t}|\mathbf{x}) - F_n(\mathbf{t}|\mathbf{x}) \} - \sigma_F(\mathcal{I}) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_F^2(\mathcal{I}) = \sup_{\mathbf{x} \in \mathcal{I}} \left\{ \frac{F(\mathbf{t}|\mathbf{x}) \{1 - F(\mathbf{t}|\mathbf{x})\}}{f_X(\mathbf{x})} \right\} \int_{\mathbf{R}^p} K^2(\mathbf{u}) d\mathbf{u}.$$

Si la fenêtre satisfait (H.3–5), nous obtenons,

$$\left| \left\{ \frac{nh_n^p}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{F}_n(\mathbf{t}|\mathbf{x}) - F_n(\mathbf{t}|\mathbf{x}) \} - \sigma_F(\mathcal{I}) \right| \stackrel{p.s.}{=} o(1).$$

Ce dernier résultat améliore les résultats antérieurs de Stute (1986) ([135] et [136]). En s'appuyant sur les travaux de Einmahl et Mason (2000), nous obtenons également le raffinement suivant :

$$\lim_{n \rightarrow \infty} \left\{ \frac{nh_n^p}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{t} \in \mathbf{R}^d} \sup_{\mathbf{x} \in \mathcal{I}} \pm \{ \hat{F}_n(\mathbf{t}|\mathbf{x}) - F_n(\mathbf{t}|\mathbf{x}) \} \stackrel{p.s.}{=} \|K\|_2 / \left\{ 2 \inf_{\mathbf{x} \in \mathcal{I}} \sqrt{f_X(\mathbf{x})} \right\}.$$

Cette dernière loi limite uniforme du logarithme est une extension directe du corollaire 2, p. 6, [42]. Ce dernier corollaire s'appuie sur leur théorème 1, p. 4–5, présenté ci-dessous dans le cas borné. Soit \mathcal{F} une classe de fonctions f mesurables et à valeurs réelles. Pour chaque fonction $f \in \mathcal{F}$ et toutes fonctions c_f et d_f définies sur un intervalle J , on pose, pour $x \in J$,

$$\begin{aligned} W_n(x, f) &= \sum_{i=1}^n \left(c_f(x) f(Y_i) + d_f(x) \right) K \left(\frac{x - X_i}{h_n} \right) \\ &\quad - n \mathbb{E} \left\{ \left(c_f(x) f(Y) + d_f(x) \right) K \left(\frac{x - X}{h_n} \right) \right\}. \end{aligned}$$

Théorème 2.5.12 Einmahl et Mason (2000)

Supposons les hypothèses du théorème 2.3.1 vérifiées. Nous supposons que \mathcal{F} est une VC classe de fonctions mesurable ponctuellement et bornée. Alors, nous avons,

$$\lim_{n \rightarrow \infty} \left\{ 2nh_n |\log h_n| \right\}^{-1/2} \sup_{f \in \mathcal{F}} \sup_{x \in I} |W_n(x, f)| \stackrel{p.s.}{=} \sigma_W(\mathcal{F}, I),$$

où

$$\sigma_W(\mathcal{F}, I) = \sup_{f \in \mathcal{F}} \sigma_W(I).$$

A présent, nous remarquons que la classe de fonctions $\mathcal{F} = \{f_t(\cdot) = \mathbb{I}\{\cdot \leq t\} : t \in \mathbb{R}\}$ est clairement une classe VC de fonctions bornée. En somme, si \mathcal{F} désigne une classe dénombrable (ou “pointwise measurable”) VC de fonctions uniformément bornée ou admettant une fonction enveloppe mesurable avec un moment d'ordre $p > 2$ fini, nous pouvons indexer notre estimateur de la régression par \mathcal{F} et obtenir également une loi limite uniforme du logarithme. La constante limite est alors le supremum sur la classe \mathcal{F} et l'intervalle I de la variance asymptotique.

2.6 Lois limites presque sûres pour les estimateurs localement polynomiaux

La méthodologie que nous utilisons permet également de traiter la convergence uniforme presque sûre d'estimateurs plus sophistiqués tels les estimateurs [PL], introduits dans la section 1.7. Nous rappelons que les estimateurs [PL] ou par polynômes locaux possèdent de meilleures propriétés théoriques et pratiques que les estimateurs des dérivées de la régression de type [NW], notamment en ce qui concerne le biais. L'idée de démonstration consiste simplement à approcher la déviation (par rapport à l'espérance modifiée) associée à ces estimateurs par une version linéarisée, équivalente au processus empirique indexé par une certaine fonction bornée.

Dans cette section, nous démontrons donc une loi limite uniforme du logarithme concernant la déviation maximale de l'estimateur localement linéaire [LL] (i.e. l'estimateur [PL](1)). Puis, nous présentons une généralisation de cette loi à l'estimation localement polynomiale d'ordre $l > k \geq 1$ des dérivées d'ordre k de la fonction de régression. Ces résultats, associés à la construction d'intervalles de confiance, peuvent donner des informations visuelles intéressantes sur les propriétés de régularité de la courbe de régression du modèle considéré. D'autre part, en notant que les estimateurs par lissage polynomial local reproduisent les polynômes, il serait intéressant de construire un test statistique asymptotique via nos lois limites.

Premièrement, nous rappelons certaines notations du premier chapitre intervenant dans la construction des estimateurs [PL] :

$$S_{n,j} = \sum_{i=1}^n (X_i - x)^j K\left(\frac{X_i - x}{h_n}\right),$$

ce qui implique

$$S_{n,j} = nh^j f_X(x) \mu_j[K] \{1 + o_{\mathbf{P}}(1)\},$$

où $\mu_j[K]$ désigne le moment d'ordre j du noyau K .

L'estimateur [LL] de la régression, noté $\hat{m}_n^{LL}(x)$, est défini par :

$$\begin{aligned} \hat{m}_n^{LL}(x) &= e_1^T S_n^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (\text{lorsque } p = 1) \\ &= \frac{1}{S_{n,0} S_{n,2} - S_{n,1}^2} \left\{ S_{n,2} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) - S_{n,1} \sum_{i=1}^n Y_i (X_i - x) K\left(\frac{X_i - x}{h_n}\right) \right\} \end{aligned}$$

Nous posons,

$$\begin{aligned} \hat{f}_n(x) &= \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \\ \hat{f}_{n,1}(x) &= \frac{1}{nh_n} \sum_{i=1}^n \left\{ \frac{x - X_i}{h_n} \right\} K\left(\frac{x - X_i}{h_n}\right), \end{aligned}$$

$$\begin{aligned}\hat{f}_{n,2}(x) &= \frac{1}{nh_n} \sum_{i=1}^n \left\{ \frac{x - X_i}{h_n} \right\}^2 K\left(\frac{x - X_i}{h_n}\right), \\ \hat{r}_n(x) &= \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right), \\ \hat{r}_{n,1}(x) &= \frac{1}{nh_n} \sum_{i=1}^n Y_i \left\{ \frac{x - X_i}{h_n} \right\} K\left(\frac{x - X_i}{h_n}\right).\end{aligned}$$

Les termes de centrage sont définis par :

$$\begin{aligned}f_n(x) &= \mathbb{E}[\hat{f}_n(x)], & r_n(x) &= \mathbb{E}[\hat{r}_n(x)], \\ f_{n,1}(x) &= \mathbb{E}[\hat{f}_{n,1}(x)], & r_{n,1}(x) &= \mathbb{E}[\hat{r}_{X;n,1}(x)] \\ f_{n,2}(x) &= \mathbb{E}[\hat{f}_{n,2}(x)].\end{aligned}$$

L'estimateur localement linéaire peut donc être écrit comme suit :

$$\hat{m}_n^{LL}(x) := \frac{\hat{r}_n(x)\hat{f}_{n,2}(x) - \hat{r}_{n,1}(x)\hat{f}_{n,1}(x)}{\hat{f}_n(x)\hat{f}_{n,2}(x) - \{\hat{f}_{n,1}(x)\}^2} \quad (2.56)$$

Nous considérons l'approximation suivante de l'espérance de $\hat{m}_n^{LL}(x)$:

$$\tilde{\mathbb{E}}[\hat{m}_n^{LL}(x)] = m_n^{LL}(x) := \frac{r_n(x)f_{n,2}(x) - r_{n,1}(x)f_{n,1}(x)}{f_n(x)f_{n,2}(x) - \{f_{n,1}(x)\}^2},$$

qui est équivalente asymptotiquement à $\mathbb{E}[\hat{m}_n^{LL}(x)]$.

Nous remarquons que, via le lemme de Bochner, sous (F.1–3) et (K.1–2) (cf. section 2.2),

$$\begin{aligned}f_{n,1}(x) &= \left\{ f_X(x) \int_{\mathbb{R}} uK(u)du \right\} (1 + o(1)) = o(1), \\ r_{n,1}(x) &= \left\{ r(x) \int_{\mathbb{R}} uK(u)du \right\} (1 + o(1)) = o(1), \\ f_{n,2}(x) &= \left\{ f_X(x) \int_{\mathbb{R}} u^2K(u)du \right\} (1 + o(1)).\end{aligned}$$

Ci-dessus, nous utilisons le fait que le moment d'ordre 1 du noyau K est toujours nul, ce qui explique intuitivement la formulation de l'estimateur [LL] (2.56). Plus précisément, les termes $\hat{r}_{n,1}$ et $\hat{f}_{n,1}$ convergent vers zéro et on retombe alors sur l'estimateur [NW].

Afin d'étudier le comportement limite de l'estimateur localement linéaire, nous introduisons le processus empirique suivant, pour tout $x \in J$ et $j = 0, 1, 2$,

$$\begin{aligned}W_{n,k,j}(x, \psi) &= \sum_{i=1}^n \left(c(x)\psi(Y_i) + d(x) \right) \left(\frac{x - X_i}{h_n} \right)^j K^{(k)}\left(\frac{x - X_i}{h_n}\right) \\ &\quad - n\mathbb{E}\left\{ \left(c(x)\psi(Y) + d(x) \right) \left(\frac{x - X}{h_n} \right)^j K^{(k)}\left(\frac{x - X}{h_n}\right) \right\}.\end{aligned} \quad (2.57)$$

En reprenant les hypothèses et notations de la section 2.2, nous obtenons le théorème suivant.

Théorème 2.6.1 *Supposons (F.1-3), (H.1-3) et (K.1-4). Nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ 2nh_n \log(1/h_n) \right\}^{-1/2} \sup_{x \in I} \{ \pm W_{n,k,j}(x, \psi) \} - \sigma_{W,j}(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_{W,j}^2(I) = \sup_{x \in I} \mathbb{E} \left[\left(c(x)\psi(Y) + d(x) \right)^2 \middle| X = x \right] f_X(x) \int_{\mathbf{R}} t^{2j} [K^{(k)}(t)]^2 dt.$$

Sous les hypothèses (F.1-3), (H.3-5) et (K.1-4), nous obtenons,

$$\left| \left\{ 2nh_n \log(1/h_n) \right\}^{-1/2} \sup_{x \in I} \{ \pm W_{n,k,j}(x, \psi) \} - \sigma_{W,j}(I) \right| \stackrel{p.s.}{=} o(1).$$

La démonstration est similaire à celle du théorème 2.3.1 et ne sera pas présentée par souci de concision. \square

Nous présentons quelques corollaires pour les cas particuliers $k = 0$ et $\psi = Id$, spécifiques à l'étude de l'estimateur localement linéaire.

Corollaire 2.6.1 *Supposons (F.1-3), (H.1-3) et (K.1-3). Nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_{n,j}(x) - f_{n,j}(x) \} - \sigma_{f,j}(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_{f,j}^2(I) = \sup_{x \in I} f_X(x) \int_{\mathbf{R}} [t^j K(t)]^2 dt.$$

Sous les hypothèses (F.1-3), (H.3-5) et (K.1-3), nous obtenons,

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_{n,j}(x) - f_{n,j}(x) \} - \sigma_{f,j}(I) \right| \stackrel{p.s.}{=} o(1).$$

Corollaire 2.6.2 *Supposons (F.1-3), (H.1-3) et (K.1-3). Nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_{n,j}(x) - r_{n,j}(x) \} - \sigma_{r,j}(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_{r,j}^2(I) = \sup_{x \in I} \sigma^2(x) f_X(x) \int_{\mathbf{R}} [t^j K(t)]^2 dt.$$

Sous les hypothèses (F.1-3), (H.3-5) et (K.1-3), nous obtenons,

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_{n,j}(x) - r_{n,j}(x) \} - \sigma_{r,j}(I) \right| \stackrel{p.s.}{=} o(1).$$

Il s'ensuit les assertions suivantes, sous les hypothèses du théorème 2.6.1,

$$\left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_n(x) - f_n(x) \} \stackrel{p.s.}{=} O(1),$$

$$\begin{aligned} & \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_{n,1}(x) - f_{n,1}(x) \} \stackrel{p.s.}{=} O(1), \\ & \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_{n,2}(x) - f_{n,2}(x) \} \stackrel{p.s.}{=} O(1), \\ & \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_n(x) - r_n(x) \} \stackrel{p.s.}{=} O(1), \\ & \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_{n,1}(x) - r_{n,1}(x) \} \stackrel{p.s.}{=} O(1). \end{aligned}$$

Notons que ces approximations sont aussi vraies pour la convergence en probabilité. Nous avons à présent tous les éléments essentiels pour la démonstration d'une loi uniforme du logarithme de l'estimateur [LL], c'est à dire nous avons établi des lois limites pour chacune des déviations de ses composantes (cf. (2.56) et les corollaires 2.6.1 et 2.6.2). Il nous reste à prouver que la déviation $\hat{m}_n^{LL}(x) - m_n^{LL}(x)$ est proche d'une certaine fonctionnelle linéaire du processus empirique. Soit

$$\epsilon_n := \sup_{x \in I} \left\{ \{ \hat{m}_n^{LL}(x) - m_n^{LL}(x) \} - \frac{1}{f_X(x)} \left\{ \{ \hat{r}_n(x) - r_n(x) \} - m(x) \{ \hat{f}_{X;n}(x) - f_{X;n}(x) \} \right\} \right\}.$$

Nous cherchons à démontrer que,

$$\epsilon_n \stackrel{p.s.}{=} o\left(\left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{-1/2} \right).$$

Nous décomposons l'erreur stochastique,

$$\begin{aligned} \hat{m}_n^{LL}(x) - m_n^{LL}(x) &= \frac{\{ \hat{r}_n(x) \hat{f}_{n,2}(x) - r_n(x) f_{n,2}(x) + r_{n,1}(x) f_{n,1}(x) - \hat{r}_{n,1}(x) \hat{f}_{n,1}(x) \}}{\hat{f}_{X;n}(x) \hat{f}_{n,2}(x) - \{ \hat{f}_{n,1}(x) \}^2} \\ &+ \frac{r_n(x) f_{n,2}(x) - r_{n,1}(x) f_{n,1}(x) \{ f_{X;n}(x) f_{n,2}(x) - \hat{f}_n(x) \hat{f}_{n,2}(x) + \{ \hat{f}_{n,1}(x) \}^2 - \{ f_{n,1}(x) \}^2 \}}{(\hat{f}_{X;n}(x) \hat{f}_{n,2}(x) - \{ \hat{f}_{n,1}(x) \}^2) (f_{X;n}(x) f_{n,2}(x) - \{ f_{n,1}(x) \}^2)}. \end{aligned}$$

Pour simplifier notre écriture, on supprime la dépendance en x momentanément,

$$\begin{aligned} \hat{m}_n^{LL} - m_n^{LL} &= \frac{\{ \hat{r}_n \hat{f}_{n,2} - r_n f_{n,2} + r_{n,1} f_{n,1} - \hat{r}_{n,1} \hat{f}_{n,1} \}}{\hat{f}_n \hat{f}_{n,2} - \{ \hat{f}_{n,1} \}^2} + \\ &\frac{r_n f_{n,2} - r_{n,1} f_{n,1} \{ f_n f_{n,2} - \hat{f}_n \hat{f}_{n,2} + \{ \hat{f}_{n,1} \}^2 - \{ f_{n,1} \}^2 \}}{(\hat{f}_n \hat{f}_{n,2} - \{ \hat{f}_{n,1} \}^2) (f_n f_{n,2} - \{ f_{n,1} \}^2)} \end{aligned}$$

Avec un peu de calculs, on obtient aisément, que la quantité ci-dessus est équivalente à,

$$\begin{aligned} &= \left\{ \frac{1}{\hat{f}_n \hat{f}_{n,2}} \times \left\{ \{ \hat{r}_n - r_n \} \hat{f}_{n,2} + r_n \{ \hat{f}_{n,2} - f_{n,2} \} + \{ r_{n,1} - \hat{r}_{n,1} \} f_{n,1} + \hat{r}_{n,1} \{ f_{n,1} - \hat{f}_{n,1} \} \right\} \right. \\ &+ \left. \frac{r_n f_{n,2}}{\hat{f}_n \hat{f}_{n,2} f_n f_{n,2}} \times \left\{ \{ f_n - \hat{f}_n \} \hat{f}_{n,2} + f_n \{ f_{n,2} - \hat{f}_{n,2} \} + \right. \right. \end{aligned}$$

$$\left. \left. \left. \left. \left. \hat{f}_{n,1} - f_{n,1} \right\} \left\{ \hat{f}_{n,1} + f_{n,1} \right\} \right\} \right\} (1 + o(1))$$

$$\begin{aligned} &\stackrel{p.s.}{=} \frac{\hat{r}_n - r_n}{\hat{f}_n} + \frac{f_n - \hat{f}_n}{\hat{f}_n f_n} \times r_n + o\left(\left\{\frac{nh_n}{2 \log(1/h_n)}\right\}^{-1/2}\right) \\ &\stackrel{p.s.}{=} \frac{1}{f} \times \{\hat{r}_n - r_n\} - \frac{m}{f} \times \{\hat{f}_n - f_n\} + o\left(\left\{\frac{nh_n}{2 \log(1/h_n)}\right\}^{-1/2}\right). \end{aligned}$$

D'après le lemme 2.4.8, nous obtenons en conséquence une loi limite uniforme du logarithme pour l'estimateur localement linéaire \hat{m}_n^{LL} .

Théorème 2.6.2 *Supposons (F.1-3), (H.1-3) et (K.1-3). Nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_n^{LL}(x) - m_n^{LL}(x) \} - \sigma_m(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_m^2(I) = \sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \int_{\mathbf{R}} K^2(t) dt.$$

Sous les hypothèses (F.1-3), (H.3-5) et (K.1-3), nous obtenons,

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_n^{LL}(x) - m_n^{LL}(x) \} - \sigma_m(I) \right| \stackrel{p.s.}{=} o(1).$$

Ce résultat s'appuie intuitivement sur l'équivalence asymptotique de la variance asymptotique des estimateurs [NW] et [LL]. En reprenant les arguments de cette section, nous pouvons généraliser la proposition ci-dessus au cas des estimateurs [PL](l) des dérivées de la régression d'ordre $k < l$, en s'appuyant sur (1.66) que l'on rappelle par convenance,

$$\text{Var} [\hat{m}_n^{(k)}(x; l) | \mathbb{X}] = \left\{ \frac{1}{nh^{1+2k}} \times \frac{\sigma^2(x)}{f_X(x)} \left\{ (k!)^2 \int_{\mathbf{R}} \{K_{k,l}^*(u)\}^2 du \right\} \right\} \{1 + o_{\mathbf{P}}(1)\},$$

Il s'ensuit le théorème suivant :

Théorème 2.6.3 *Supposons (F.1-3), (H.1-3) et (K.1-3). Nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_n^{(k)}(x; l) - m_n^{(k)}(x; l) \} - \sigma_{m,l}(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_{m,l}^2(I) = \sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \left\{ (k!)^2 \int_{\mathbf{R}} \{K_{k,l}^*(u)\}^2 du \right\}.$$

Sous les hypothèses (F.1-3), (H.3-5) et (K.1-3), nous obtenons,

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_n^{(k)}(x; l) - m_n^{(k)}(x; l) \} - \sigma_{m,l}(I) \right| \stackrel{p.s.}{=} o(1). \quad (2.58)$$

Ce théorème se généralise également au cadre multidimensionnel, en reprenant les arguments présentés dans les sections précédentes de ce chapitre.

2.7 Applications statistiques

2.7.1 Un critère simple de choix de fenêtre pour la convergence uniforme presque sûre

L'objet de ce paragraphe est la présentation d'une nouvelle procédure de sélection de la fenêtre, appropriée à la convergence uniforme presque sûre. En s'appuyant sur les travaux de Stute ([133] et [134]), il est possible de formuler une fonction de risque globale adaptée à la convergence uniforme presque sûre, similaire au risque quadratique ou MISE pour la convergence L_2 .

Le couple de variable aléatoire (X, Y) est supposé à valeurs dans $\mathbb{R} \times \mathbb{R}$ et nous nous concentrons sur la fenêtre optimale liée à l'estimation de la simple fonction de régression. Notons que notre approche du choix optimal de la fenêtre reste valide lorsque les variables aléatoires sont multivariées et également pour l'estimation des dérivées de la régression. Nous définissons la fonction de risque suivante, pour un estimateur $\hat{m}_n(x)$ de la régression, uniformément en $x \in I$, où $I \subset \mathbb{R}$ dénote un intervalle compact.

$$[\text{RPS}]\{\hat{m}_n(x)\} = \sup_{x \in I} \left\{ \hat{m}_n(x) - \mathbb{E}[\hat{m}_n(x)] \right\}^2 + \sup_{x \in I} \left\{ \mathbb{E}[\hat{m}_n(x)] - m(x) \right\}^2. \quad (2.59)$$

Exposons notre idée plus précisément et expliquons rapidement pourquoi cette définition du risque est censée. Nous considérons l'estimateur localement linéaire $\hat{m}_n^{LL}(x)$ qui constitue un estimateur performant de la régression. D'après le théorème 2.6.2, sous les hypothèses (F.1–3), (H.3–5) et (K.1–3), nous avons,

$$\lim_{n \rightarrow \infty} \left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} |\hat{m}_n^{LL}(x) - m_n^{LL}(x)| \stackrel{p.s.}{=} \left\{ \sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \int_{\mathbb{R}} K^2(t) dt \right\}^{1/2}. \quad (2.60)$$

A présent, nous supposons le noyau K d'ordre 2 et (F.5) vérifiée pour $k = 2$ (cf. le début de la section 2.2, ainsi que les paragraphes 1.3.2 et 1.7.2 consacrés au biais). Alors, le biais de l'estimateur localement linéaire est contrôlé par,

$$\lim_{n \rightarrow \infty} 2h_n^{-2} \sup_{x \in I} |m_n^{LL}(x) - m(x)| = \sup_{x \in I} |m''(x)| \int_{\mathbb{R}} t^2 K(t) dt. \quad (2.61)$$

Par la suite, lorsque la fenêtre h_n satisfait

$$\frac{nh_n^5}{\log(h_n^{-1})} \rightarrow 0, \quad \text{c'est à dire } h_n^2 = o\left(\left\{ \frac{\log(h_n^{-1})}{nh_n} \right\}^{1/2}\right),$$

le terme déterministe de biais est asymptotiquement négligeable. Il s'ensuit,

$$\lim_{n \rightarrow \infty} \left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} |\hat{m}_n^{LL}(x) - m(x)| \stackrel{p.s.}{=} \left\{ \sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \int_{\mathbb{R}} K^2(t) dt \right\}^{1/2}.$$

De l'autre côté, si h_n vérifie

$$\frac{nh_n^5}{\log(h_n^{-1})} \rightarrow \infty, \quad \text{c'est à dire } \left\{ \frac{\log(h_n^{-1})}{nh_n} \right\}^{1/2} = o(h_n^2),$$

nous obtenons,

$$\lim_{n \rightarrow \infty} 2\{h_n\}^{-2} \sup_{x \in I} |\hat{m}_n^{LL}(x) - m(x)| \stackrel{p.s.}{=} \sup_{x \in I} |m''(x)| [\mu_2(K)].$$

Ci-dessus, le terme stochastique est négligeable et la loi limite ne dépend plus de la variance asymptotique mais du biais asymptotique. Nous savons que la fenêtre optimale est obtenue en équilibrant le biais et la variance. En conséquence, il paraît raisonnable de déterminer la fenêtre optimale, en minimisant asymptotiquement le risque presque sûr défini en (2.59), c'est à dire

$$[\text{RPS}] \{ \hat{m}_n^{LL}(x) \} = \sup_{x \in I} \{ \hat{m}_n^{LL}(x) - m_n^{LL}(x) \}^2 + \sup_{x \in I} \{ m_n^{LL}(x) - m(x) \}^2.$$

D'après (2.60) et (2.61), nous cherchons à minimiser suivant h_n la quantité suivante,

$$\left\{ \left\{ \frac{2 \log(h_n^{-1})}{nh_n} \right\} \sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \mu_0(K^2) + \frac{h_n^4}{4} \sup_{x \in I} \{ m''(x) \}^2 \{ \mu_2(K) \}^2 \right\},$$

qui est presque sûrement égale au risque $[\text{RPS}] \{ \hat{m}_n^{LL}(x) \}$ asymptotiquement. Si on note $h_{n,opt}^{\text{RPS}}(K)$ la fenêtre optimale, il s'ensuit, après calculs,

$$h_{n,opt}^{\text{RPS}}(K) = h^{\text{RPS}}(K) = \left\{ \frac{\log n}{n} \right\}^{1/5} \left\{ \frac{\sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \mu_0(K^2)}{\sup_{x \in I} \{ m''(x) \}^2 [\mu_2(K)]^2} \right\}^{1/5} \quad (2.62)$$

L'étape suivante consiste à remplacer les termes inconnus par des estimateurs uniformément consistants. On obtient alors une fenêtre aléatoire de type plug-in, asymptotiquement optimale. Il serait intéressant de continuer ce travail, en démontrant par exemple la vitesse de convergence de cette fenêtre plug-in vers la fenêtre optimale théorique (2.62). Enfin, notons qu'il est possible de formuler un risque presque sûr local, en utilisant des résultats de convergence ponctuelle de type lois du logarithme itéré (cf. théorème 2.5.9, par exemple). La fenêtre optimale est alors de l'ordre $\{\log_2 n/n\}^{1/5}$, lorsque le noyau $K \in \mathcal{K}[2]$.

2.7.2 Fenêtre adaptative et intervalles de confiance

Cette sous-section propose une méthodologie, inspirée fortement par les travaux de Deheuvels et Mason (2004) [29], qui permet la construction d'intervalles de confiance uniformes et asymptotiquement optimaux pour différents paramètres fonctionnels de la distribution. Dans la littérature statistique classique, on utilise souvent la normalité asymptotique et les lois qui en découlent afin de construire des intervalles de confiance. Nous remarquons que nos lois limites uniformes du logarithme ainsi que les lois ponctuelles du logarithme itéré sont des extensions des lois asymptotiques normales à des modes de convergence plus forts (probabilité et presque sûre). Evidemment, du point de vue statistique, la convergence en probabilité est une notion suffisante. La convergence presque sûre, bien que plus

raffinée, nous oblige à supposer des conditions additionnelles sur la fenêtre. De plus, le mode de convergence presque sûre n'est pas naturel, par rapport à la définition classique des intervalles de confiance. En conséquence, les lois limites uniformes du logarithme que nous avons présentées dans ce chapitre seront utilisées pour le mode de convergence en probabilité, afin de déterminer des intervalles de confiance. Notons également que ce mode de convergence est particulièrement appropriée pour déterminer des bornes de confiance et nécessite des hypothèses peu restrictives sur la fenêtre.

Nous introduisons un estimateur consistant de la variance conditionnelle $\sigma_\psi^2(x)$, défini par,

$$\hat{\sigma}_{\psi;n}^2(x) = \frac{\sum_{i=1}^n \left\{ \psi(Y_i) - \hat{m}_{\psi;n}(x) \right\}^2 K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} \mathbb{I}\left\{ \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \neq 0 \right\}.$$

Sous les hypothèses du théorème 2.3.2, cet estimateur est bien consistant uniformément sur l'intervalle I . Nous avons, pour tout $\epsilon > 0$, lorsque $n \rightarrow \infty$,

$$\mathbb{P}\left\{ \sup_{x \in I} \left| \frac{\hat{\sigma}_{\psi;n}^2(x)}{\sigma_\psi^2(x)} - 1 \right| \geq \epsilon \right\} \rightarrow 0.$$

De la même façon, nous obtenons, pour tout $\epsilon > 0$, lorsque $n \rightarrow \infty$,

$$\mathbb{P}\left\{ \sup_{x \in I} \left| \frac{\hat{f}_{X;n}(x)}{f_X(x)} - 1 \right| \geq \epsilon \right\} \rightarrow 0.$$

Il s'ensuit le corollaire suivant, via Slutsky.

Corollaire 2.7.1 *Nous supposons les hypothèses (F.1-3), (H.1-3) et (K.1-3) vérifiées. Alors, nous obtenons, lorsque $n \rightarrow \infty$,*

$$\left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \left\{ \left\{ \frac{\hat{f}_{X;n}(x)}{\hat{\sigma}_{\psi;n}^2(x)} \right\}^{1/2} \pm \left\{ \hat{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\hat{m}_{n;\psi}^{(k)}(x)] \right\} \right\} \xrightarrow{\mathbb{P}} \int_{\mathbf{R}} [K^{(k)}(u)]^2 du.$$

La construction d'intervalles de confiance à partir de nos lois uniformes du logarithme implique implicitement la négligence du biais, car ces lois ne concernent que la déviation maximale par rapport à l'espérance. Afin de traiter le terme de biais, nous introduisons l'hypothèse suivante :

- (F.6) (i) f_X admet des dérivées continues jusqu'à l'ordre l sur l'intervalle J ;
 (ii) $f_{X,Y}$ est l -fois continûment différentiable sur $J \times \mathbb{R}$.

Lorsque la distribution du couple (X, Y) satisfait (F.6) et que le noyau K est d'ordre $l > k$ ou de manière équivalente si le noyau $K^{(k)}$ est d'ordre (k, l) , il s'ensuit

$$\sup_{x \in I} \left\{ \tilde{\mathbb{E}}[\hat{m}_{n;\psi}^{(k)}(x, c) - m_\psi^{(k)}(x)] \right\} = O(h_n^{l-k}),$$

Ainsi le biais est négligeable lorsque h_n est de l'ordre $n^{-\delta}$ avec $(2l+1)^{-1} \leq \delta \leq 1$. En conséquence, si nous supposons la fenêtre telle que $h_n = n^{-1/(2l+1)}$, sous les hypothèses (F.1–3), (F.6) et (K.1–3) (et K comme ci-dessus), nous obtenons, lorsque $n \rightarrow \infty$,

$$\left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \frac{\hat{f}_{X;n}(x)}{\hat{\sigma}_{\psi;n}^2(x)} \right\}^{1/2} \left\{ \hat{m}_{\psi;n}^{(k)}(x) - m_{\psi}^{(k)}(x) \right\} \xrightarrow{\mathbf{P}} \int_{\mathbf{R}} [K^{(k)}(u)]^2 du.$$

Cette version de la loi limite du logarithme nous permet de construire directement des intervalles de confiance pour la dérivée d'ordre k de la fonction de régression, uniformément en $x \in I$. Par convenance, nous posons,

$$L_n(x) := \left\{ \frac{2 \log(1/h_n)}{nh_n^{2k+1}} \times \frac{\hat{\sigma}_{\psi;n}^2(x)}{\hat{f}_{X;n}(x)} \right\}^{1/2} \left[\int_{\mathbf{R}} [K^{(k)}(u)]^2 du \right]^{1/2}.$$

Pour tout $0 < \epsilon < 1$, lorsque $n \rightarrow \infty$, il s'ensuit

$$\mathbb{P} \left\{ m_{\psi}^{(k)}(x) \in [\hat{m}_{\psi;n}^{(k)}(x) - (1+\epsilon)L_n(x), \hat{m}_{\psi;n}^{(k)}(x) + (1+\epsilon)L_n(x)], \forall x \in I \right\} \rightarrow 1,$$

et

$$\mathbb{P} \left\{ m_{\psi}^{(k)}(x) \in [\hat{m}_{\psi;n}^{(k)}(x) - (1-\epsilon)L_n(x), \hat{m}_{\psi;n}^{(k)}(x) + (1-\epsilon)L_n(x)], \forall x \in I \right\} \rightarrow 0.$$

En conséquence, nous dirons que les intervalles

$$\left[\hat{m}_{\psi;n}^{(k)}(x) - L_n(x), \hat{m}_{\psi;n}^{(k)}(x) + L_n(x) \right] \tag{2.63}$$

constituent des bornes de confiance asymptotiquement optimales (ou à un niveau de confiance asymptotique de 100 %) pour $m_{\psi}^{(k)}(x)$, uniformément en $x \in I$.

Il est possible également de déterminer des intervalles de confiance de la forme (2.63) lorsque la fenêtre est dépendante des données et donc aléatoire. Plus précisément, nous choisissons la fenêtre associée à nos estimateurs de la forme

$$H_n(x) = H_n(X_1, \dots, X_n; x) \quad \text{pour } n \geq 1.$$

Afin de contrôler le comportement limite des estimateurs à fenêtre aléatoire $H_n(x)$, nous admettrons qu'elle est assez proche de la fenêtre classique h_n dans un certain sens (cf. hypothèses (B.1–2) ci-dessous).

Pour chaque $n \geq 1$, nous supposons que la fenêtre adaptative $H_n(X_1, \dots, X_n; x)$ est une fonction mesurable des X_1, \dots, X_n et de $x \in I$. Comme ceci ne garantit pas la mesurabilité de

$$\inf_{x \in I} H_n(X_1, \dots, X_n; x) \quad \text{et} \quad \sup_{x \in I} H_n(X_1, \dots, X_n; x),$$

par rapport aux X_1, \dots, X_n , nous utiliserons la convention suivante. On désigne par $(\Omega, \mathcal{A}, \mathbb{P})$ l'espace de probabilité sur lequel nos variables aléatoires sont définies. Lorsque $\{A_n : n \geq 1\}$ sont des sous-ensembles (éventuellement non-mesurables) de Ω , nous écrivons $\mathbb{P}(A_n) \rightarrow 1$ (ou bien $\mathbb{P}(\bar{A}_n) \rightarrow 0$, avec $\bar{A} := \Omega - A$ le complémentaire de A), lorsqu'il

existe une suite $\{A_n : n \geq 1\} \subseteq \mathcal{A}$, telle que $\bar{A}_n \subseteq B_n$ pour chaque $n \geq 1$, et $\mathbb{P}(B_n) \rightarrow 0$. Cette convention est proche de la notion de \mathbb{P} -mesurabilité définie en annexe et rejoint celle de complétion d'un espace.

Avec ces conventions, nous supposons que $H_n(x) =: h_n C_n(x)$, $x \in I$, vérifie certaines hypothèses parmi (B.1) – (B.2) ci-dessous. Soient $0 < c_1 \leq c_2 < \infty$ deux constantes et soit $\{C(x) : x \in I\}$ une fonction positive fixée, continue et différente de 0 sur l'intervalle I .

$$(B.1) \quad \mathbb{P}\left(c_1 h_n \leq \inf_{x \in I} H_n(x) \leq \sup_{x \in I} H_n(x) \leq c_2 h_n\right) \rightarrow 1, \text{ quand } n \rightarrow \infty;$$

$$(B.2) \quad \mathbb{P}\left(\sup_{x \in I} \left| \frac{H_n(x)}{h_n} - C(x) \right| \geq \epsilon\right) \rightarrow 0, \text{ lorsque } n \rightarrow \infty, \text{ pour } \epsilon > 0.$$

Lorsque (B.1) ou (B.2) est vérifiée, nous pouvons alors définir des intervalles de confiance asymptotiquement optimaux en remplaçant dans (2.63) h_n par $H_n(x)$. Pour cela, il faut étendre le théorème 2.3.2 au cadre suivant. Soit $c > 0$ un nombre réel fixé. Nous désignons par $\hat{m}_{n,\psi}^{(k)}(x, c)$ l'estimateur [NW] de la dérivée k -ième de la régression construit avec une fenêtre de taille ch_n , c'est à dire

$$\hat{m}_{\psi;n}^{(k)}(x, c) = \sum_{j=1}^k \left\{ \frac{\hat{r}_{\psi;n}^{(j)}(x, c)}{\hat{f}_{X;n}^{(k-j)}(x, c)} \right\},$$

lorsque

$$\hat{f}_{X;n}(x, c) = \frac{1}{nch_n} \sum_{i=1}^n K\left(\frac{x - X_i}{ch_n}\right) \quad \text{et} \quad \hat{r}_{\psi;n}(x, c) = \frac{1}{nch_n} \sum_{i=1}^n \psi(Y_i) K\left(\frac{x - X_i}{ch_n}\right).$$

En s'appuyant sur l'article de Deheuvels et Mason [29], l'idée est de faire varier la constante $c > 0$ dans un certain intervalle qui délimitera alors la zone de variation de la fenêtre h_n . Nous présentons donc une extension du théorème 2.3.2, dans le cadre où la fenêtre h_n n'est plus strictement fixée.

Théorème 2.7.1 *Nous supposons les hypothèses (F.1–3), (H.1–3), (K.1–3) vérifiées et nous fixons $0 < c_1 \leq 1 \leq c_2 < \infty$. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\sup_{c_1 \leq c \leq c_2} \left| \left\{ \frac{n(ch_n)^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{m}_{\psi;n}^{(k)}(x, c) - \tilde{\mathbb{E}}[\hat{m}_{n,\psi}^{(k)}(x, c)] \right\} - \sigma_m(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_m^2(I) = \sup_{x \in I} \left\{ \frac{\sigma_{\psi}^2(x)}{f_X(x)} \right\} \int_{\mathbf{R}} [K^{(k)}(u)]^2 du.$$

La démonstration de ce théorème reprend les arguments principaux de la démonstration du théorème 2.3.2 mais en ajoutant une dimension supplémentaire à la discrétisation. Nous nous référons également à la démonstration du théorème 3.1 de [29]. \square

Chapitre 3

Maximum de vraisemblance local et régression non-paramétrique

Ce travail se situe dans le cadre de l'estimation d'un paramètre fonctionnel lié à une fonction de régression. Nous considérons des estimateurs à noyaux d'une fonction de régression particulière, fondés sur un maximum de vraisemblance pondérée. En s'appuyant sur les résultats du précédent chapitre (en particulier la section 2.2), nous obtenons des lois exactes concernant la convergence uniforme presque sûre de ces estimateurs. Ces lois limites permettent la construction de bornes de confiance uniformes et asymptotiquement optimales pour certains paramètres de la distribution dans un cadre semi-paramétrique. Ce dernier chapitre montre une fois de plus la puissance du formalisme de la théorie des processus empiriques. A cet effet, on cite en avant-propos le fameux livre de Sara van de Geer (2000) [144] qui présente de nombreuses applications en statistique non-paramétrique, en arguant de la théorie récente sur les processus empiriques indexés par des classes de fonctions. Ses investigations concernent les propriétés asymptotiques des M -estimateurs, et plus particulièrement les estimateurs du maximum de vraisemblance et des moindres carrés. Les ouvrages de référence des auteurs Van der Vaart et Wellner (1996) (§.3, [145]) ainsi que Van der Vaart (1998) [146] illustrent également l'utilité de la théorie moderne des processus empiriques, ou théorie des processus empiriques indexés par des classes de fonctions, pour démontrer des résultats de nature statistique. Ces livres présentent des applications variées dans de nombreux domaines de la statistique, notamment en estimation semi-paramétrique et M -estimation.

3.1 Introduction

L'estimation du maximum de vraisemblance local est un sujet relativement peu abordé dans la littérature statistique. Le cadre de notre travail sur l'estimation du maximum de vraisemblance local étant très proche de celui de la M -estimation, il nous paraît intéressant de rappeler certaines notions clés de la M -estimation, afin de recentrer notre propos. Ces notions rejoignent également l'estimation non-paramétrique de la régression et ouvrent de larges perspectives d'études.

La méthode la plus importante de construction d'estimateurs statistiques consiste à choisir

un estimateur minimisant ou maximisant un certain critère fonctionnel. De tels estimateurs sont appelés **M-estimateurs** dans la littérature statistique. Premièrement, notons que la dénomination de M -estimation vient du simple fait que l'on cherche à minimiser ou maximiser une certaine fonctionnelle, c'est à dire M correspond à un minimum ou un maximum. Dans de nombreuses situations, ces estimateurs, qui maximisent ou minimisent une certaine application, sont aussi solutions d'un système d'équations. Par exemple, dans ce chapitre, nous considérons les propriétés asymptotiques d'une certaine statistique, l'estimateur du maximum de vraisemblance local, solution d'une équation (cf. (3.8) ci-après). Dans le cadre d'observations i.i.d., les M -estimateurs sont donc simplement les zéros d'application du type :

$$\theta \rightarrow \mathbb{P}_n \psi_\theta. \quad (3.1)$$

Ce type d'estimateurs sont aussi appelés Z -estimateurs (cf. section 3.3 dans [146]). Ils sont définis par une équation de la forme :

$$\psi_n(\hat{\theta}_n) = 0,$$

où ψ_n désigne une application aléatoire définie sur l'espace produit de l'espace des paramètres Θ et un certain espace de probabilité. La forme des M ou Z -estimateurs, présentée ci-dessus en (3.1), nous incite à orienter la recherche des propriétés générales de ces estimateurs via la théorie des processus empiriques indexés par des classes de fonctions.

-Présentation générale de la M -estimation

Pour toute fonction $\psi(x, t)$, nous pouvons associer une fonctionnelle T_ψ , définie sur sur les fonctions de répartition F , telle que $T_\psi(F)$ soit la solution t_0 de l'équation suivante :

$$\int \psi(x, t_0) dF(x) = 0. \quad (3.2)$$

Nous appelons $T_\psi(\cdot)$ la **M -fonctionnelle correspondant à ψ** . En suivant (3.1), à partir d'un n échantillon de variables aléatoires $\{X_i : 1 \leq i \leq n\}$, le **M -estimateur correspondant à ψ** est donc la statistique $T_\psi(\mathbb{P}_n) = T_n$ solution de l'équation :

$$\sum_{i=1}^n \psi(X_i, T_n) = 0. \quad (3.3)$$

Notons que les équations (3.2) et (3.3) peuvent admettre plusieurs solutions.

Dans le cadre standard, l'équation (3.2) correspond à la réalisation d'une condition de premier ordre liée à la minimisation ou maximisation d'une certaine fonctionnelle

$$\int \rho(x, t_0) dF(x).$$

La fonction ψ peut donc être regardée comme la dérivée d'une certaine fonction $\rho(x, \cdot)$ dérivable, telle que

$$\psi(x, t) = c \frac{\partial}{\partial t} \{ \rho(x, t) \},$$

où c désigne une constante arbitraire.

Exemple 3.1.1 Estimation du maximum de vraisemblance

Soit $\mathcal{F} = \{F(\cdot; \theta) : \theta \in \Theta\}$ une famille paramétrique de distributions. Soit $\psi = \psi(x, t)$ une fonction telle que

$$\int \psi(x, \theta) dF(x; \theta) = 0.$$

Cet exemple est bien une M -fonctionnelle au sens de (3.2) pour $F = F(\cdot; \theta)$ et la solution de (3.2) coïncide alors avec θ . En d'autres termes, la M -fonctionnelle T_ψ satisfait $T_\psi(F(\cdot; \theta)) = \theta$. L'estimateur naturel de θ est donné par $\hat{\theta}_n = T_\psi(\mathbb{P}_n)$ d'après (3.3). Suivant le choix de ψ , nous obtenons différents estimateurs. Lorsque les distributions $F(\cdot; \theta)$ admettent des densités $f(\cdot; \theta)$, l'**estimateur du maximum de vraisemblance** correspond aux choix :

$$\begin{aligned} \rho(x, \theta) &= -\log f(x; \theta), \\ \psi(x, \theta) &= -\frac{\partial}{\partial \theta} \log f(x; \theta). \end{aligned}$$

- M -estimation et localisation de paramètre

Un autre cas particulièrement intéressant de la M -estimation, qui englobe le problème de l'estimation non-paramétrique de la régression, est le suivant. Supposons que la fonction ψ est de la forme $\psi(x, t) = \psi(x - t)$, alors la M -fonctionnelle associée $T_\psi(F)$ est appelée **paramètre de localisation** (ou "*location parameter*"). La M -fonctionnelle $T_\psi(F)$ est alors la solution θ de l'équation :

$$\int \psi(x - \theta) dF(x) = 0.$$

A présent, si la fonction de répartition F est symétrique par rapport à θ (ici, θ désigne un paramètre informel), tout choix de fonction $\psi(\cdot)$ antisymétrique nous donne clairement $T_\psi(F) = \theta$. Ainsi, lorsque nous disposons d'une classe de fonctions $\psi(\cdot)$ antisymétriques, en remplaçant la mesure F par la mesure empirique \mathbb{P}_n , nous obtenons une classe d'estimateurs de θ . Pour un choix de ψ convenable, cet estimateur possède des propriétés de robustesse, ou résistance aux données aberrantes ("*outliers*"). En effet, une propriété fondamentale de la M -estimation est sa capacité à transférer au M -estimateur les propriétés intrinsèques de la fonction ψ qui lui est associée. Comme nous allons le voir ci-dessous, le choix de la fonction ψ nous conduit également à des estimateurs variés et robustes dans le cadre de l'estimation d'une fonction de régression. Un exemple célèbre d'estimation robuste est donné par les estimateurs de Huber (cf. [77] et [78]), solutions de

$$\sum_{i=1}^n \psi(X_i - \theta) = 0,$$

lorsque $\psi(\cdot)$ est de la forme suivante

$$\psi(x) = [x]_{-k}^k := \begin{cases} -k & \text{si } x \leq -k, \\ x & \text{si } |x| \leq k, \\ k & \text{si } x \geq k. \end{cases}$$

Ces estimateurs sont robustes et, suivant la valeur de k , se comportent comme la moyenne classique (k grand) ou comme la médiane (k petit). Ainsi, les estimateurs de Huber vont de la moyenne non-robuste à la médiane très robuste. Ce point de vue a été repris par Arcones dans un article récent [5] concernant la convergence du M -estimateur optimal parmi une famille paramétrique de M -estimateurs. Nous nous référons aux pages 246-248 de Serfling (1980) pour une présentation d'autres exemples liés au modèle de localisation de paramètre ou "*location parameter estimation*" et à l'exemple 5.4, p. 42-44, de van der Vaart (1998) [146].

-Régression et M -estimation

Soient (X, Y) , (X_1, Y_1) , $(X_2, Y_2), \dots$, des couples de variables aléatoires à valeurs réelles, indépendants et identiquement distribués. Soit $F = F_{(X,Y)}$ la fonction de répartition jointe associée au couple générique (X, Y) . Le problème de la régression non-paramétrique consiste à estimer la courbe de régression de Y sachant X . Il nous faut donc déterminer $m(x) = m_{\psi, F}(x)$ à partir des données $\{(X_i, Y_i) : 1 \leq i \leq n\}$. Ici, le paramètre fonctionnel ψ est lié à la forme de la courbe de régression, voir (3.4) ci-dessous. Suivant le choix de ψ , nous obtenons comme fonction de régression, par exemple, la moyenne conditionnelle ou la médiane conditionnelle. Plus précisément, la fonction $m_{\psi, F}$ vérifie

$$\mathbb{E}[\psi(Y - m(x)) | X = x] = 0. \quad (3.4)$$

Si $\psi(u) = u$ dans (3.4), il s'ensuit comme définition de la courbe de régression $m(x) = \mathbb{E}[Y | X = x]$, c'est à dire la moyenne conditionnelle $\mathbb{E}[Y | X = x]$ minimise la perte L_2 par rapport à la distribution conditionnelle de $Y | X = x$. Pour le choix de $\psi(u) = 1/2 - \mathbb{I}\{u \leq 0\}$ ou $\psi(u) = \text{signu}$, nous obtenons

$$m(x) = \text{med}[Y | X = x], \quad \text{la médiane conditionnelle.}$$

La médiane conditionnelle minimise la perte L_1 par rapport à la distribution conditionnelle de $Y | X = x$. Notons que l'estimateur de la médiane conditionnelle (cf. [141]) est plus robuste que l'estimateur [NW]. Par exemple, en présence de données aberrantes ou extrêmes, il est plus approprié d'utiliser comme estimateur de la régression la moyenne conditionnelle. Suivant le poids de la queue de distribution ou le type de distribution, l'estimateur de la moyenne ou de la médiane conditionnelle ont des performances bien distinctes. Dans le cadre où la distribution est dans le domaine d'attraction gaussien, il est préférable d'utiliser la moyenne mais lorsque la distribution est de type exponentielle, la médiane s'avère plus efficace. En reprenant le point de vue d'Arcones [5], il serait intéressant de proposer un estimateur de la régression adaptatif dans ce sens, suivant un critère d'erreur spécifique tel l'erreur moyenne quadratique intégrée.

En conséquence, nous pouvons estimer, de manière générale, la fonction de régression $m(x)$ par l'estimateur $m_n(x)$, solution (par rapport à θ) de

$$\sum_{i=1}^n W_{ni}(x) \psi(Y_i - \theta) = 0,$$

avec $W_{ni}(\cdot)$ fonction de poids arbitraire. Nous remarquons immédiatement que lorsque

$$\psi(u) = u \quad \text{et} \quad W_{ni}(x) = K\left(\frac{x - X_i}{h}\right),$$

nous retombons sur l'estimateur [NW] classique.

On se réfère aux travaux de Härdle (1984) [65], Härdle et Luckaus (1984) [72], Härdle, Janssen et Serfling (1988) [68], Härdle et Tsybakov (1988) [74], ainsi que Truong (1989) [141] pour une exposition des différentes propriétés de ces estimateurs robustes de la régression et un approfondissement des notions présentées ci-dessus.

3.2 Hypothèses de travail

Soient (X, Y) , (X_1, Y_1) , $(X_2, Y_2), \dots$, des couples de variables aléatoires à valeurs réelles, indépendants et identiquement distribués. Le couple (X, Y) est supposé admettre une densité jointe sur \mathbb{R}^2 notée $f_{X,Y}$ et nous désignons toujours par f_X la densité marginale de X . Soit Θ désignant une collection de fonctions $\theta : \mathbb{R} \rightarrow \mathbb{R}$. Pour $(x, y) \in \mathbb{R}^2$, la loi conditionnelle de Y sachant $X = x$ est définie par sa densité, notée $f_{Y|X}(y, x)$, et supposée de la forme :

$$f_{Y|X}(y, x) := g(y; \theta(x)), \tag{3.5}$$

où $g(\cdot; \cdot)$ est une fonction supposée de forme connue. Cette hypothèse est fondamentale pour l'exposition de nos travaux et souligne le caractère semi-paramétrique de notre étude.

Exemple 3.2.1 Lorsque la loi conditionnelle de Y sachant $X = x$ est une Exponentielle de paramètre fonctionnel $\theta(x)$ inconnu, alors g est de la forme :

$$g(y; \theta(x)) = \frac{1}{\theta(x)} e^{-y/\theta(x)}, \quad y > 0 \quad \text{et} \quad \theta(x) > 0.$$

Nous avons alors $\mathbb{E}[Y|X = x] = \theta(x)$, donc l'estimation du paramètre fonctionnel $\theta(x)$ est ici équivalent à l'estimation de la courbe de régression classique. De même, lorsque

$$g(y; \theta(x)) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y - \theta(x))^2\right\}.$$

Le paramètre $\theta \in \Theta$ désigne donc une fonction à valeurs réelles, supposée deux fois continûment différentiable (voir (F.5) ci-après). Dans ce chapitre, nous travaillons avec $x \in J$ intervalle compact de \mathbb{R} et nous posons $T := \{\theta(x) : \theta \in \Theta, x \in J\}$, qui constitue un intervalle compact de \mathbb{R} .

Les résultats seront établis uniformément sur un intervalle compact $I \subset J$, comme dans le précédent chapitre.

Rappelons que la fonction $\theta(x)$ vérifie par définition, en tant que paramètre de la distribution conditionnelle,

$$\theta(x) = \arg \max_{t \in T} \{\mathbb{E}_x[\log g(Y; t)]\}, \quad \text{ou bien} \quad \mathbb{E}_x[\psi(Y; \theta(x))] = 0, \tag{3.6}$$

avec $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot | X = x]$ et $\psi(y; t) = \frac{\partial}{\partial t} \log g(y; t)$.

Remarque 3.2.1 L'équation (3.6) est équivalente à

$$\int_{\mathbf{R}} \psi(y; \theta(x)) f_{Y|X}(y|x) dy = 0.$$

D'après la formulation en (3.2), nous sommes exactement dans le contexte de la Z -estimation ou M -estimation.

Nous utiliserons souvent la variance conditionnelle de la variable $\psi(Y; \theta(x))$, notée

$$I_{\theta}(x) := \mathbb{E}_x \left[- \left\{ \frac{\partial \psi}{\partial t}(Y; t) \right\} \right]_{t=\theta(x)} = \mathbb{E}_x \left[\left\{ \psi(Y; t) \right\}^2 \right]_{t=\theta(x)}, \quad (3.7)$$

qui désigne également l'Information de Fisher locale.

Nous imposons certaines conditions sur la distribution du couple (X, Y) , parmi les hypothèses (F.1–6), présentées ci-dessous.

(F.1) $f_X(\cdot)$ et $I_{\theta}(\cdot)$ sont continues et strictement positives sur J ;

(F.2) $Y \mathbb{I}\{X \in J\}$ est bornée.

(F.3) Les dérivées partielles d'ordre 1, 2, 3 (par rapport à t) de $\log g(y, t)$ existent et sont continues sur $\mathbf{R} \times T$.

(F.4) Il existe des fonctions $H_i(y)$ intégrables telles que

$$\left| \frac{\partial^i \log g(y; t)}{\partial t^i} \right| \leq H_i(y), \quad \text{pour } i = 1, 2.$$

(F.5) Les dérivées $f'(x)$, $I'(x)$, $\theta'(x)$ et $\theta''(x)$ sont continues et bornées.

(F.6) Il existe des constantes positives C_1 et C_2 telles que

$$\inf_{x \in I} \mathbb{E}_x \left[- \psi'(Y; \theta(x) + \epsilon) \right] > C_2 > 0, \quad \text{lorsque } |\epsilon| \leq C_1.$$

L'hypothèse (F.1) est fondamentale car si la densité marginale f_X ou l'information de Fisher locale $I_{\theta}(x)$ sont nulles, le paramètre fonctionnel $\theta(x)$ ne peut pas être estimé. Par contre, nous remarquons que la continuité et la bornitude de $I_{\theta}(x)$ sont impliquées par (F.2–3–4). L'hypothèse (F.2) est classique en régression non-paramétrique. Notons qu'il est possible également de supposer un moment d'ordre $s > 2$, en utilisant un argument de troncation combiné à une hypothèse supplémentaire sur la fenêtre. Les hypothèses (F.3–4) sont des extensions naturelles de conditions nécessaires à la théorie du maximum de vraisemblance pour obtenir les propriétés usuelles de consistance et normalité asymptotique (cf. § 4.2.2, p. 144-149, [123]) dans les modèles paramétriques. Plus précisément, la condition (F.3) nous assure que la fonction de score $\psi(y; t)$ admet un développement de Taylor, comme fonction de t . L'hypothèse (F.4) permet la différentiation par rapport à t sous le signe intégrale et justifie les formules (3.6) et (3.7).

$$\mathbb{E}_x [\psi(y; \theta(x))] = \int_{\mathbf{R}} \frac{1}{g(y; t)} \frac{\partial g(y; t)}{\partial t} g(y; t) dy$$

$$= \int_{\mathbf{R}} \frac{\partial g(y; t)}{\partial t} dy = \frac{\partial}{\partial t} \int_{\mathbf{R}} g(y; t) dy = \frac{\partial}{\partial t}(1) = 0.$$

Nous obtenons également

$$\begin{aligned} \mathbb{E}_x \left[\left\{ \frac{\partial \psi}{\partial t}(Y; t) \right\} \right] &= \int_{\mathbf{R}} \left\{ \frac{1}{g(y; t)} \frac{\partial^2 g(y; t)}{\partial t^2} - \left(\frac{1}{g(y; t)} \frac{\partial g(y; t)}{\partial t} \right)^2 \right\} g(y; t) dy \\ &= \int_{\mathbf{R}} \frac{\partial^2 g(y; t)}{\partial t^2} dy - \int_{\mathbf{R}} \left(\frac{1}{g(y; t)} \frac{\partial g(y; t)}{\partial t} \right)^2 g(y; t) dy \\ &= 0 - \mathbb{E}_x \left[\psi(y; t)^2 \right] = -\mathbb{E}_x \left[\psi(y; t)^2 \right]. \end{aligned}$$

L'hypothèse (F.5) est utile pour contrôler le biais et la condition (F.6) est raisonnable d'après les définitions en (3.6).

L'estimation de $\theta(x)$ est fondée sur la maximisation d'une vraisemblance locale (cf. (3.8) ci-dessous). Cette technique d'estimation dénommée "*local likelihood estimation*" a pour origine une idée développée par Tibshirani et Hastie (1987) [140]. La normalité asymptotique a été discutée par Staniswalis (1989) [128] dans le cadre du plan fixe, c'est à dire lorsque les $X_i = x_i$ sont déterministes. En s'appuyant sur la méthodologie de Härdle, Jansen et Serfling (1988) [68], Zhao (1994) [151] a démontré la convergence uniforme avec vitesse optimale de l'estimateur $\hat{\theta}_{n,h}(x)$, défini ci-dessous.

Théorème 3.2.1 Zhao (1994) *Sous les hypothèses du théorème 2.2, p. 82, [151].*

$$\sup_{x \in I} \left| \hat{\theta}_{n,h}(x) - \theta(x) \right| \stackrel{p.s.}{=} O \left(\left\{ \frac{\log n}{nh} \right\}^{1/2} + h^2 \right).$$

D'après (3.6), l'estimateur du maximum de vraisemblance local $\hat{\theta}_{n,h}(x)$ est solution (par rapport à t) de l'équation suivante

$$\hat{r}_{n,h}(x, t) = \frac{1}{nh} \sum_{i=1}^n \psi(Y_i; t) K \left(\frac{x - X_i}{h} \right) = 0, \quad (\text{condition du premier ordre}) \quad (3.8)$$

où $K(\cdot)$ est un noyau et h désigne la fenêtre ou paramètre de lissage (par la suite, nous supprimerons la dépendance en h). Le noyau $K(\cdot)$ est supposé satisfaire :

(K.1) K est continue et à variation bornée sur \mathbb{R} ;

(K.2) K est à support compact ;

(K.3) K noyau d'ordre 2.

Pour être plus précis, la fonction K est seulement continue par morceaux mais continue sur son support compact. D'une manière générale, une grande majorité des noyaux sont des fonctions polynomiales par morceaux. Notons enfin que l'hypothèse (K.3) est assez arbitraire, elle nous servira notamment à expliciter le biais asymptotique. Il est possible de choisir un noyau avec des conditions de régularité différentes mais pour le praticien un noyau d'ordre 2 est souvent suffisant.

Afin d'éviter les valeurs négatives du logarithme, nous considérons la notation suivante $\log u = \log(u \vee e)$. Nous travaillerons avec une fenêtre $h_n > 0$, indexée par $n = 1, 2, \dots$, vérifiant certaines conditions parmi les hypothèses (H.1–5) ci-dessous.

- (H.1) $h_n \rightarrow 0$, lorsque $n \rightarrow \infty$;
- (H.2) $nh_n / \log n \rightarrow \infty$, lorsque $n \rightarrow \infty$;
- (H.3) $\log(1/h_n) / nh_n^5 \rightarrow \infty$, lorsque $n \rightarrow \infty$;
- (H.4) $h_n \searrow 0$ et $nh_n \nearrow \infty$, lorsque $n \rightarrow \infty$;
- (H.5) $\log(1/h_n) / \log \log n \rightarrow \infty$, lorsque $n \rightarrow \infty$.

Remarque 3.2.2 Sous certaines conditions de régularités, l'hypothèse (H.3) permet de négliger asymptotiquement le terme de biais qui est de l'ordre $O(h_n^2)$, via la condition de moment (K.3). Cette condition est nécessaire pour l'obtention d'une loi limite concernant la déviation maximale $\sup_{x \in I} \{\hat{\theta}_{n,h_n}(x) - \theta(x)\}$. Si nous choisissons le noyau K d'ordre $q > 2$, l'hypothèse (H.3) devient

$$\log(1/h_n) / nh_n^{2q+1} \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty.$$

Dans la prochaine section, nous montrerons une nouvelle loi limite uniforme pour la convergence en probabilité et presque sûre de la déviation $\{\hat{\theta}_n(x) - \theta(x)\}$. Cette loi limite uniforme du logarithme, présenté dans le théorème 3.3.1 ci-dessous, permet la construction d'intervalles de confiance asymptotiquement optimaux pour le paramètre fonctionnel $\theta(x)$ et raffine les précédents résultats de Zhao (1994) [151]. On note toutefois que la construction de ces intervalles de confiance nous amène à négliger le biais ou terme déterministe (cf. remarque 3.2.2), afin d'obtenir une loi uniforme exacte. La section 3.4 présente une extension du théorème 3.3.1 au cadre multidimensionnel, avec quelques applications statistiques intéressantes. La section 3.5 est consacrée à la démonstration de nos résultats.

3.3 Résultats

L'estimateur du maximum de vraisemblance local est solution de l'équation (3.8). Nos hypothèses garantissent l'existence d'une racine à cette équation. Afin de se prémunir d'éventuelles solutions multiples, nous supposons que la suite de solutions $\hat{\theta}_n(x)$ satisfait

$$\sup_{x \in I} \left| \hat{\theta}_n(x) - \theta(x) \right| < \alpha \quad \text{presque sûrement,}$$

où α désigne une constante suffisamment petite. Ainsi, toutes les racines de l'équation (3.8) sont proches les unes des autres, voire égales. Pour plus de détails concernant cet argument, nous renvoyons à Zhao, p. 83, [151].

Théorème 3.3.1 *Supposons que les hypothèses (F.1–6), (H.1–3) et (K.1–3) soient vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{\theta}_n(x) - \theta(x) \} - \sigma_\theta(I) \right| = o_{\mathbf{P}}(1), \quad (3.9)$$

où

$$\sigma_\theta(I) = \sup_{x \in I} \left\{ \frac{1}{f_X(x)I_\theta(x)} \int_{\mathbf{R}} [K^2(u)] du \right\}^{1/2} =: \sup_{x \in I} \{V_\theta(x)\}^{1/2}.$$

Si la fenêtre satisfait (H.2-5) nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{\hat{\theta}_n(x) - \theta(x)\} - \sigma_\theta(I) \right| = o(1) \quad \text{presque sûrement.} \quad (3.10)$$

Construction d'intervalles de confiance

Par convenance, nous introduisons l'estimateur à noyau $V_\theta(x)$, lorsque $K \geq 0$,

$$\hat{V}_n(x) = \left\{ \frac{1}{nh_n} \sum_{i=1}^n \psi^2(Y_i; t) K\left(\frac{x - X_i}{h_n}\right) \right\}^{-1} \times \int_{\mathbf{R}} [K^2(u)] du.$$

Sous les hypothèses du théorème 3.3.1, cet estimateur est uniformément consistant sur I , c'est à dire

$$\sup_{x \in I} \left| \frac{\hat{V}_n(x)}{V_\theta(x)} - 1 \right| \xrightarrow{\mathbf{P}} 0, \quad \text{ou} \quad \sup_{x \in I} \left| \frac{\hat{V}_n(x)}{V_\theta(x)} - 1 \right| \xrightarrow{p.s.} 0.$$

En s'appuyant sur un argument du type Slutsky, nous obtenons le corollaire suivant.

Corollaire 3.3.1 *Sous les hypothèses (F.1-6), (H.1-3) et (K.1-3), nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \{\hat{V}_n(x)\}^{-1/2} \times \{\hat{\theta}_n(x) - \theta(x)\} \right\} - 1 \right| = o_{\mathbf{P}}(1). \quad (3.11)$$

Si la fenêtre satisfait (H.2-5) nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \{\hat{V}_n(x)\}^{-1/2} \times \{\hat{\theta}_n(x) - \theta(x)\} \right\} - 1 \right| \stackrel{p.s.}{=} o(1). \quad (3.12)$$

Soit

$$\tilde{R}_{n,h_n}(x) = \left\{ \left\{ \frac{2 \log(1/h_n)}{nh_n} \right\} \times \left\{ \hat{V}_n(x) \right\} \right\}^{1/2}$$

D'après (3.11), il est possible de construire des intervalles de confiance pour $\theta(x)$, uniformément en $x \in I$. Précisément, nous avons, pour chaque $\varepsilon > 0$, lorsque $n \rightarrow \infty$,

$$\mathbb{P} \left\{ \theta(x) \in [\hat{\theta}_n(x) - (1 + \varepsilon)\tilde{R}_{n,h_n}(x), \hat{\theta}_n(x) + (1 + \varepsilon)\tilde{R}_{n,h_n}(x)], \forall x \in I \right\} \rightarrow 1,$$

et

$$\mathbb{P} \left\{ \theta(x) \in [\hat{\theta}_n(x) - (1 - \varepsilon)\tilde{R}_{n,h_n}(x), \hat{\theta}_n(x) + (1 - \varepsilon)\tilde{R}_{n,h_n}(x)], \forall x \in I \right\} \rightarrow 0.$$

Ainsi les intervalles $[\hat{\theta}_n(x) - \tilde{R}_{n,h_n}(x), \hat{\theta}_n(x) + \tilde{R}_{n,h_n}(x)]$ constituent des intervalles de confiance pour la fonction $\theta(x)$ ($x \in I$), à un niveau de confiance asymptotique de 100%.

3.4 Extension multidimensionnelle

Dans de nombreux cas d'estimation par la méthode du maximum de vraisemblance, le paramètre $\theta(x)$ est à valeurs dans \mathbb{R}^p , avec $p > 1$. Nous considérons alors la déviation unidimensionnelle $\sup_{x \in I} \pm \mathbf{u}^T \{\hat{\theta}_n(x) - \theta(x)\}$ où $\mathbf{u} \in \mathbb{R}^p$ est fixé et \mathbf{u}^T dénote son transposé.

Afin de proposer une extension multidimensionnelle du théorème 3.3.1, nous introduisons quelques notations. Soit la matrice $\mathbf{I}_\theta(x)$, supposée définie positive,

$$\mathbf{I}_\theta(x) = \left[\mathbb{E}_x \left[\left\{ \psi_i(Y; \theta(x)) \psi_j(Y; \theta(x)) \right\} \right] \right]_{p \times p}, \quad (\text{équivalente à } I_\theta(x) \text{ lorsque } p = 1)$$

où $\psi_i(Y; t)$ désigne la i -ième composante du vecteur des dérivées partielles $\psi(Y; t) \in \mathbb{R}^p$. Soit

$$\Sigma_x := \{\mathbf{I}_\theta(x)\}^{-1} \times \frac{1}{f_X(x)}.$$

En adaptant de manière convenable les hypothèses sur la distribution au cadre multivarié, nous obtenons le théorème suivant.

Théorème 3.4.1 *Supposons que les hypothèses (F.1-6), (H.1-3) et (K.1-3) soient vérifiées. Alors, nous avons, lorsque $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \mathbf{u}^T \{\hat{\theta}_n(x) - \theta(x)\} - \sigma_{\theta, \mathbf{u}}(I) \right| = o_{\mathbf{P}}(1), \quad (3.13)$$

où

$$\sigma_{\theta, \mathbf{u}}(I) = \sup_{x \in I} \left\{ \mathbf{u}^T \Sigma_x \mathbf{u} \int_{\mathbf{R}} [K^2(v)] dv \right\}^{1/2}.$$

Si la fenêtre satisfait (H.2-5) nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \mathbf{u}^T \{\hat{\theta}_n(x) - \theta(x)\} - \sigma_{\theta, \mathbf{u}}(I) \right| = o(1) \quad \text{presque sûrement.}$$

Applications statistiques

Lorsque $p = 2$, nous appliquons le théorème 3.4.1 et (3.13) pour les choix particuliers de vecteur $\mathbf{u}_1 = (1, 0)^T$ et $\mathbf{u}_2 = (0, 1)^T$ afin d'obtenir des intervalles de confiance uniformes en $x \in I$ pour chacune des composantes du vecteur de paramètre $\theta(x) = (\theta_1(x), \theta_2(x))^T$.

Exemple 3.4.1 Si nous supposons que $Y|X = x$ suit une loi Gamma $(\alpha(x), \beta(x))$ à deux paramètres. Nous posons alors $\theta(x) = (\alpha(x), \beta(x))$, avec $\alpha(x) > 0$ et $\beta(x) > 0$. Nous avons,

$$g(y; \theta(x)) = \frac{(y)^{\alpha(x)-1} \exp[-y/\beta(x)]}{\beta(x)^{\alpha(x)} \Gamma(\alpha(x))},$$

lorsque $x \in I$ et $y > 0$.

Exemple 3.4.2 Lorsque $Y|X = x$ suit une loi Weibull $\theta(x) = (c(x), \alpha(x))$ à deux paramètres. La densité conditionnelle est donc de la forme suivante

$$g(y; \theta(x)) = \frac{c(x)}{\alpha(x)} \left(\frac{y}{\alpha(x)} \right)^{c(x)-1} \exp \left(- \left\{ \frac{y}{\alpha(x)} \right\}^{c(x)} \right),$$

avec $c(x) > 0$, $\alpha(x) > 0$, $x \in I$ et $y > 0$.

En conclusion, si nous travaillons dans un cadre semi-paramétrique, où la loi conditionnelle de Y sachant $X = x$ est supposée proche d'une certaine famille de distributions, notre méthode permet de construire des intervalles de confiance asymptotiquement optimaux pour les différents paramètres de la distribution modélisée. Nous notons également que les hypothèses faites sur la distribution sont faciles à vérifier en pratique et les conditions sur le noyau et la fenêtre ne sont pas restrictives.

3.5 Démonstration

Posons, pour $x \in I$ et $t \in T$,

$$\hat{r}'_n(x, t) = \frac{\partial}{\partial t} \hat{r}_n(x, t) = \frac{1}{nh_n} \sum_{i=1}^n \psi'(Y_i; t) K\left(\frac{x - X_i}{h_n}\right),$$

et

$$\hat{r}''_n(x, t) = \frac{\partial^2}{\partial t^2} \hat{r}_n(x, t) = \frac{1}{nh_n} \sum_{i=1}^n \psi''(Y_i; t) K\left(\frac{x - X_i}{h_n}\right),$$

avec $\psi'(y; t) = \frac{\partial}{\partial t} \psi(y; t)$ et $\psi''(y; t) = \frac{\partial^2}{\partial t^2} \psi(y; t)$ proprement définis via (F.3).

La démonstration est fondée sur un résultat remarquable de Einmahl et Mason (2000) [42]. Par convenance, nous désignons par $\mathcal{F} = \{f(\cdot; t) : t \in T\}$ une classe de fonctions indexée par le paramètre $t \in T$. Nous posons,

$$\hat{l}_n(x, t) := \sum_{i=1}^n f(Y_i; t) K\left(\frac{x - X_i}{h_n}\right)$$

et

$$l_n(x, t) := \mathbb{E} \left[\hat{l}_n(x, t) \right].$$

Théorème 3.5.1 *Supposons les hypothèses (K.1–3), (H.2–4–5) vérifiées. Nous supposons également que*

$$\begin{aligned} f_X &\text{ est continue et strictement positive sur } J, \\ f_{X,Y} &\text{ est continue sur } J \times \mathbb{R}. \end{aligned}$$

Si la classe de fonctions \mathcal{F} est bornée, nous avons

$$\lim_{n \rightarrow \infty} \sup_{t \in T} \sup_{x \in I} \frac{\pm \left\{ \hat{l}_n(x, t) - l_n(x, t) \right\}}{\sqrt{2nh_n \log(1/h_n)}} \stackrel{p.s.}{=} \sigma_{\mathcal{F}},$$

où

$$\sigma_{\mathcal{F}}^2 := \sup_{t \in T} \sup_{x \in I} \mathbb{E}_x \left[\{f(Y; t)\}^2 \right] f_X(x) \|K\|_2^2.$$

Ce théorème est une simple conséquence du théorème 1, p. 4, [42]. \square

A présent, nous considérons les classes de fonctions suivantes

$$\Psi_0 := \left\{ \psi(\cdot; t) : t \in T \right\},$$

$$\Psi_1 := \left\{ \psi'(\cdot; t) : t \in T \right\},$$

$$\Psi_2 := \left\{ \psi''(\cdot; t) : t \in T \right\}.$$

Les hypothèses (F.2-3-4) nous assurent que ces classes sont bornées. Nous pouvons donc appliquer le théorème 3.5.1 et conclure que, sous les hypothèses du théorème 3.3.1, nous avons, lorsque $n \rightarrow \infty$,

$$\sup_{t \in T} \sup_{x \in I} \pm \left\{ \hat{r}_n(x, t) - r_n(x, t) \right\} \stackrel{p.s.}{=} O \left(\left\{ \frac{\log(1/h_n)}{nh_n} \right\}^{1/2} \right), \quad (3.14)$$

$$\sup_{t \in T} \sup_{x \in I} \pm \left\{ \hat{r}'_n(x, t) - r'_n(x, t) \right\} \stackrel{p.s.}{=} O \left(\left\{ \frac{\log(1/h_n)}{nh_n} \right\}^{1/2} \right), \quad (3.15)$$

$$\sup_{t \in T} \sup_{x \in I} \pm \left\{ \hat{r}''_n(x, t) - r''_n(x, t) \right\} \stackrel{p.s.}{=} O \left(\left\{ \frac{\log(1/h_n)}{nh_n} \right\}^{1/2} \right). \quad (3.16)$$

Ces résultats nous seront utiles au cours de la démonstration.

Dans le cadre de l'estimation par la méthode du maximum de vraisemblance ou plus généralement de la M -estimation, l'obtention de lois limites ou l'étude du comportement limite de M -estimateurs se décompose en trois étapes principales :

- consistance
- vitesse de convergence
- loi limite exacte.

Sous les hypothèses du théorème 3.3.1, il est facile de montrer l'existence d'une suite d'estimateurs $\hat{\theta}_n(x)$ solutions de (3.8) et consistants (voir, par exemple, Serfling (1980), p.147-148, [123]), i.e. telle que

$$\hat{\theta}_n(x) \xrightarrow{\mathbf{P}} \theta(x), \text{ lorsque } n \rightarrow \infty. \quad (3.17)$$

Ci-dessous, nous démontrerons l'existence et la consistance forte de $\hat{\theta}_n(x)$, avec une vitesse de convergence préliminaire.

-Consistance de l'estimateur $\hat{\theta}_n(x)$

Nous introduisons une sous-classe de Ψ_0 , définie par,

$$\Psi_\alpha := \left\{ \psi(\cdot, \theta(x) + t) : |t| \leq \alpha \right\}.$$

D'après l'hypothèse (F.1) nous admettons l'existence de

$$m_0 := \inf_{x \in J} \min\{f_X(x), I_\theta(x)\} > 0.$$

Par convenance, nous rappelons l'hypothèse (F.6) : il existe des constantes positives C_1 et C_2 telles que

$$\inf_{x \in J} \mathbb{E}_x \left[-\psi'(Y; \theta(x) + \epsilon) \right] > C_2 > 0, \quad \text{lorsque } |\epsilon| \leq C_1.$$

Comme la fonction $\theta(\cdot)$ est continûment dérivable sur J , nous avons, uniformément en $x, z \in I$,

$$w_\theta(h) = \sup_{|x-z| \leq h} |\theta(x) - \theta(z)| = O(h).$$

D'après (3.14), lorsque $n \rightarrow \infty$, nous obtenons

$$\left| \hat{r}_n(x, \theta(x) \pm \epsilon_n) - r_n(x, \theta(x) \pm \epsilon_n) \right| \stackrel{p.s.}{\leq} C_\alpha \times \left\{ \frac{\log(1/h_n)}{nh_n} \right\}^{1/2} =: L_{n,\alpha}. \quad (3.18)$$

Ci-dessus, nous posons $\epsilon_n = \max\{3L_{n,\alpha}/(m_0C_2), 2w_\theta(h_n)\}$ qui tend vers zéro lorsque $n \rightarrow \infty$. En utilisant un argument de conditionnement, nous avons

$$r_n(x, \theta(x) + \epsilon_n) = \int_{\mathbf{R}} h^{-1} \mathbb{E}_t \left[\psi(Y, \theta(x) + \epsilon_n) \right] K\left(\frac{x-t}{h}\right) f_X(t) dt.$$

Nous effectuons à présent un développement de Taylor de la fonction $\psi(y, \theta(x) + \epsilon_n)$ autour de $\theta(t)$. Nous obtenons,

$$r_n(x, \theta(x) + \epsilon_n) = - \int_{\mathbf{R}} h^{-1} \mathbb{E}_t \left[-\psi'(Y, \theta(t) + \xi) \right] K\left(\frac{x-t}{h}\right) f_X(t) \{\epsilon_n - (\theta(t) - \theta(x))\} dt,$$

où $|\xi| \leq \epsilon_n + w_\theta(h) \leq C_1$, lorsque n suffisamment grand. En conséquence, comme le noyau K est à support compact, i.e. $|t-x| = O(h) = o(1)$, nous concluons que

$$r_n(x, \theta(x) + \epsilon_n) \leq -m_0C_2\{\epsilon_n - w_\theta(h)\} \int_{\mathbf{R}} h^{-1} K\left(\frac{x-t}{h}\right) dt \leq -m_0C_2\epsilon_n/2, \quad (3.19)$$

d'après la définition de ϵ_n . De la même manière, nous obtenons, lorsque $n \rightarrow \infty$,

$$r_n(x, \theta(x) - \epsilon_n) \geq m_0C_2\epsilon_n/2. \quad (3.20)$$

Maintenant, en combinant (3.18) avec (3.19) et (3.20), il s'ensuit les inégalités suivantes, lorsque $n \rightarrow \infty$, pour tout $x \in I$, avec probabilité un,

$$\hat{r}_n(x, \theta(x) + \epsilon_n) \leq L_{n,\alpha} - m_0C_2\epsilon_n/2 < 0,$$

$$\hat{r}_n(x, \theta(x) - \epsilon_n) \geq L_{n,\alpha} - m_0C_2\epsilon_n/2 > 0.$$

Les deux inégalités ci-dessus impliquent que, pour tout $x \in I$, presque sûrement, il existe une solution $\hat{\theta}_n(x) \in [\theta(x) - \epsilon_n, \theta(x) + \epsilon_n]$ de (3.8) telle que $\hat{r}_n(x, \hat{\theta}_n(x)) = 0$. En conclusion, nous avons prouvé l'existence d'une suite $\hat{\theta}_n(x)$ de solutions de l'équation (3.8), telle que

$$\sup_{x \in I} \left| \hat{\theta}_n(x) - \theta(x) \right| \leq \epsilon_n, \quad \text{presque sûrement.}$$

Cette dernière inégalité est équivalente à

$$\sup_{x \in I} \left| \hat{\theta}_n(x) - \theta(x) \right| \stackrel{p.s.}{=} O \left(\left\{ \frac{\log(h_n^{-1})}{nh_n} \right\}^{1/2} + h_n \right) = o(1).$$

Sous des hypothèses un peu moins fortes sur la fenêtre, nous obtenons également

$$\sup_{x \in I} \left| \hat{\theta}_n(x) - \theta(x) \right| \stackrel{\mathbf{P}}{=} O \left(\left\{ \frac{\log(h_n^{-1})}{nh_n} \right\}^{1/2} + h_n \right) = o(1), \quad (3.21)$$

qui implique bien (3.17).

-Vitesse de convergence pour l'estimateur $\hat{\theta}_n(x)$

La fonction de vraisemblance locale définie en (3.8) admet un développement de Taylor autour de $\theta(x)$ d'après (F.3). Nous avons, lorsque $|\xi| \leq |\hat{\theta}_n(x) - \theta(x)|$,

$$\begin{aligned} \hat{r}_n(x, \hat{\theta}_n(x)) &= \hat{r}_n(x, \theta(x)) + \hat{r}'_n(x, \theta(x)) \{ \hat{\theta}_n(x) - \theta(x) \} \\ &\quad + \{ \hat{\theta}_n(x) - \theta(x) \}^2 \hat{r}''_n(x, \theta(x) + \xi), \end{aligned}$$

où le dernier terme constitue le reste de Lagrange d'ordre 1. Nous obtenons, via la propriété de consistance de notre estimateur (3.21), lorsque $n \rightarrow \infty$,

$$\hat{r}_n(x, \hat{\theta}_n(x)) = \hat{r}_n(x, \theta(x)) + \hat{r}'_n(x, \theta(x)) \{ \hat{\theta}_n(x) - \theta(x) \} + o_{\mathbf{P}}(1). \quad (3.22)$$

D'après (3.8), nous pouvons écrire (3.22) ainsi

$$\{ \hat{\theta}_n(x) - \theta(x) \} = - \frac{\hat{r}_n(x, \theta(x))}{\hat{r}'_n(x, \theta(x))} + o_{\mathbf{P}}(1). \quad (3.23)$$

Remarque 3.5.1 L'approximation (3.23) est un argument essentiel de notre démonstration. Le comportement asymptotique associée à la déviation $\{ \hat{\theta}_n(x) - \theta(x) \}$ se réduit à l'étude d'un terme de régression. Ainsi, nous pouvons appliquer la méthodologie développée dans la section 2.2 afin de démontrer (3.9) ou (3.10). Notons enfin que le terme en $o_{\mathbf{P}}(1)$ ci-dessus est tel que l'approximation (3.23) est équivalente à

$$\{ \hat{\theta}_n(x) - \theta(x) \} = - \frac{\hat{r}_n(x, \theta(x))}{\hat{r}'_n(x, \theta(x))} + o_{\mathbf{P}} \left(\left\{ \frac{\log(h_n^{-1})}{nh_n} \right\}^{1/2} \right).$$

En conséquence, nous avons

$$\left\{ \frac{nh_n}{\log(h_n^{-1})} \right\}^{1/2} \{ \hat{\theta}_n(x) - \theta(x) \} = \left\{ \frac{nh_n}{\log(h_n^{-1})} \right\}^{1/2} \left\{ \frac{\hat{r}_n(x, \theta(x))}{-\hat{r}'_n(x, \theta(x))} \right\} + o_{\mathbf{P}}(1). \quad (3.24)$$

Dans un premier temps, nous considérons le terme du numérateur $\hat{r}_n(x, \theta(x))$. Nous avons clairement d'après (3.6) combinée au lemme de Bochner,

$$\mathbb{E}[\hat{r}_n(x, \theta(x))] = o(h_n).$$

Une étude plus précise du biais est présentée dans le lemme 3.5.1 ci-dessous. Nous posons $v_\theta(x) := f_X(x)I_\theta(x)$, $s_\theta(x) := \mathbb{E}_x[\psi''(Y; \theta(x))]$, et

$$b(x) := \left\{ \theta''(x)v_\theta(x) + 2v'_\theta(x)\theta'(x) + \{\theta'(x)\}^2 f_X(x)s_\theta(x) \right\} \mu_2(K).$$

Lemme 3.5.1 *Sous les hypothèses du théorème 3.3.1, nous obtenons, lorsque le noyau K est d'ordre 2,*

$$\sup_{x \in I} \left| \mathbb{E}[\hat{r}_n(x, \theta(x))] \right| = O(h_n^2). \quad (3.25)$$

Si les conditions de régularité en (F.5) sont vérifiées, il s'ensuit,

$$\left| \mathbb{E}[\hat{r}_n(x, \theta(x))] \right| = \frac{1}{2} b(x) h_n^2 (1 + o(1)). \quad (3.26)$$

$$\begin{aligned} \mathbb{E}[\hat{r}_n(x, \theta(x))] &= \int_{\mathbf{R}} h_n^{-1} K\left(\frac{v-x}{h_n}\right) f_x(v) \mathbb{E}_v[\psi(Y, \theta(x))] dv \\ &= \int_{\mathbf{R}} K(u) f_X(x + uh_n) \mathbb{E}_{x+uh_n}[\psi(Y, \theta(x))] du. \end{aligned}$$

En développant $\psi(y, \theta(x))$ autour de $\theta(x + uh_n)$:

$$\begin{aligned} \psi(y, \theta(x)) &= \psi(y, \theta(x + uh_n)) + \{\theta(x) - \theta(x + uh_n)\} \psi'(y, \theta(x + uh_n)) \\ &\quad + \frac{1}{2} \{\theta(x) - \theta(x + uh_n)\}^2 \psi''(y, \theta(x + uh_n) + \epsilon), \end{aligned}$$

avec $|\epsilon| \leq |\theta(x) - \theta(x + uh_n)| \rightarrow 0$. Il s'ensuit, via (3.6) et (3.7),

$$\begin{aligned} \mathbb{E}[\hat{r}_n(x, \theta(x))] &= \int_{\mathbf{R}} K(u) v_\theta(x + uh_n) \{\theta(x + uh_n) - \theta(x)\} du \\ &\quad + \frac{1}{2} \int_{\mathbf{R}} K(u) f_X(x + uh_n) \mathbb{E}_{x+uh_n}[\psi''(Y, \theta(x) + \epsilon)] \{\theta(x) - \theta(x + uh_n)\}^2 du \\ &= (I) + (II) \end{aligned}$$

Nous avons,

$$\theta(x + uh_n) - \theta(x) = uh_n \theta'(x) + \frac{(uh_n)^2}{2} \times \theta''(x + \epsilon_1),$$

et

$$v_\theta(x + uh_n) = v_\theta(x) + uh_n v'_\theta(x) + \frac{(uh_n)^2}{2} \times v''_\theta(x + \epsilon_2).$$

avec $\epsilon_i \rightarrow 0$, $i = 1, 2$. En utilisant les propriétés du noyau $K(\cdot)$, nous obtenons finalement,

$$(I) = \frac{h_n^2}{2} \{ \theta''(x) v_\theta(x) + 2v'_\theta(x) \theta'(x) \} \int_{\mathbf{R}} u^2 K(u) du + o(h_n^2),$$

et

$$(II) = \frac{h_n^2}{2} \{ (\theta'(x))^2 f_X(x) s_\theta(x) \} \int_{\mathbf{R}} u^2 K(u) du + o(h_n^2).$$

□

Le lemme 3.5.1 ou (3.25) combiné à l'hypothèse sur la fenêtre (H.3) nous donnent l'approximation suivante

$$\left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \hat{r}_n(x, t) = \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \{ \hat{r}_n(x, t) - r_n(x, t) \} + o(1), \quad (3.27)$$

où $r_n(x, t) = \mathbb{E}[\hat{r}_n(x, t)]$. La démonstration de (3.9) repose maintenant sur la proposition 3.5.1 ci-dessous, qui établit une loi uniforme du logarithme sur l'intervalle I concernant la déviation $\{ \hat{r}_n(x, \theta(x)) - r_n(x, \theta(x)) \}$.

Proposition 3.5.1 *Supposons (F.1-4), (H.1-2), (K.1-3).*

Alors, nous avons, lorsque $n \rightarrow \infty$

$$\left| \left\{ \frac{nh_n}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_n(x, \theta(x)) - r_n(x, \theta(x)) \} - \sigma_r(I) \right| = o_{\mathbf{P}}(1),$$

où

$$\sigma_r^2(I) = \sup_{x \in I} \left\{ I_\theta(x) f_X(x) \right\} \int_{\mathbf{R}} K^2(u) du.$$

La démonstration est une légère modification de la preuve du corollaire 2.3.2. En effet, en reprenant les notations de la section 2.2, nous avons

$$\hat{r}_n(x, \theta(x)) = \hat{r}_{\psi, n}(x),$$

pour le choix particulier de $\psi(\cdot) = \psi(\cdot; t)$ avec $t = \theta(x)$. La déviation $\{ \hat{r}_n(x, t) - r_n(x, t) \}$ peut être vue comme un certain processus empirique indexé par une classe de fonctions dépendante de t ou indexée par le paramètre t . Soit $\alpha_n(\cdot)$ le processus empirique bivarié basé sur les couples d'observations $(X_1, Y_1), \dots, (X_n, Y_n)$ et indexé par une fonction $m : \mathbb{R}^2 \rightarrow \mathbb{R}$ à déterminer :

$$\alpha_n(m) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ m(X_i, Y_i) - \mathbb{E}[m(X_i, Y_i)] \right\}.$$

Lorsque $(u, v) \in I \times \mathbb{R}$, pour le choix de $m(u, v) = m_t(u, v) := \psi(v; t) K\left(\frac{x-u}{h_n}\right)$,

$$\{ \hat{r}_n(x, t) - r_n(x, t) \} = \frac{1}{n^{1/2} h_n} \times \alpha_n(m_t).$$

Les arguments utilisés pour conclure sont identiques à ceux du chapitre précédent. Ils ne seront pas répétés, par souci de concision. □

-Obtention de la loi limite uniforme du logarithme

Lemme 3.5.2 *Sous (F.1-4), (H.1-2), (K.1-3), nous avons, lorsque $n \rightarrow \infty$,*

$$\sup_{x \in I} \left| \left\{ \frac{\hat{r}'_n(x, \theta(x))}{I_\theta(x) f_X(x)} \right\} + 1 \right| \xrightarrow{\mathbf{P}} 0.$$

Nous remarquons que, via le lemme de Bochner,

$$\sup_{x \in I} \mathbb{E} \left[\hat{r}'_n(x, \theta(x)) \right] = \sup_{x \in I} r'_n(x, \theta(x)) = -I_\theta(x) f_X(x) + O(h).$$

De plus, nous avons, lorsque $n \rightarrow \infty$,

$$\sup_{x \in I} \left\{ \hat{r}'_n(x, \theta(x)) - r'_n(x, \theta(x)) \right\} \stackrel{\mathbf{P}}{=} O \left(\left\{ \frac{\log(1/h_n)}{nh_n} \right\}^{1/2} \right),$$

d'après la version en probabilité de (3.15). □

En combinant (3.24) et (3.27) puis le lemme 3.5.2 et la proposition 3.5.1, nous obtenons clairement la borne asymptotique (3.9). Le passage à la convergence presque sûre (3.10) s'appuie sur le lemme de Borel-Cantelli (voir, par exemple, [42]). Le théorème 3.4.1 est une conséquence directe du théorème 3.3.1.

Annexe A

A.1 Processus empirique et estimation fonctionnelle non-paramétrique

Le but de cette section est d'exhiber le lien qui existe entre l'estimation non-paramétrique de certaines fonctionnelles de distribution et l'étude du processus empirique. De manière générale, la théorie sur les processus empiriques est très utile car de nombreuses statistiques peuvent s'exprimer comme des fonctionnelles de la fonction de répartition empirique notée F_n . Soit $\{X_i : i \geq 1\}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathcal{X} = \mathbb{R}^p$, définis sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Plus précisément, on peut voir la variable X_i comme une application telle que $X_i : \Omega \rightarrow \mathcal{X}$, pour chaque $i \geq 1$. La **fonction de répartition empirique** basée sur les X_1, \dots, X_n est définie par

$$F_n(\mathbf{t}) := \frac{1}{n} \#\{X_i \leq \mathbf{t} : 1 \leq i \leq n\} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq \mathbf{t}\}, \quad \mathbf{t} \in \mathbb{R}^p. \quad (\text{A.1})$$

Pour insister sur le fait que la fonction F_n est aléatoire, c'est à dire dépendante de $\omega \in \Omega$, on peut utiliser l'écriture suivante :

$$F_n(\mathbf{t}, \omega) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i(\omega) \leq \mathbf{t}\}, \quad \mathbf{t} \in \mathbb{R}^p.$$

La théorie sur la fonction de répartition empirique a été essentiellement élaboré pour $p = 1$, i.e. pour des variables aléatoires réelles. On se réfère à l'article de Gaenssler et Stute (1979) [51] et au livre de Shorack et Wellner (1986) [124] pour une exposition complète des propriétés de F_n dans le cadre univarié. On remarque, dans un premier temps, que F_n est la fonction de répartition associée à la mesure empirique du n -échantillon $\{X_i : 1 \leq i \leq n\}$, définie par,

$$\mathbb{P}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i},$$

où $\delta_{\mathbf{x}}$ dénote la mesure de Dirac au point $\mathbf{x} \in \mathbb{R}^p$. Lorsqu'on regarde F_n comme une mesure aléatoire discrète, il s'ensuit, pour une fonction de score φ donnée,

$$\int_{\mathbb{R}^p} \varphi(\mathbf{t}) F_n(d\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i).$$

Ainsi, pour φ intégrable, on obtient clairement,

$$\int_{\mathbf{R}^p} \varphi(\mathbf{t}) F_n(d\mathbf{t}) \rightarrow \int_{\mathbf{R}^p} \varphi(\mathbf{t}) F(d\mathbf{t}) = \mathbb{E}[\varphi(Y)], \quad \text{presque sûrement.}$$

Exemple A.1.1 Pour le choix particulier de $\varphi(\cdot) = h^{-p} K\left(\frac{\mathbf{x} - \cdot}{h}\right)$, nous retrouvons l'estimateur à noyau [PR] multivarié de la densité (cf. (1.9)) :

$$\int_{\mathbf{R}^p} \varphi(\mathbf{t}) F_n(d\mathbf{t}) = \frac{1}{nh^p} \sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right).$$

Le **processus empirique** α_n est défini par,

$$\alpha_n(\mathbf{t}) := n^{1/2} \{F_n(\mathbf{t}) - F(\mathbf{t})\}.$$

En s'appuyant sur le théorème des quantiles, on peut restreindre l'étude du processus empirique au cas où la distribution des variables est uniforme sur $[0, 1]^p$.

Soit U_1, U_2, \dots , une suite de vecteurs aléatoires i.i.d. uniformément distribués sur $[0, 1]^p$. D'après le théorème des quantiles, nous avons,

$$X_i \leq \mathbf{t} \Leftrightarrow U_i \leq F(\mathbf{t}).$$

Nous dénotons par \bar{F}_n et $\bar{\alpha}_n$ la fonction de répartition empirique et le processus empirique fondés sur le n -échantillon $\{U_i : 1 \leq i \leq n\}$. Il s'ensuit,

$$F_n(\mathbf{t}) = \bar{F}_n(F(\mathbf{t})) \quad \text{et} \quad \alpha_n(\mathbf{t}) = \bar{\alpha}_n(F(\mathbf{t})).$$

Ainsi, lorsque F est continue,

$$\sup_{\mathbf{t} \in \mathbf{R}^p} |\alpha_n(\mathbf{t})| = \sup_{\mathbf{u} \in [0, 1]^p} |\bar{\alpha}_n(\mathbf{u})|.$$

Nous pouvons donc travailler avec $U_i : \Omega \rightarrow [0, 1]^p$ et,

$$\mathbb{P}\{U_i \leq \mathbf{t}\} = \bar{F}(\mathbf{t}) := \prod_{j=1}^p t_j, \quad \forall \mathbf{t} = (t_1, \dots, t_p) \in [0, 1]^p.$$

Le processus stochastique $\bar{\alpha}_n = \{\bar{\alpha}_n(\mathbf{t}) : \mathbf{t} \in [0, 1]^p\}$ est alors appelé **processus empirique uniforme multivarié** de taille n . L'étude du processus empirique uniforme permet d'obtenir des lois limites du type loi du logarithme itéré pour l'estimateur [PR] de la densité. La méthodologie employée s'appuie sur l'étude du comportement limite du module de continuité associé au processus empirique uniforme. En suivant les travaux de Stute [132], on introduit le module d'oscillation (ou de continuité) du processus empirique, qui est défini, pour chaque $0 < h < 1$, par

$$w_n(h) := \sup \left\{ \bar{\alpha}_n(t+s) - \bar{\alpha}_n(t) : 0 \leq t, t+s \leq 1, 0 \leq s \leq h \right\}.$$

ou, de manière équivalente,

$$w_n(h) := \sup \left\{ \bar{\alpha}_n(t) - \bar{\alpha}_n(s) : |t - s| \leq h \right\}.$$

Sous les hypothèses [C-R-S] sur h_n , Stute obtient la loi uniforme du logarithme suivante, concernant le module d'oscillation,

$$\lim_{n \rightarrow \infty} \frac{w_n(h_n)}{\sqrt{2h_n \log(h_n^{-1})}} = 1, \quad \text{presque sûrement.}$$

Deheuvels et Mason [26] ont amélioré les travaux de Stute, en démontrant une loi uniforme fonctionnelle du logarithme pour un certain processus d'incrément défini sur $[0, 1]$ par,

$$\left\{ \xi_n(h_n, t; \cdot) : 0 \leq t \leq 1 - h_n \right\}$$

où, pour chaque $0 \leq t \leq 1 - h_n$, $\xi_n(h_n, t; \cdot)$ est la fonction définie défini sur $[0, 1]$ par,

$$\xi_n(h_n, t; s) = \bar{\alpha}_n(t + h_n s) - \bar{\alpha}_n(t), \quad 0 \leq s \leq 1.$$

Ils démontrent que le processus d'incrément a pour ensemble limite la boule de Strassen, presque sûrement. Cette loi fonctionnelle leur permet de déterminer la vitesse exacte de convergence presque sûre de certains estimateurs non-paramétriques de la densité, parmi lesquels l'estimateur à noyau et l'estimateur par la méthode des plus proches voisins. Plus récemment, en s'appuyant sur une idée développée par Deheuvels et Mason [27], Einmahl et Mason [40] et [41] ont obtenu une loi du logarithme itéré pour l'estimateur à noyau de la régression, via l'étude du processus empirique local.

Le processus empirique local indexé par des ensembles

Soit U_1, U_2, \dots , une suite de vecteurs aléatoires indépendants et uniformément distribués sur $[0, 1]^d$. En suivant les notations de Deheuvels et Mason (1994) [27], soit \mathbb{B} la classe des Boréliens sur $[0, 1]^d$ et soit \mathbb{D} une sous-classe quelconque de \mathbb{B} . On introduit alors le **processus empirique uniforme indexé par l'ensemble \mathbb{D}** , défini par

$$\alpha_n(D) := n^{1/2} \{ \lambda_n(D) - \lambda(D) \}, \quad D \in \mathbb{D},$$

où λ désigne la mesure de Lebesgue sur \mathbb{R}^d et λ_n est **la mesure empirique uniforme indexée par \mathbb{B}** , telle que,

$$\lambda_n(B) := n^{-1} \sum_{i=1}^n \mathbf{I}\{U_i \in B\}, \quad B \in \mathbb{B}.$$

Soit $\mathbf{t} \in [0, 1]^d$ un point de \mathbb{R}^d fixé et soit \mathbb{C} une classe particulière de Boréliens de $[a, b]^d$, lorsque $a < b$ avec $b - a = 1$, $a, b \in \mathbb{R}$. Nous considérons alors la classe $\mathbb{D} = \{ \mathbf{t} + C : C \in \mathbb{C} \}$ telle que $\mathbf{t} + C \subseteq [0, 1]^d$, $\forall C \in \mathbb{C}$. Ces définitions nous permettent d'introduire le **processus empirique local au point \mathbf{t} indexé par l'ensemble \mathbb{C}** :

$$\Theta_n(C) := \Theta_n(C, h_n) = h_n^{-1/2} \alpha_n(\mathbf{t} + h_n^{1/d} C)$$

$$= h_n^{-1/2} \left\{ \sum_{i=1}^n \mathbb{I}\{U_i \in \mathbf{t} + h_n^{1/d} C\} - nh_n \lambda(C) \right\}, \quad C \in \mathbb{C}.$$

L'étude de ce processus par Deheuvels et Mason a permis de généraliser la loi fonctionnelle du logarithme itéré de Mason (1988) et notamment d'améliorer les résultats de Hall (1981) concernant la consistance ponctuelle presque sûre de l'estimateur [PR] multivarié de la densité. Entre autres, ils établissent également une notion d'indépendance asymptotique intéressante (cf. théorème 1.2, p. 1622, [27]).

L'approche du processus empirique local

Soient ξ_j , $j \in \mathbb{N}$, des vecteurs aléatoires i.i.d. à valeurs dans \mathbb{R}^d , définis sur un espace de probabilité arbitraire $(\Omega, \mathcal{A}, \mathbb{P})$ et de fonction de répartition commune $G(\cdot)$. On fixe $\mathbf{t} \in \mathbb{R}^d$ et $\mathbf{J} \in \mathcal{B}(\mathbb{R}^d) \supset \mathbb{B}$, c'est à dire l'ensemble \mathbf{J} est dans la classe des boréliens de \mathbb{R}^d . Alors, pour toute transformation bimesurable inversible $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$, on pose

$$A(h) := \mathbf{t} + h\mathbf{J} \quad (\text{avec } h\mathbf{J} := \{h(\mathbf{x}) : \mathbf{x} \in \mathbf{J}\}).$$

On peut visualiser $A(h)$ comme un voisinage du point \mathbf{t} de la forme $h\mathbf{J}$. Par exemple, si $d = 2$, $J = \mathcal{B}_2$ (la boule unité de \mathbb{R}^2) et $h = Id$, alors $A(h)$ est la boule unité de centre \mathbf{t} .

A présent, soit $(h_n)_{n \in \mathbb{N}}$ une suite de transformations bimesurables inversibles et supposons que, pour

$$A_n := A(h_n) \quad \text{et} \quad a_n := \mathbb{P}\{\xi_j \in A_n\}, \quad n \in \mathbb{N},$$

les conditions suivantes sont vérifiées

$$(A.1) \quad a_n > 0, \quad \forall n \in \mathbb{N};$$

$$(A.2) \quad na_n \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty;$$

$$(A.3) \quad a_n \rightarrow 0.$$

Pour chaque $n \in \mathbb{N}$, nous pouvons alors définir la **mesure empirique locale** au point \mathbf{t} ,

$$\nu_n(\mathbf{t}, \mathbf{B}) := \frac{1}{na_n} \sum_{i=1}^n \mathbb{I}\{\xi_i \in \mathbf{t} + h_n(\mathbf{J} \cap \mathbf{B})\}, \quad \mathbf{B} \in \mathcal{B}(\mathbb{R}^d).$$

Il apparaît clairement que, par sa forme, la mesure empirique locale est un outil approprié pour l'étude des estimateurs à noyaux de la densité ou de la régression.

Maintenant, nous considérons \mathcal{F} une classe de fonctions mesurables $f : \mathbb{R}^d \rightarrow \mathbb{R}$ de supports contenus dans \mathbf{J} (i.e. $\forall f \in \mathcal{F}$, $f(\mathbf{x}) = 0$ lorsque $\mathbf{x} \in \{\mathbb{R}^d / \mathbf{J}\}$). On peut alors introduire la mesure empirique locale indexée par \mathcal{F} ,

$$\nu_n(\mathbf{t}, f) := \int_{\mathbf{J}} f(\mathbf{x}) \nu_n(\mathbf{t}, d\mathbf{x}) = \frac{1}{na_n} \sum_{i=1}^n f(h_n^{-1}(\xi_i - \mathbf{t})), \quad f \in \mathcal{F},$$

où h_n^{-1} dénote l'inverse de h_n . En suivant Einmahl et Mason (1997) [40], on définit le **processus empirique local** au point \mathbf{t} , indexé par \mathcal{F} ,

$$L_n(\mathbf{t}, f, h_n) := (na_n)^{1/2} \left\{ \nu_n(\mathbf{t}, f) - \mathbb{E}[\nu_n(\mathbf{t}, f)] \right\}. \quad (\text{A.2})$$

Exemple A.1.2 Soient ξ_1, ξ_2, \dots , des variables aléatoires réelles i.i.d., de densité $f_\xi(\cdot)$ continue et positive dans un voisinage d'un point $x \in \mathbb{R}$ fixé. On pose $J = [-1/2, 1/2]$ et on définit $h_n(x) := h_n \times x$, telle que $h_n > 0$ et $h_n \rightarrow 0$ lorsque $n \rightarrow 0$. On considère comme classe de fonctions $\mathcal{F} = \{K\}$, avec $K(\cdot)$ fonction noyau vérifiant les hypothèses classiques. Alors, d'après (A.2),

$$L_n(x, K, h_n) = \frac{1}{na_n} \sum_{i=1}^n K\left(\frac{x - \xi_i}{h_n}\right) = h_n a_n^{-1} \hat{f}_{\xi;n}(x).$$

Exemple A.1.3 Soient $d = 2$ et $\xi_i = (X_i, Y_i)$, $i \in \mathbb{N}$, des variables aléatoires i.i.d. à valeurs dans \mathbb{R}^2 , admettant une densité jointe $f_{X,Y}$ et des densités marginales f_X et f_Y . On pose $J = [-1/2, 1/2] \times \mathbb{R}$, $\mathbf{t} = (t, 0)$, $t \in \mathbb{R}$, et, pour $(x, y) \in \mathbb{R}^2$, $h_n(x, y) := (h_n x, y)$, telle que $h_n > 0$ et $h_n \rightarrow 0$ lorsque $n \rightarrow 0$. Nous considérons alors comme classe de fonctions $\mathcal{F} = \{R\}$, avec

$$R(x, y) = y \times K(x), \quad (x, y) \in \mathbb{R}^2.$$

Il s'ensuit, via la définition (A.2),

$$L_n(x, R, h_n) = \frac{1}{na_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right) = h_n a_n^{-1} \hat{r}_n(x).$$

A.2 Le lemme de Bochner

Le résultat présenté ci-dessous est une version du lemme de Bochner (cf. Bosq et Lecoutre (1987) [13], Einmahl et Mason (2000) [42], p. 27, lemme 2.9), qui constitue un outil classique pour traiter le biais d'estimateurs à noyaux. On rappelle la définition de l'uniforme équicontinuité, qui est une généralisation de l'uniforme continuité.

Définition A.2.1 Soit \mathcal{F} une famille d'applications $f : X \rightarrow Y$ où X et Y sont des espaces métriques. On dit que \mathcal{F} est **uniformément équicontinue** si, pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que, pour tout couple $(x, z) \in X^2$ vérifiant $d(x, z) < \eta$, et toute fonction $f \in \mathcal{F}$, on ait $d(f(x), f(z)) < \epsilon$.

Les classes de fonctions rencontrées sont toujours relativement compactes par rapport à la topologie de la norme du supremum, ce qui est équivalent à l'uniforme équicontinuité d'après le théorème d'Arzelà-Ascoli. Si les classes de fonctions sont définies sur des ensembles compacts, ceci implique également l'uniforme bornitude.

Soit \mathcal{I} un pavé compact de \mathbb{R}^p fixé. Pour un certain $\tau > 0$, nous désignons par $\mathcal{J} = \mathcal{I}^\tau$ le τ -voisinage du pavé \mathcal{I} dans \mathbb{R}^p défini par :

$$\mathcal{J} = \{\mathbf{u} \in \mathbb{R}^p, \exists \mathbf{v} \in \mathcal{I} : \|\mathbf{u} - \mathbf{v}\|_+ \leq \tau\},$$

où $\|\cdot\|_+$ désigne la norme maximum sur \mathbb{R}^p , i.e. $\|\mathbf{u}\|_+ = \max_{1 \leq i \leq p} |u_i|$.

Résultat A.2.1 Lemme de Bochner : version multidimensionnelle

Soit \mathcal{F} une classe de fonctions $f : \mathbb{R}^p \rightarrow \mathbb{R}^d$, uniformément équicontinue et bornée sur le pavé \mathcal{J} . Soit $K : \mathbb{R}^p \rightarrow \mathbb{R}$ une fonction intégrable (i.e. $K \in L_1(\mathbb{R}^p)$). Nous avons, uniformément en $f \in \mathcal{F}$,

$$D(f, h) := \sup_{\mathbf{x} \in \mathcal{I}} \left| f * K_h(\mathbf{x}) - f(\mathbf{x}) \int_{\mathbb{R}^p} K(\mathbf{u}) d\mathbf{u} \right| \rightarrow \mathbf{0}_d, \text{ lorsque } h \searrow 0,$$

où

$$f * K_h(\mathbf{x}) := h^{-p} \int_{\mathbb{R}^p} f(\mathbf{t}) K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) d\mathbf{t} \quad \text{et} \quad \mathbf{0}_d \text{ dénote le vecteur nul de } \mathbb{R}^d.$$

Premièrement, nous remarquons que

$$\left| h^{-p} \int_{\mathbb{R}^p} f(\mathbf{t}) K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) d\mathbf{t} - f(\mathbf{x}) \int_{\mathbb{R}^p} K(\mathbf{u}) d\mathbf{u} \right| = \left| h^{-p} \int_{\mathbb{R}^p} \{f(\mathbf{x} - \mathbf{z}) - f(\mathbf{x})\} K\left(\frac{\mathbf{z}}{h}\right) d\mathbf{z} \right|,$$

suite aux changements de variable $\mathbf{z} = \mathbf{x} - \mathbf{t}$ et $\mathbf{z} = h \times \mathbf{u}$ respectivement.

Par la suite, en séparant le domaine d'intégration, lorsque $h \rightarrow 0$, uniformément en $f \in \mathcal{F}$,

$$\begin{aligned} D(f, h) &\leq \sup_{\mathbf{x} \in \mathcal{I}} \sup_{\|\mathbf{z}\|_+ \leq \delta} |f(\mathbf{x} - \mathbf{z}) - f(\mathbf{x})| \int_{\|\mathbf{z}\|_+ \leq \delta} h^{-p} K\left(\frac{\mathbf{z}}{h}\right) d\mathbf{z} \\ &\quad + 2 \sup_{\mathbf{x} \in \mathcal{I}} |f(\mathbf{x})| \int_{\|\mathbf{z}\|_+ \geq \delta} h^{-p} K\left(\frac{\mathbf{z}}{h}\right) d\mathbf{z} \\ &\leq \sup_{\mathbf{x} \in \mathcal{I}} \sup_{\|\mathbf{z}\|_+ \leq \delta} |f(\mathbf{x} - \mathbf{z}) - f(\mathbf{x})| \times \int_{\mathbb{R}^p} |K(\mathbf{u})| d\mathbf{u} \\ &\quad + 2 \sup_{\mathbf{x} \in \mathcal{I}} |f(\mathbf{x})| \times \int_{\|\mathbf{u}\|_+ > \delta h^{-p}} |K(\mathbf{u})| d\mathbf{u}, \end{aligned}$$

ce qui tend bien vers $\mathbf{0}_d$, d'après l'hypothèse d'équicontinuité (pour le premier terme) et d'uniforme bornitude sur le pavé \mathcal{J} de la classe de fonctions \mathcal{F} combinée à la décroissance vers zéro de h et l'intégrabilité de $K(\cdot)$ (pour le deuxième terme). \square

Remarque A.2.1 Les fonctions $f_X(\cdot)$, $r_\psi(\cdot)$ et $m_\psi(\cdot)$ sont uniformément continues (cf. section A.5) et bornées sur le pavé \mathcal{J} d'après les hypothèses (F.1–3) dans le cadre multi-varié ($X \in \mathbb{R}^p$, $Y \in \mathbb{R}^d$ et $\psi(Y) \in \mathbb{R}^q$) :

(F.1) $f_{X,Y}(\cdot, \cdot)$ est continue sur $\mathcal{J} \times \mathbb{R}^d$;

(F.2) $f_X(\cdot)$ est continue et strictement positive sur \mathcal{J} ;

(F.3) $Y \mathbb{I}\{X \in \mathcal{J}\}$ est bornée.

On rappelle que la fonction $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^q$ est borélienne et bornée sur les compacts de \mathbb{R}^d . La suite $\{h_n : n \geq 1\}$ est supposée vérifier simplement $h_n \searrow 0$, lorsque $n \rightarrow \infty$. Nous supposons les hypothèses (F.1–3) vérifiées, ainsi que

(F.5) $f_{X,Y}(\cdot, \cdot)$ et $f_X(\cdot)$ sont k -fois continûment différentiables sur $\mathcal{J} \times \mathbb{R}^d$;

Nous obtenons alors, comme conséquence du lemme A.2.1 de Bochner ,

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{I}} |f_{X;n}^{(\mathbf{k})}(\mathbf{x}) - f_X^{(\mathbf{k})}(\mathbf{x})| &= o(1) \\ \sup_{\mathbf{x} \in \mathcal{I}} |r_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - r_{\psi}^{(\mathbf{k})}(\mathbf{x})| &= \mathbf{o}(1) \\ \sup_{\mathbf{x} \in \mathcal{I}} |m_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - m_{\psi}^{(\mathbf{k})}(\mathbf{x})| &= \mathbf{o}(1). \end{aligned}$$

Ci-dessus, la notation $\mathbf{o}(1)$ désigne l'extension multidimensionnelle ou q -dimensionnelle de $o(1)$.

A.3 Inégalités exponentielles en dimension infinie

Cette section est consacrée à la présentation de résultats nécessaires à nos démonstrations. Les résultats principaux sont une inégalité exponentielle de type Borell-Bernstein en dimension infinie et une borne de moment, concernant la norme du supremum du processus empirique indexé par une classe de fonctions. Par souci de clarté, nous présentons un résumé des étapes essentielles qui permettent de démontrer de telles inégalités, en introduisant certaines notions clés.

Premièrement, on se place dans un cadre très général. Soient X, X_1, \dots, X_n des variables aléatoires i.i.d. définis sur un espace de probabilité $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ de distribution commune notée μ . Afin d'obtenir les propriétés asymptotiques du processus empirique indexé par une classe de fonctions, une des approches principales est dénommée la **symétrisation** (*symmetrization* ou *randomization*, en anglais). L'idée sous-jacente de la symétrisation ou du principe de symétrisation consiste à remplacer le processus empirique classique par une version "symétrisée" proche d'un certain processus sous-Gaussien, le **processus de Rademacher**. Avant de présenter la définition formelle du processus de Rademacher et ses propriétés fondamentales, nous introduisons quelques notations nécessaires à son introduction. Soit δ_t la mesure de Dirac au point t .

Définition A.3.1 Toute suite de variables aléatoires $\{\varepsilon_i : 1 \leq i \leq n\}$ i.i.d. de distribution

$$\mathcal{L}(\varepsilon_i) = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1},$$

est appelée **suite de Rademacher**.

Nous notons que la distribution des ε_i est symétrique, telle que $\mathbb{E}[\varepsilon_i] = 0, 1 \leq i \leq n$.

Soit $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ la mesure empirique associée à l'échantillon de variables aléatoires $\{X_1, \dots, X_n\}$. La mesure empirique \mathbb{P}_n peut être regardée comme une mesure aléatoire discrète mettant le poids $1/n$ à chaque observation, c'est à dire une combinaison linéaire des mesures de Dirac associées aux observations.

Soit \mathcal{F} une classe de fonctions mesurables $f : \mathcal{X} \rightarrow \mathbb{R}$. En référence à [145], nous utiliserons la notation suivante : pour une fonction f mesurable et Q une mesure signée, on note

$Qf = \int f dQ$ (en particulier $\mathbb{P}f = \mathbb{E}[f(X)]$). La symétrisation est fondée sur l'argument principal suivant : au lieu du processus empirique, de la forme (modulo \sqrt{n})

$$f \rightarrow (\mathbb{P}_n - \mathbb{P})f = \frac{1}{n} \sum_{i=1}^n \{f(X_i) - \mathbb{P}f\},$$

on considère le processus symétrisé suivant

$$f \rightarrow \mathbb{P}_n^o f = \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i),$$

où $(\varepsilon_i)_{i \leq n}$ désigne une suite de Rademacher indépendante des $\{X_1, \dots, X_n\}$. Remarquons dans un premier temps que ces deux processus sont naturellement centrés (car $\mathbb{E}[\varepsilon_i f(X_i) | X_i] = 0$, via la propriété de symétrie des ε_i). Puis, en conditionnant le processus symétrisé $\mathbb{P}_n^o f$ par rapport aux variables X_i , nous obtenons clairement un processus de Rademacher, c'est à dire un processus de la forme

$$X_{\mathbf{a}} = \sum_{i=1}^n a_i \varepsilon_i, \text{ avec } \mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n. \quad (\text{A.3})$$

Soit $x_i \in \mathcal{X}$ la réalisation de la variable X_i . Plus précisément, nous obtenons,

$$n\{\mathbb{P}_n^o f | X_i, 1 \leq i \leq n\} = X_{f\{\mathbf{x}\}} = \sum_{i=1}^n f(x_i) \varepsilon_i, \text{ avec } f\{\mathbf{x}\} = (f(x_1), \dots, f(x_n)) \in \mathbb{R}^n.$$

On rappelle la définition d'un processus sous-Gaussien :

Définition A.3.2 Un processus stochastique $\{X_t : t \in T\}$ est appelé **sous-Gaussien** par rapport à la semi-métrique d si il vérifie l'inégalité suivante sur ses incréments :

$$\mathbb{P}\{|X_s - X_t| > x\} \leq 2 \exp\left\{-\frac{x^2}{2d^2(s, t)}\right\}, \text{ pour tout } s, t \in T, x > 0.$$

Le processus de Rademacher vérifie, via l'inégalité de Hoeffding (cf. p. 100-101, [145]),

$$\mathbb{P}\left\{\left|\sum_{i=1}^n a_i \varepsilon_i\right| > x\right\} \leq 2 \exp\left\{-\frac{x^2}{2\|\mathbf{a}\|^2}\right\},$$

où ici $\|\cdot\|$ dénote la norme euclidienne sur \mathbb{R}^n . Il s'ensuit, via (A.3),

$$\mathbb{P}\{|X_{\mathbf{a}} - X_{\mathbf{b}}| > x\} \leq 2 \exp\left\{-\frac{x^2}{2\|\mathbf{a} - \mathbf{b}\|^2}\right\}.$$

Cette dernière inégalité nous assure que le processus de Rademacher est sous-Gaussien par rapport à la métrique ou distance Euclidienne sur \mathbb{R}^n . Cette propriété est particulièrement intéressante car les processus sous-Gaussiens ont des bornes connues sur leurs incréments qui vont faire apparaître la fameuse entropie (c'est à dire le logarithme du nombre de recouvrement). A ce propos, nous citons le corollaire 2.2.8, p. 101, dans [145]. Notons que, via une inégalité liée à la norme d'Orlicz du maximum d'un nombre fini de variables aléatoires, combinée aux contrôle des incréments ("*chaining method*" ou chaînage) décrit ci-dessus, nous pouvons alors obtenir une majoration du supremum d'un nombre infini de variables.

Définition A.3.3 Nombre de recouvrement

Soit $\{\mathcal{F}, d\}$ un espace métrique. Le **nombre de recouvrement**, noté $N(\epsilon, \mathcal{F}, d)$, est le nombre minimal de boules de d -rayon ϵ (c'est à dire, par rapport à la métrique d) nécessaires pour recouvrir \mathcal{F} .

$$N(\epsilon, \mathcal{F}, d) = \arg \min_{n \in \mathbb{N}} \left\{ \bigcup_{i=1}^n B(x_i, \epsilon) \supseteq \mathcal{F} : x_i \in \mathcal{F} \right\}.$$

(Dans les démonstrations, nous considérons des espaces normés, du type $L_2(Q)$, Q désignant une mesure de probabilité quelconque).

Aparté Il existe une notion similaire au nombre de recouvrement, appelée *packing number* ou nombre d'emboîtement, qui est rencontrée fréquemment dans la littérature.

Définition A.3.4 Nombre d'emboîtement

Soit $\{\mathcal{F}, d\}$ un espace métrique. Le **nombre d'emboîtement**, noté $D(\epsilon, \mathcal{F}, d)$, est le nombre maximal de points ϵ -séparés dans \mathcal{F} . On appelle $x, y \in \mathcal{F}$ deux points ϵ -séparés, dès lors que $d(x, y) > \epsilon$.

Pour clore cette digression, nous présentons un petit lemme qui confirme la similitude de ces deux notions.

Lemme A.3.1 *Soit $\{\mathcal{F}, d\}$ un espace métrique. Alors, $N(\epsilon, \mathcal{F}, d)$ et $D(\epsilon, \mathcal{F}, d)$ sont deux fonctions décroissantes en ϵ satisfaisant, pour chaque $\epsilon > 0$,*

$$N(\epsilon, \mathcal{F}, d) \leq D(\epsilon, \mathcal{F}, d) \leq N(\epsilon/2, \mathcal{F}, d).$$

Nous faisons l'hypothèse suivante, sans perte de généralité, il existe au plus $n = D(\epsilon, \mathcal{F}, d)$ points ϵ -séparés $\{x_1, \dots, x_n\} \in \mathcal{F}$. Alors, les boules

$$B(x_i, \epsilon) = \{x \in \mathcal{F} : d(x, x_i) \leq \epsilon\}, \quad i = 1, \dots, n,$$

doivent recouvrir l'espace \mathcal{F} . Sinon, il existerait un point $y \in \mathcal{F}$ tel que $d(y, x_i) > \epsilon$, pour $i = 1, \dots, n$. Cette dernière assertion est en contradiction avec l'hypothèse de départ, car x_1, \dots, x_n, y seraient alors ϵ -séparés. Donc

$$\mathcal{F} \subseteq \bigcup_{i=1}^n B(x_i, \epsilon),$$

ce qui implique la partie gauche de l'encadrement $N(\epsilon, \mathcal{F}, d) \leq D(\epsilon, \mathcal{F}, d)$. Maintenant, via l'inégalité triangulaire, nous avons, pour $1 \leq i \neq j \leq n$,

$$\epsilon < d(x_i, x_j) \leq d(x_i, x) + d(x, x_j) \leq 2 \max \{d(x_i, x), d(x, x_j)\}.$$

En conséquence, chaque boule $B(x, \epsilon/2)$ contient au plus un des n points $\{x_1, \dots, x_n\}$. Il s'ensuit la partie droite de l'encadrement, $D(\epsilon, \mathcal{F}, d) \leq N(\epsilon/2, \mathcal{F}, d)$. \square

Reprenons le cours de l'exposition des étapes permettant d'obtenir la borne exponentielle de type Bernstein concernant le supremum du processus empirique indexé par une classe

de fonctions. Il serait intéressant de pouvoir borner la norme sup du processus $\mathbb{P}_n - \mathbb{P}$ par celle du processus symétrisé puis grâce aux propriétés énoncées ci-dessus, d'appliquer une inégalité maximale de type Bernstein, en conditionnant toujours par rapport aux $\{X_1, \dots, X_n\}$. Avant de présenter ce résultat, il faut prendre garde à la possible non-mesurabilité du supremum $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$. Les résultats seront alors formulés en termes d'espérances extérieures ou *outer expectations*.

Définition A.3.5 Soit (Ω, \mathcal{A}, P) un espace de probabilité arbitraire et $T : \Omega \rightarrow \bar{\mathbb{R}}$ une application quelconque. L'**intégrale extérieure** de T par rapport à P est définie par :

$$\mathbb{E}^*[T] = \inf \{ \mathbb{E}[U] : U \geq T, U : \Omega \rightarrow \bar{\mathbb{R}} \text{ mesurable et telle que } \mathbb{E}[U] \text{ existe} \}$$

Nous pouvons maintenant énoncer le lemme central de notre argumentation.

Lemme A.3.2 (Symétrisation)

$$\mathbb{E}^* \left[\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \right] \leq \mathbb{E}^* \left[2 \|\mathbb{P}_n^o\|_{\mathcal{F}} \right], \quad (\text{A.4})$$

Le lemme de symétrisation est valable quelle que soit la classe de fonctions \mathcal{F} considérée. Après avoir approché le supremum $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$ par le supremum du processus symétrisé $\|\mathbb{P}_n^o\|_{\mathcal{F}}$, il reste à appliquer une inégalité maximale au membre de droite de l'inégalité (A.4), conditionnellement aux $\{X_1, \dots, X_n\}$. En d'autres termes, on peut continuer à majorer la norme $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$ en utilisant les propriétés structurelles des processus sous-Gaussiens. A ce stade du développement, il manque dans la littérature une version générale du théorème de Fubini pour les intégrales extérieures. Pour contourner ce problème, on utilise alors une hypothèse de mesurabilité appropriée. Il faut s'assurer que l'intégrande $\|\mathbb{P}_n^o\|_{\mathcal{F}}$ est conjointement mesurable en $\{X_1, \dots, X_n, \varepsilon_1, \dots, \varepsilon_n\}$. Comme les variables de Rademacher sont discrètes, ceci est vérifié si et seulement si les applications

$$(X_1, \dots, X_n) \rightarrow \left\| \sum_{i=1}^n e_i f(X_i) \right\|_{\mathcal{F}}$$

sont mesurables pour chaque n -uplet $(e_1, \dots, e_n) \in \{-1, 1\}^n$. Il s'ensuit la définition suivante.

Définition A.3.6 Une classe \mathcal{G} de fonctions mesurables $g : \mathcal{X} \rightarrow \mathbb{R}$ sur un espace de probabilité $\{\mathcal{X}, \mathcal{A}, \mathbb{P}\}$ est appelée une classe **\mathbb{P} -mesurable** si l'application

$$(X_1, \dots, X_n) \rightarrow \left\| \sum_{i=1}^n e_i f(X_i) \right\|,$$

est mesurable sur la complétion de l'espace produit $\{\mathcal{X}^n, \mathcal{A}^n, \mathbb{P}^n\}$, pour tout choix de n et tout vecteur $\{e_1, \dots, e_n\} \in \mathbb{R}^n$.

Pour nos applications, nous utilisons la notion suivante qui implique la \mathbb{P} -mesurabilité.

Définition A.3.7 Supposons qu'il existe une sous-classe dénombrable \mathcal{G}_0 contenue dans \mathcal{G} telle que pour chaque fonction $g \in \mathcal{G}$, il existe une suite de fonctions $\{g_n : n \geq 1\}$ appartenant à \mathcal{G}_0 vérifiant

$$\lim_{n \rightarrow \infty} g_n = g \quad (\text{i.e. } \lim_{n \rightarrow \infty} g_n(x) = g(x), \text{ pour chaque } x \in \mathcal{X}).$$

La classe \mathcal{F} est alors appelée **pointwise mesurable** ou **mesurable ponctuellement** (ou encore mesurable point par point) et notée [m.p.].

Lorsque la classe \mathcal{F} est mesurable (dans le sens décrit ci-dessus), les mesures extérieures redeviennent des mesures et l'enveloppe mesurable de \mathcal{F} coïncide alors presque partout avec le supremum. En somme, le lemme de symétrisation permet de trouver des bornes pour le supremum du processus empirique indexé par \mathcal{F} . La présence d'une hypothèse spécifique sur la mesurabilité est une conséquence de la façon dont l'entropie uniforme est utilisée pour contrôler le supremum via randomisation ou symétrisation associé au manque d'une version générale du théorème de Fubini pour les intégrales extérieures. Dans la plupart des cas, il suffirait de supposer la classe dénombrable afin de simplifier l'exposition des résultats.

Nous sommes maintenant prêts pour présenter l'inégalité exponentielle en dimension infinie qui est l'outil central de nos démonstrations. L'inégalité suivante est due à Talagrand (1994) [137] (voir aussi Ledoux (1996) [88]). Comme nous allons le constater, les conditions d'entropie servent pour borner l'espérance du processus mais n'ont aucune incidence sur le contrôle de la déviation du processus par rapport à l'espérance. Dans la forme, cette inégalité est aussi proche de l'inégalité exponentielle de Borell concernant la probabilité de déviation par rapport à l'espérance de suprema de processus Gaussiens séparables. On reprend les notations usuelles pour le processus empirique.

Résultat A.3.1 Soit \mathcal{G} une classe [m.p.] (i.e. mesurable ponctuellement) de fonctions $g : \mathcal{X} \rightarrow \mathbb{R}$ satisfaisant, pour un certain $0 < M < \infty$,

$$\|g\|_\infty := \sup_{x \in \mathcal{X}} |g(x)| \leq M, \quad \forall g \in \mathcal{G}.$$

Alors, pour tout $t > 0$, nous avons, avec $A_1, A_2 > 0$ des constantes convenablement choisies,

$$\mathbb{P} \left\{ \|\alpha_n\|_{\mathcal{G}} \geq A_1 \left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} + t \right) \right\} \leq 2 \left\{ \exp \left(- \frac{A_2 t^2}{n \sigma_{\mathcal{G}}^2} \right) + \exp \left(- \frac{A_2 t}{M} \right) \right\}, \quad (\text{A.5})$$

où $\sigma_{\mathcal{G}}^2 = \sup_{x \in \mathcal{G}} \text{Var}[g(X)]$.

Le résultat A.3.1 est une conséquence du théorème 2.14.25, p. 255, de Van der Vaart et Wellner (1996), lui-même version du théorème 3.5, p. 45, [137]. Ce résultat est valable pour toute classe de fonctions uniformément bornée et la taille de la classe de fonctions n'intervient qu'à travers la norme $L_1 : \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}}$. Bien sur, sans l'hypothèse de mesurabilité, il faut remplacer \mathbb{P} et \mathbb{E} par \mathbb{P}^* et \mathbb{E}^* .

Pour la convergence presque sûre, on utilise la version suivante de l'inégalité (A.5), en combinaison avec l'inégalité de Ottaviani :

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

Le résultat A.3.1 permet de réduire de nombreux problèmes de convergence presque sûre à l'étude de la quantité de moment suivante :

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}}.$$

L'obtention de bornes pour cette quantité est liée à certaines conditions d'entropie, comme nous l'avons remarqué dans le paragraphe concernant la symétrisation (et plus particulièrement, le chaînage).

Soit $G(\cdot)$ une fonction **enveloppe** mesurable à valeurs finies satisfaisant, pour tout $x \in \mathcal{X}$,

$$G(x) \geq \sup_{g \in \mathcal{G}} |g(x)|.$$

Nous introduisons le nombre de recouvrement lié au nombre d'entropie.

$$N(\epsilon, \mathcal{G}) = \sup_Q N(\epsilon \sqrt{Q(G^2)}, \mathcal{G}, d_Q),$$

où le supremum est pris parmi toutes les mesures de probabilités Q sur $(\mathcal{X}, \mathcal{A})$ pour lesquelles $0 < Q(G^2) < \infty$. Ainsi les résultats présentés sont valables indépendamment de la mesure de probabilité \mathbb{P} associée aux variables aléatoires X_i , $1 \leq i \leq n$. La distance d_Q est la métrique sur $L^2(Q)$, définie par

$$d_Q(f, g) = d_{Q,2}(f, g) := \left\{ \int (f - g)^2 dQ \right\}^{1/2},$$

et $N(\epsilon, \mathcal{G}, d)$ désigne le nombre minimum de boules $B(g, \epsilon) := \{f : d(g, f) < \epsilon\}$ pour recouvrir l'ensemble \mathcal{G} (cf. la définition A.3.3). Nous pouvons remarquer que le nombre de recouvrement $N(\epsilon, \mathcal{G}, d)$ est proche du nombre de points (ou de fonctions $g \in \mathcal{G}$) constituant un ϵ -réseau de l'ensemble \mathcal{G} .

D'après Einmahl et Mason (2000) [42], la quantité de moment intervenant dans (A.5) est bornée sous certaines conditions précisées ci-dessous.

Résultat A.3.2 Einmahl et Mason (2000)

Soit \mathcal{G} une classe [m.p.] de fonctions bornées, mesurables et à valeurs réelles. La classe de fonctions \mathcal{G} est telle que, pour certaines constantes $\beta, \nu, C > 1, \sigma \leq 1/8C$, les quatre conditions suivantes sont vérifiées :

$$(C.1) \quad \mathbb{E}[G^2(X)] \leq \beta^2;$$

$$(C.2) \quad N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}, \quad 0 < \epsilon < 1;$$

$$(C.3) \quad \sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E}[g^2(X)] \leq \sigma^2;$$

$$(C.4) \quad \sup_{g \in \mathcal{G}} \|g\|_\infty \leq \frac{1}{2\sqrt{\nu+1}} \sqrt{n\sigma^2 / \log(\beta \vee 1/\sigma)}.$$

Alors nous avons, avec $A_3 > 0$ constante universelle,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq A_3 \sqrt{\nu n \sigma^2 \log(\beta \vee 1/\sigma)}. \quad (\text{A.6})$$

La démonstration s'appuie sur le lemme 5.2, p. 963–964, de Giné et Zinn (1984) [58] et certains résultats bien connus concernant les processus subgaussiens. Nous nous référons au prochain résultat pour une idée de démonstration. Nous introduisons la distance $d_{n,2}^{\mathbf{x}}$, métrique sur $L_2(\mathbb{P}_n)$, avec \mathbb{P}_n mesure empirique associée à $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$,

$$d_{n,2}^{\mathbf{x}}(g_1, g_2) := \left\{ \frac{1}{n} \sum_{i=1}^n \left\{ g_1(x_i) - g_2(x_i) \right\}^2 \right\}^{1/2}, \quad g_1, g_2 \in \mathcal{G}.$$

Nous rappelons que la distribution commune des variables $\{X_i : i \geq 1\}$ est notée μ .

Résultat A.3.3 Giné et Zinn (1984)

Soit \mathcal{G} une classe [m.p.] uniformément bornée de fonctions à valeurs réelles et définies sur \mathcal{X} . C'est à dire, pour tout $g \in \mathcal{G}$,

$$\|g\|_\infty \leq M.$$

Alors, lorsque $t \geq 32\sqrt{n}\sigma_0^2 > 0$ et $m \geq 1$, on a

$$\mathbb{P} \left\{ \sup_{g \in \mathcal{G}} \sum_{i=1}^n g^2(X_i) > t\sqrt{n} \right\} \leq 4\mu^n \{ \mathbf{x} : N(\rho/n^{1/4}, \mathcal{G}, d_{n,2}) > m \} + 8m \exp(-t\sqrt{n}/64M^2),$$

où $\sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E}[g^2(X)]$, $\rho = \min\{\sqrt{t}/8, n^{1/4}\} > 0$.

En reprenant les arguments de la démonstration du résultat A.3.2 avec quelques modifications mineures, Einmahl et Mason (2005) [43] obtiennent une borne similaire mais sous des hypothèses plus souples.

Résultat A.3.4 Soit \mathcal{G} une classe [m.p.] de fonctions mesurables bornées et à valeurs réelles telle que, pour certaines constantes $C, \nu \geq 1$ et $0 < \sigma \leq \beta$, les quatre conditions suivantes sont vérifiées :

$$(C.1) \quad \mathbb{E} \left[(G(X))^2 \right] \leq \beta^2;$$

$$(C.2) \quad N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}, \quad 0 < \epsilon < 1;$$

$$(C.3) \quad \sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E} \left[(g(X))^2 \right] \leq \sigma^2;$$

$$(C.4) \quad \sup_{g \in \mathcal{G}} \|g\|_\infty \leq \frac{1}{4} \sqrt{\frac{n\sigma^2}{\nu \log(C_1\beta/\sigma)}}, \quad \text{avec } C_1 = C^{1/\nu} \vee e.$$

Alors nous avons, pour une certaine constante $A_7 > 0$,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq A_7 \sigma \sqrt{\nu n \log(C_1 \beta / \sigma)}. \quad (\text{A.7})$$

En utilisant l'inégalité de Hoffman-Jørgensen sur l'intégrabilité de somme de variables aléatoires indépendantes (voir, par exemple, la proposition 6.8, p.156, dans Ledoux et Talagrand (1991) [89] ou le résultat A.3.1 ci-après), nous pouvons borner $\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}}$ (norme L_1) lorsque cette somme est bornée en probabilité (norme L_0). Ainsi démontrer l'inégalité (A.6) est équivalent à prouver que, pour une certaine constante A_4 , nous avons

$$t_n \leq A_4 \sigma \sqrt{\nu n \log(C_1 \beta / \sigma)},$$

où

$$t_n = \inf \left\{ t > 0 : \mathbb{P} \left\{ \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} > t \right\} \leq \frac{1}{24} \right\}.$$

En bref, il nous faut démontrer que, lorsque t est de l'ordre $\sigma \sqrt{\nu n \log(C_1 \beta / \sigma)}$, la probabilité ci-dessus est inférieure ou égale à $1/24$. Ceci implique que, pour n suffisamment grand, l'inégalité (A.7) est bien vérifiée, en remarquant que le premier terme de la borne [H-J] sera négligeable d'après les conditions (C.3-4) par exemple. Plus précisément, le terme

$$6 \times \mathbb{E} \left[\max_{1 \leq i \leq n} \left\{ \sup_{g \in \mathcal{G}} g(X_i) \right\} \right]$$

est négligeable asymptotiquement face à $6t_n$.

Nous commençons par séparer le domaine d'intégration via

$$F_n := \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \sup_{g \in \mathcal{G}} \sum_{i=1}^n g^2(x_i) \leq 64 \sigma^2 \right\}$$

et

$$G_n := \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \sum_{i=1}^n G^2(x_i) \leq 256 \beta^2 \right\}.$$

Ainsi, pour $t > 0$, nous avons

$$\begin{aligned} \mathbb{P} \left\{ \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} > t \right\} \leq \\ \int_{F_n \cap G_n} \mathbb{P} \left\{ \left\| \sum_{i=1}^n \varepsilon_i g(x_i) \right\|_{\mathcal{G}} > t \right\} \mu^n(d\mathbf{x}) + \mu^n(F_n^c) + \mu^n(G_n^c), \end{aligned} \quad (\text{A.8})$$

où μ^n désigne la mesure produit associée au n -échantillon (X_1, \dots, X_n) . Dans un premier temps, nous cherchons à borner le terme de gauche dans (A.8). En appliquant un résultat de Jain et Marcus (1978) [79] concernant les processus subgaussiens (ici de Rademacher), nous obtenons que, pour n'importe quel vecteur de réalisations $\mathbf{x} \in \mathcal{X}^n$,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(x_i) \right\|_{\mathcal{G}} \leq \mathbb{E} \left| \sum_{i=1}^n \varepsilon_i g_0(x_i) \right| + L \sqrt{n} \int_0^\infty \sqrt{\log N(\epsilon, \mathcal{G}, d_{n,2})} d\epsilon, \quad (\text{A.9})$$

avec L constante universelle et g_0 fonction arbitraire appartenant à \mathcal{G} . Nous pouvons également trouver une version de ce résultat dans le livre de Van der Vaart et Wellner (1996), corollaire 2.2.8, p. 101, [145]. Via l'inégalité de Cauchy-Schwartz, nous avons sur le sous-ensemble F_n de \mathcal{X}^n ,

$$\mathbb{E} \left| \sum_{i=1}^n \varepsilon_i g_0(x_i) \right| \leq \left\{ \sum_{i=1}^n g_0^2(x_i) \right\}^{1/2} \leq 8\sqrt{n}\sigma. \quad (\text{A.10})$$

De plus, pour $\mathbf{x} \in F_n$ et $g_1, g_2 \in \mathcal{G}$,

$$\begin{aligned} \{d_{n,2}^{\mathbf{x}}(g_1, g_2)\}^2 &= \frac{1}{n} \sum_{i=1}^n \{g_1(x_i) - g_2(x_i)\}^2 \\ &\leq \frac{2}{n} \sum_{i=1}^n \{g_1^2(x_i) + g_2^2(x_i)\} \\ &\leq 4 \times 64 \sigma^2. \end{aligned}$$

En d'autres termes, la distance $L_2(\mathbb{P}_n)$ au carré entre 2 fonctions choisies arbitrairement dans \mathcal{G} est inférieure ou égale à $256 \sigma^2$, lorsque $\mathbf{x} \in F_n$. Il s'ensuit, $N(\epsilon, \mathcal{G}, d_{n,2}) = 1$ pour $\epsilon > 16 \sigma$, lorsque $\mathbf{x} \in F_n$.

Soit $Q_n = 1/n (\delta_{x_1} + \dots + \delta_{x_n}) \equiv \mathbb{P}_n$ la mesure empirique associée à l'échantillon $\mathbf{x} = (x_1, \dots, x_n)$. Lorsque $\mathbf{x} \in G_n$, nous avons la borne $\sqrt{Q_n(G^2)} \leq 16\beta$. En remarquant que $N(\epsilon, \mathcal{G})$ décroît en ϵ , nous obtenons, pour tout $\mathbf{x} \in G_n$,

$$\begin{aligned} N(\epsilon, \mathcal{G}, d_{n,2}) &= N\left(\sqrt{Q_n(G^2)} \epsilon / \sqrt{Q_n(G^2)}, \mathcal{G}, d_{n,2}\right) \\ &\leq N\left(\sqrt{Q_n(G^2)} \epsilon / 16\beta, \mathcal{G}, d_{n,2}\right) \quad (\text{car } \mathbf{x} \in G_n \Rightarrow \sqrt{Q_n(G^2)} \leq 16\beta) \\ &\leq \sup_Q N\left(\sqrt{Q(G^2)} \epsilon / 16\beta, \mathcal{G}, d_Q\right) \\ &= N(\epsilon / 16\beta, \mathcal{G}). \end{aligned} \quad (\text{A.11})$$

Nous rappelons l'hypothèse (C.2),

$$N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu} \quad \text{avec } 0 < \epsilon < 1.$$

On pose $\epsilon_1 := \epsilon / 16\beta$ par convenance. Pour $0 < \epsilon \leq 16\sigma$, nous avons bien l'encadrement suivant

$$0 < \epsilon_1 \leq \frac{\sigma}{\beta} \leq 1.$$

Nous concluons que, via (A.11), pour $\mathbf{x} \in G_n$ et $0 < \epsilon \leq 16\sigma$,

$$N(\epsilon, \mathcal{G}, d_{n,2}) \leq C\epsilon_1^{-\nu} = C \left\{ \frac{16\beta}{\epsilon} \right\}^{\nu}.$$

On rappelle que $C_1 = C^{1/\nu} \vee e$. Ainsi sur $F_n \cap G_n$, nous avons

$$\int_0^\infty \sqrt{\log N(\epsilon, \mathcal{G}, d_{n,2})} d\epsilon$$

$$\begin{aligned}
&= \int_0^{16\sigma} \sqrt{\log N(\epsilon, \mathcal{G}, d_{n,2})} d\epsilon + \int_{16\sigma}^{\infty} \sqrt{\log N(\epsilon, \mathcal{G}, d_{n,2})} d\epsilon \\
&\leq \int_0^{16\sigma} \sqrt{\log \left(C \left\{ \frac{16\beta}{\epsilon} \right\}^\nu \right)} d\epsilon \quad (\epsilon > 16\sigma \Rightarrow N(\epsilon, \mathcal{G}, d_{n,2}) = 1 \text{ sur } F_n) \\
&\leq \sqrt{\int_0^{16\sigma} \log \left(\left\{ \frac{C_1 16\beta}{\epsilon} \right\}^\nu \right) d\epsilon} \\
&\leq A_5 \sigma \sqrt{\nu \log(C_1 \beta / \sigma)}, \tag{A.12}
\end{aligned}$$

où la dernière inégalité suit après quelque intégration par partie.

Remarque A.3.1 On note immédiatement que l'expression $\log(C_1 \beta / \sigma)$ nous permet d'obtenir du $\log(1/h_n)$ lors de l'application du résultat A.3.2 dans nos démonstrations. Pour cela, il suffit de contrôler en h_n la variance de la classe \mathcal{G} (cf. p. 1406 de [99]).

En combinant (A.9) avec (A.10) et (A.12), il s'ensuit, sur $F_n \cap G_n$,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(x_i) \right\|_{\mathcal{G}} \leq A_6 \sigma \sqrt{\nu n \log(C_1 \beta / \sigma)}, \tag{A.13}$$

où A_6 désigne une constante absolue. Cette dernière inégalité (A.13) entraîne, pour $t \geq 96 A_6 \sigma \sqrt{\nu n \log(C_1 \beta / \sigma)}$ et lorsque $\mathbf{x} = \{x_1, \dots, x_n\} \in F_n \cap G_n$,

$$\begin{aligned}
\mathbb{P} \left\{ \left\| \sum_{i=1}^n \varepsilon_i g(x_i) \right\|_{\mathcal{G}} \geq t \right\} &\leq \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(x_i) \right\|_{\mathcal{G}} \times t^{-1} \\
&\leq \frac{1}{96}.
\end{aligned}$$

En conséquence, d'après (A.8), il reste à démontrer que

$$\mu^n(F_n^c) + \mu^n(G_n^c) \leq \frac{1}{32}.$$

Et alors, nous pouvons poser $t_n = 96 A_6 \sigma \sqrt{\nu n \log(C_1 \beta / \sigma)}$ ou $A_7 \geq 96 A_6$ puis conclure.

Pour borner $\mu^n(G_n^c)$, nous utilisons l'inégalité de Markov,

$$\begin{aligned}
\mu^n(G_n^c) &= \mathbb{P} \left\{ \sum_{i=1}^n G^2(X_i) > n 256 \beta^2 \right\} \\
&\leq \frac{\mathbb{E}[G^2(X)]}{256 \beta^2} \leq \frac{1}{256} \quad (\text{par définition de } \beta).
\end{aligned}$$

Enfin, il faut démontrer l'inégalité suivante :

$$\mu^n(F_n^c) = \mathbb{P} \left\{ \sup_{g \in \mathcal{G}} \sum_{i=1}^n g^2(X_i) > n 64 \sigma^2 \right\} \leq \frac{7}{256}.$$

C'est maintenant que nous allons faire usage de l'obscur condition (C.4) et du lemme de Giné et Zinn (*the "square root trick"*) :

$$\sup_{g \in \mathcal{G}} \|g\|_\infty \leq \frac{1}{4} \sqrt{\frac{n\sigma^2}{\nu \log(C_1\beta/\sigma)}} =: M.$$

En appliquant le résultat A.3.3 pour le choix de $t = 64\sqrt{n}\sigma^2$ et en ajustant les diverses constantes convenablement, nous obtenons, pour $m \geq 1$,

$$\begin{aligned} \mu^n(F_n^c) &\leq 4\mu^n\{\mathbf{x} : N(\rho/n^{1/4}, \mathcal{G}, d_{n,2}) \geq m\} + 8m \exp\left\{-t\sqrt{n}/(64M^2)\right\} \\ &= 4\mu^n\{\mathbf{x} : N(\rho/n^{1/4}, \mathcal{G}, d_{n,2}) \geq m\} + 8m \exp\left\{-16\nu \log(C_1\beta/\sigma)\right\} \end{aligned}$$

On rappelle que,

$$\rho = \min\{\sqrt{t}/8, n^{1/4}\} = \min\{n^{1/4}\sigma, n^{1/4}\},$$

ce qui entraîne, pour $\sigma < 1$,

$$\mu^n(F_n^c) \leq 4\mu^n\{\mathbf{x} : N(\sigma, \mathcal{G}, d_{n,2}) \geq m\} + 8m \exp\left\{-4(\nu+1) \log(C_1\beta/\sigma)\right\}.$$

Ensuite, sur l'événement ou sous-ensemble G_n , nous avons

$$N(\epsilon, \mathcal{G}, d_{n,2}) \leq C \left\{\frac{16\beta}{\epsilon}\right\}^\nu.$$

En choisissant $m = \frac{3}{2} \times C \left\{\frac{16\beta}{\epsilon}\right\}^\nu$.

$$\begin{aligned} \mu^n(F_n^c) &\leq 4\mu^n\{\mathbf{x} : N(\sigma, \mathcal{G}, d_{n,2}) \geq m\} + 12C \left\{\frac{16\beta}{\epsilon}\right\}^\nu \exp\left\{-4\nu \log(C_1\beta/\sigma)\right\} \\ &\leq \frac{1}{64} + 12C (C_1\beta/\sigma)^{-4} \\ &\leq \frac{1}{64} + \frac{1}{256} = \frac{7}{256} \end{aligned}$$

□

Il est possible d'étendre ce résultat en affaiblissant la condition (C.4), qui est quelque peu restrictive ou du moins plus difficile à manipuler. Il s'ensuit le corollaire suivant, toujours d'après Einmahl et Mason (2005) [43] :

Corollaire A.3.1 Einmahl et Mason (2005)

Soit \mathcal{G} une classe de fonctions comme ci-dessus, satisfaisant les conditions (C.1–3) et, à la place de (C.4),

$$(C.5) \sup_{g \in \mathcal{G}} \|g\|_\infty \leq U, \text{ où } \sigma_0 \leq U \leq C_2\sqrt{n}\beta \text{ et } C_2 = \frac{1}{\sqrt{2\nu \log C_1}}.$$

Nous avons,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq A_7 \left\{ \sigma_0 \sqrt{\nu n \log(C_1\beta/\sigma_0)} + 2\nu U \log(C_3 n(\beta/U)^2) \right\},$$

avec $C_3 = C_1^2/16\nu$ et A_7 définie en (A.7).

La démonstration se scinde en deux parties.

1] Lorsque

$$U \leq \frac{1}{4\sqrt{\nu}} \times \sqrt{\frac{n\sigma_0^2}{\log(C_1\beta/\sigma_0)}},$$

en appliquant le résultat A.3.4 avec $\sigma = \sigma_0$, nous obtenons

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq A_7 \sqrt{\nu n \sigma_0^2 \log(C_1\beta/\sigma_0)}. \quad (\text{A.14})$$

2] Lorsque

$$\frac{1}{4\sqrt{\nu}} \times \sqrt{\frac{n\sigma_0^2}{\log(C_1\beta/\sigma_0)}} < U \leq C_2 \sqrt{n}\beta.$$

En remarquant que la fonction $t \rightarrow t^2 \{ \log(t^{-1}) \}^{-1}$ est monotone, il est possible de trouver un unique $\sigma \in]\sigma_0, \beta]$ tel que

$$U = \frac{1}{4\sqrt{\nu}} \times \sqrt{\frac{n\sigma^2}{\log(C_1\beta/\sigma)}} = \frac{\sqrt{n}}{4\sqrt{\nu}} \times \frac{\sigma}{\sqrt{\log(C_1\beta/\sigma)}}. \quad (\text{A.15})$$

Il s'ensuit, via (A.15),

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq A_7 \sqrt{\nu n \sigma^2 \log(C_1\beta/\sigma)} = 4A_7 \nu U \times \log(C_1\beta/\sigma). \quad (\text{A.16})$$

On rappelle que, par définition, $C_1 \geq e$ et donc $C_1\beta/\sigma \geq e$ d'après la définition de $\sigma \in]\sigma_0, \beta]$. On obtient alors,

$$\frac{1}{\sigma} \leq \frac{1}{\sigma} \times \sqrt{\log(C_1\beta/\sigma)} = \frac{\sqrt{n}}{4U\sqrt{\nu}} \quad \text{en utilisant encore la définition de } U \text{ en (A.15).}$$

Il s'ensuit,

$$C_1 \frac{\beta}{\sigma} \leq \frac{C_1}{4\sqrt{\nu}} \times \frac{\sqrt{n}\beta}{U} := \sqrt{C_3} \times \frac{\sqrt{n}\beta}{U}.$$

D'après (A.16) et l'inégalité ci-dessus, nous avons :

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq 2A_7 \nu U \times \log(C_3 n (\beta/U)^2). \quad (\text{A.17})$$

En combinant (A.14) et (A.17) la démonstration du corollaire est achevée. \square

Ce dernier corollaire peut être combiné au résultat A.3.1 pour obtenir des versions directement applicables (cf. théorème 2.1 et corollaire 2.2, p. 909-910, [56] et surtout l'inégalité 1, p. 1406, de Mason [99]). Nous citons également la section 19.6, de Van der Vaart (1998) [146], pour d'autres résultats concernant les inégalités maximales mais avec des conditions sur le nombre de crochets ou "*bracketing number*" (cf. lemme 19.34).

Exemple de classe mesurable ponctuellement

Dans cette sous-section, nous allons donner un exemple de classe [m.p.] suffisant pour la plupart de nos applications. Soit $\mathcal{K}(\cdot)$ une fonction à valeurs réelles, continue à droite et définie sur \mathbb{R} . Nous introduisons la classe de fonctions

$$\mathcal{G} = \{\mathcal{K}(\gamma \cdot + t) : \gamma \in \mathbb{R}, t \in \mathbb{R}\}.$$

D'après le lemme 5.1 de Deheuvels et Mason (2004) [29], nous avons l'assertion suivante :

Lemme A.3.3 *La classe de fonctions \mathcal{G} est mesurable ponctuellement.*

D'après la définition A.3.7, il faut démontrer l'existence d'une sous-classe dénombrable de fonctions \mathcal{G}_0 , telle que

$$\forall g \in \mathcal{G}, \exists (g_m)_{m \geq 1} \in \mathcal{G}_0 \text{ vérifiant } g_m(x) \rightarrow g(x), x \in \mathbb{R}.$$

On dénote par \mathbb{Q} l'ensemble des nombres rationnels et on introduit la classe de fonction

$$\mathcal{G}_0 = \{\mathcal{K}(\gamma \cdot + t) : \gamma \in \mathbb{Q}, t \in \mathbb{Q}\},$$

cette classe étant clairement dénombrable et contenue dans \mathcal{G} . Pour n'importe quelle fonction $g(u) = \mathcal{K}(\gamma u + t) \in \mathcal{G}$, $u \in \mathbb{R}$, nous posons, pour $m \geq 1$,

$$g_m(u) = \mathcal{K}(\gamma_m u + t_m), \text{ avec } \gamma_m = \frac{1}{m^2} \lfloor m^2 \gamma \rfloor + \frac{1}{m^2} \text{ et } t_m = \frac{1}{m} \lfloor m t \rfloor + \frac{2}{m}.$$

Soient $\varepsilon_m = \gamma_m - \gamma$ et $\delta_m = t_m - t$. On peut alors décomposer la différence

$$\Delta_m := \gamma_m u + t_m - (\gamma u + t) = \varepsilon_m u + \delta_m,$$

pour $u \in \mathbb{R}$ fixé. En utilisant les encadrements suivants,

$$\lfloor m^2 \gamma \rfloor \leq m^2 \gamma < \lfloor m^2 \gamma \rfloor + 1 \quad \text{et}$$

$$\lfloor m t \rfloor \leq m t < \lfloor m t \rfloor + 1,$$

nous obtenons,

$$0 < \varepsilon_m \leq \frac{1}{m^2} \quad \text{et} \quad \frac{1}{m} < \delta_m \leq \frac{2}{m}.$$

Ainsi, pour m suffisamment grand, il s'ensuit

$$\Delta_m = \delta_m (1 + o(1)) > 0.$$

Donc $\gamma_m u + t_m$ se trouve bien à droite de $\gamma u + t$. Ceci, combiné avec la continuité à droite de la fonction \mathcal{K} et le fait que $\gamma_m u + t_m \rightarrow \gamma u + t$, implique

$$\lim_{m \rightarrow \infty} g_m(u) = g(u), \quad \forall u \in \mathbb{R}.$$

La démonstration est alors achevée. □

Ce lemme se généralise aisément pour $\mathcal{K}(\cdot)$ fonction à valeurs réelles, continue à droite et définie sur \mathbb{R}^p . En conclusion, la continuité à droite de la fonction $K^{(\mathbf{k})}(\cdot)$ nous assure que la classe de fonctions

$$\mathbb{K}_{\mathbf{k}} = \left\{ K^{(\mathbf{k})} \left(\frac{\mathbf{x} - \cdot}{h} \right) : \mathbf{x} \in \mathbb{R}^p, h > 0 \right\}$$

est bien mesurable ponctuellement. Lorsqu'on passe aux autres classes de fonctions (cf. démonstration de la borne supérieure, arguments précédents le Lemme 2.4.5), on s'appuie sur la continuité des fonctions $c(\cdot)$ et $d(\cdot)$.

Condition d'entropie pour le produit de classes de fonctions

Nous présentons un lemme utile concernant l'entropie ou nombre de recouvrement du produit de classes de fonctions mesurables. En suivant précisément l'argumentation de Einmahl et Mason (2000), nous obtenons que le produit de deux classes de fonctions mesurables à nombre de recouvrement polynomial, l'une étant uniformément bornée et l'autre possédant une fonction enveloppe à valeurs finies, reste une classe de fonctions à nombre de recouvrement polynomial. Cette dernière propriété est très utile, au cours de la démonstration de nos théorèmes, afin d'appliquer la borne exponentielle de Talagrand combinée au contrôle de la norme L_1 du supremum du processus empirique symétrisé démontré par Einmahl et Mason.

Soient \mathcal{F} et \mathcal{G} deux classes de fonctions mesurables, à valeurs réelles et définies sur \mathcal{X} . La classe \mathcal{F} est supposée uniformément bornée, c'est à dire il existe une constante $M_{\mathcal{F}} > 0$ telle que

$$\sup_{x \in \mathcal{X}} |f(x)| = \|f\|_{\infty} \leq M_{\mathcal{F}}, \quad \text{pour chaque } f \in \mathcal{F}.$$

La classe \mathcal{G} vérifie la condition d'entropie notée $[\mathcal{E}]$, i.e.

$$\sup_{g \in \mathcal{G}} |g(x)| \leq G(x), \quad x \in \mathcal{X},$$

avec $G : \mathcal{X} \rightarrow \mathbb{R}$ fonction enveloppe mesurable à valeurs finies. Nous avons également, pour $\nu_{\mathcal{G}}, \nu_{\mathcal{F}}, C_{\mathcal{G}}, C_{\mathcal{F}}$, des constantes convenablement choisies,

$$N(\varepsilon [Q(G^2)]^{1/2}, \mathcal{G}, d_Q) \leq C_{\mathcal{G}} \varepsilon^{-\nu_{\mathcal{G}}}, \quad 0 < \varepsilon < 1, \quad (\text{A.18})$$

et

$$N(\varepsilon M_{\mathcal{F}}, \mathcal{F}, d_Q) \leq C_{\mathcal{F}} \varepsilon^{-\nu_{\mathcal{F}}}, \quad 0 < \varepsilon < 1, \quad (\text{A.19})$$

où la première inégalité est valable pour toutes les mesures de probabilité Q telles que $0 < Q(G^2) < \infty$ d'après (2.20), et la seconde est vraie pour toute mesure de probabilité Q . Nous rappelons que les hypothèses (A.18) et (A.19) spécifient exactement que les classes \mathcal{G} et \mathcal{F} sont à nombre de recouvrement polynomial.

Lemme A.3.4 *Sous les hypothèses précédentes, nous avons,*

$$N(\varepsilon M_{\mathcal{F}} [Q(G^2)]^{1/2}, \mathcal{F}\mathcal{G}, d_Q) \leq C_{\mathcal{F}\mathcal{G}} \varepsilon^{-(\nu_{\mathcal{G}} + \nu_{\mathcal{F}})}, \quad 0 < \varepsilon < 1,$$

pour une certaine constante $C_{\mathcal{F}\mathcal{G}} > 0$ finie.

Nous suivons la démonstration du lemme A.1, p. 35-36, de Einmahl et Mason (2000) [42]. Les deux inégalités (A.18) et (A.19) nous assurent l'existence d'un nombre $n_G \leq C_G \varepsilon^{-\nu_G}$ de fonctions g_1, \dots, g_{n_G} et d'un nombre $n_{\mathcal{F}} \leq C_{\mathcal{F}} \varepsilon^{-\nu_{\mathcal{F}}}$ de fonctions $f_1, \dots, f_{n_{\mathcal{F}}}$ tels que

$$\sup_{g \in \mathcal{G}} \min_{1 \leq j \leq n_G} d_Q(g, g_j) \leq \varepsilon [Q(G^2)]^{1/2}, \quad (\text{A.20})$$

et

$$\sup_{f \in \mathcal{F}} \min_{1 \leq i \leq n_{\mathcal{F}}} d_Q(f, f_i) \leq \varepsilon M_{\mathcal{F}}. \quad (\text{A.21})$$

Ces inégalités sont une conséquence directe de la définition du nombre de recouvrement.

Remarquons que (A.21) est vraie pour n'importe quelle mesure de probabilité, en particulier on peut écrire

$$\sup_{f \in \mathcal{F}} \min_{1 \leq i \leq n_{\mathcal{F}}} d_{\tilde{Q}}(f, f_i) \leq \varepsilon M_{\mathcal{F}},$$

où \tilde{Q} est la mesure associée à la Q -fonction de densité $x \rightarrow G^2(x)/Q(G^2)$. Il s'ensuit,

$$\begin{aligned} \sup_{f, g} \min_{i, j} d_Q(f \cdot g, f_i \cdot g_j) &\leq \sup_{f, g} \min_{1 \leq j \leq n_G} d_Q(f \cdot g, f \cdot g_j) + \sup_{f, g} \min_{1 \leq i \leq n_{\mathcal{F}}} d_Q(f \cdot g, f_i \cdot g) \\ &\leq M_{\mathcal{F}} \sup_{g \in \mathcal{G}} \min_{1 \leq j \leq n_G} d_Q(g, g_j) + \sup_{f \in \mathcal{F}} \min_{1 \leq i \leq n_{\mathcal{F}}} d_Q(G \cdot f, G \cdot f_i) \\ &\leq M_{\mathcal{F}} \times \varepsilon [Q(G^2)]^{1/2} + [Q(G^2)]^{1/2} \times \sup_{f \in \mathcal{F}} \min_{1 \leq i \leq n_{\mathcal{F}}} d_{\tilde{Q}}(f, f_i) \\ &\leq 2\varepsilon M_{\mathcal{F}} [Q(G^2)]^{1/2}. \end{aligned}$$

La dernière inégalité implique,

$$N(2\varepsilon M [Q(G^2)]^{1/2}, \mathcal{FG}, d_Q) \leq C_{\mathcal{F}} C_G \varepsilon^{-(\nu_G + \nu_{\mathcal{F}})},$$

ce qui clôt la démonstration. \square

Nous terminons cette section par quelques inégalités qui sont utiles à nos démonstrations.

-Quelques inégalités utiles

Pour l'étude de l'intégrabilité de sommes de variables aléatoires indépendantes, il existe plusieurs types d'inégalités, dont certaines célèbres sont isopérimétriques (inégalités de concentration). Pour notre part, nous présentons une version de la fameuse inégalité de Hoffman-Jørgensen, notée inégalité [H-J], qui introduit une notion d'équivalence entre différents modes de convergence. Plus précisément, si les sommes de variables aléatoires sont contrôlées en probabilité, on peut en déduire des bornes pour la norme L_p ($p > 0$) de cette somme, en s'assurant que le maximum parmi les variables aléatoires est bien borné pour la norme L_p . La version énoncée de l'inégalité [H-J] concerne des variables aléatoires symétriques, ce qui est bien le cas dans (A.6), d'après la définition des variables de Rademacher. Ci-dessous, B désigne un espace de Banach.

Proposition A.3.1 Inégalité [H-J]

Soit $0 < p < \infty$ et soient $\{X_i\}_{i \leq N}$ des variables aléatoires indépendantes dans $L_p(B)$.

Posons $S_N = \sum_{i=1}^N X_i$. Lorsque les X_i sont symétriques, et $t_0 = \inf \{t > 0; \mathbb{P}\{S_N > t\} \leq (8 \cdot 3^p)^{-1}\}$, alors, nous avons,

$$\mathbb{E}\|S_N\|^p \leq 2 \cdot 3^p \mathbb{E} \left[\max_{i \leq N} \|X_i\|^p \right] + 2(3t_0)^p.$$

Confer les pages 156 et 157, [89]. □

La proposition A.3.1 est utilisée dans la démonstration de (A.6) avec $p = 1$. Cette proposition permet d'établir aisément des équivalences de moments comme le montre le théorème 6.11, p. 158-159, [89]. Ces équivalences incluent notamment la partie indépendante du lemme de Borel-Cantelli.

L'inégalité maximale de Montgomery-Smith, (1993) [101] :

$$\mathbb{P} \left\{ \max_{1 \leq m \leq n} \|\alpha_m\|_{\mathcal{G}} > t \right\} \leq 9\mathbb{P} \left\{ \|\alpha_n\|_{\mathcal{G}} > \frac{t}{30} \right\}, \quad \text{pour tout } t > 0. \quad (\text{A.22})$$

Cette inégalité est, par nature, applicable lors de démonstration de résultats presque sûrs. Par exemple, nous avons l'inégalité suivante

$$\mathbb{P} \left\{ \max_{n_{k-1} < n \leq n_k} \|\alpha_n\|_{\mathcal{G}} > t \right\} \leq 9\mathbb{P} \left\{ \|\alpha_{n_k}\|_{\mathcal{G}} > \frac{t}{30} \right\}, \quad \text{pour tout } t > 0.$$

A.4 La loi du logarithme itéré multidimensionnelle

Le but de cette section est de présenter la démonstration complète de **la Loi du Logarithme Itéré de Hartman-Wintner** dans le cadre multivarié. Soit $\mathbf{y} \in \mathbb{R}^d$, un vecteur arbitraire. On note \mathbf{y}^T le vecteur transposé de \mathbf{y} vérifiant

$$\mathbf{y}^T \mathbf{y} = \sum_{i=1}^d y_i^2 = \|\mathbf{y}\|^2,$$

où $\|\cdot\| := \|\cdot\|_d$ dénote la norme euclidienne dans \mathbb{R}^d et y_i désigne la i -ième composante du vecteur \mathbf{y} .

Le lemme qui suit est un outil essentiel de la démonstration de la loi du logarithme itéré pour la fonction de répartition empirique. Cette loi fonctionnelle du logarithme itéré complète notamment les travaux de Strassen (1964) [131]. Soient $\mathbf{0}_d$ et \mathbf{I}_d le vecteur nul et la matrice identité d -dimensionnels respectivement.

Lemme A.4.1 Finkelstein (1971)

Soient Y, Y_1, Y_2, \dots , des vecteurs aléatoires à valeurs dans \mathbb{R}^d indépendants et identiquement distribués, tels que $\mathbb{E}[Y] = \mathbf{0}_d$ et $\mathbb{E}[Y^T Y] = \mathbf{I}_d$. Nous posons, pour $n \geq 3$,

$$Z_n := \sum_{i=1}^n Y_i / \sqrt{2n \log_2 n}.$$

Alors, la suite $\{Z_n : n \geq 3\}$ est presque sûrement relativement compacte et a pour ensemble limite $\mathcal{L} = \mathcal{B}_d = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\| \leq 1\}$.

Lorsque $d = 1$ les variables Y_i sont à valeurs réelles et nous pouvons donc appliquer le théorème de Hartman-Wintner qui spécifie : pour X, X_1, X_2, \dots des variables aléatoires réelles i.i.d. telles que $\mathbb{E}[X] = 0$ et $\text{Var}[X] = \mathbb{E}[X^2] = \sigma^2$, nous avons la loi du logarithme itéré suivante

$$\frac{\sum_{i=1}^n X_i}{\sqrt{2n \log_2 n}} \xrightarrow{p.s.} [-\sigma, \sigma],$$

où la notation $\xrightarrow{p.s.}$ signifie presque sûrement relativement compacte, c'est à dire la suite ci-dessus est p.s. relativement compacte avec comme ensemble limite l'intervalle $[-\sigma, \sigma]$. A présent, pour tout vecteur \mathbf{y} dans \mathbb{R}^d , les variables $\mathbf{y}^T Y, \mathbf{y}^T Y_1, \mathbf{y}^T Y_2, \dots$, sont i.i.d. à valeurs réelles. En conséquence, nous obtenons, d'après la LLI de Hartman-Wintner et la propriété de séparabilité de \mathbb{R}^d ,

$$\limsup_{n \rightarrow \infty} \mathbf{y}^T Z_n \stackrel{p.s.}{=} \|\mathbf{y}\|, \quad \forall \mathbf{y} \in \mathbb{R}^d. \quad (\text{A.23})$$

Il s'ensuit, d'après (A.23),

$$\text{comme } \langle \mathbf{y}, Z_n \rangle = \mathbf{y}^T Z_n \leq \|\mathbf{y}\| \times \|Z_n\|$$

$$\|\mathbf{y}\| \leq \limsup_{n \rightarrow \infty} \|Z_n\| \times \|\mathbf{y}\| \quad (\text{en passant à la limsup})$$

$$\text{donc } \limsup_{n \rightarrow \infty} \|Z_n\| \geq 1,$$

toutes les inégalités ci-dessus étant vraies presque sûrement.

La prochaine étape consiste à démontrer que, via (A.23), $\limsup_{n \rightarrow \infty} \|Z_n\| \stackrel{p.s.}{=} 1$ exactement. Nous supposons que $\limsup_{n \rightarrow \infty} \|Z_n(\omega)\| = 1 + \eta$. Il est alors possible de choisir une suite $(\mathbf{y}_n)_{n \geq 3}$ de vecteurs dans \mathbb{R}^d de norme 1 tels que $\forall n \geq 3, \cos(\mathbf{y}_n, Z_n) = 1$ (i.e. chaque direction \mathbf{y}_i est orthogonal au d -vecteur Z_i). Nous avons

$$\limsup_{n \rightarrow \infty} \mathbf{y}_n^T Z_n = \limsup_{n \rightarrow \infty} \|Z_n\| \times \|\mathbf{y}_n\| \times \cos(\mathbf{y}_n, Z_n) = 1 + \eta.$$

L'ensemble $\{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\| = 1\}$ est compact donc la suite $(\mathbf{y}_n)_{n \geq 3}$ admet une limite \mathbf{l} de norme 1 également. En utilisant (A.23), il s'ensuit

$$\limsup_{n \rightarrow \infty} \mathbf{y}_n^T Z_n = \|\mathbf{l}\| = 1 \text{ p.s.},$$

donc $\eta = 0$. En conséquence

$$\limsup_{n \rightarrow \infty} \|Z_n\| = 1 \text{ p.s.},$$

c'est à dire l'ensemble de points limites \mathcal{L} de la suite $\{Z_n : n \geq 3\}$ est contenu dans la boule unité \mathcal{B}_d d -dimensionnelle presque sûrement.

Pour la borne inférieure, nous utiliserons un argument de récurrence sur la dimension après avoir démontré que chaque point de la sphère unité est atteint par l'ensemble limite. Soit $\mathcal{S}_d = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\| = 1\}$ la sphère unité dans \mathbb{R}^d . D'après (A.23), $\forall \mathbf{y}_0 \in \mathcal{S}_d$,

$$\limsup_{n \rightarrow \infty} \mathbf{y}_0^T Z_n \stackrel{p.s.}{=} \|\mathbf{y}_0\| = 1.$$

On évalue à présent la distance entre \mathbf{y}_0 et Z_n ,

$$\|Z_n - \mathbf{y}_0\|^2 \leq \|\mathbf{y}_0\|^2 + \|Z_n\|^2 - 2\mathbf{y}_0^T Z_n.$$

En passant à la limite supérieure des deux cotés, nous obtenons clairement que cette distance tend presque sûrement vers 0. Donc chaque point de la sphère est atteint par la suite $(Z_n)_{n \geq 3}$, avec probabilité 1.

Nous considérons alors des variables auxiliaires V_1, V_2, \dots , i.i.d., à valeurs réelles, centrées réduites et indépendantes des vecteurs Y_i et nous construisons des vecteurs aléatoires à valeurs dans \mathbb{R}^{d+1} centrés et de matrice de variance-covariance identité,

$$R_i := (Y_i, V_i), \forall i \in \mathbb{N}.$$

Soit $Z_n^* = \sum_{i=1}^n R_i / \sqrt{2n \log_2 n}$. En utilisant un argument de récurrence, chaque point de la sphère de dimension $d+1$ est presque sûrement atteint par la suite $\{Z_n^* : n \geq 3\}$. Ainsi, l'ensemble de points limites \mathcal{L}^* associé à la suite $\{Z_n^* : n \geq 3\}$ contient \mathcal{S}_{d+1} . En appliquant la projection π définie par

$$\begin{aligned} \pi : \mathbb{R}^{d+1} &\longrightarrow \mathbb{R}^d \\ (x_1, \dots, x_{d+1}) &\longrightarrow (x_1, \dots, x_d) \end{aligned}$$

nous obtenons que l'ensemble de points limites de la suite $\{\pi(Z_n^*) : n \geq 3\}$ contient $\pi(\mathcal{S}_{d+1})$, i.e. l'ensemble de points limites \mathcal{L} de la suite $\{Z_n : n \geq 3\}$ contient \mathcal{B}_d (ces assertions étant vraies presque sûrement). \square

Lorsque la matrice de variance-covariance associée aux Y_i est supposée strictement définie positive ou inversible, il s'ensuit le corollaire suivant.

Corollaire A.4.1 *Soient Y, Y_1, Y_2, \dots des vecteurs aléatoires i.i.d. à valeurs dans \mathbb{R}^d tels que $\mathbb{E}[Y] = \mathbf{0}_d$ et $\mathbb{E}[Y^T Y] = \Sigma$ avec Σ strictement définie positive. Soit*

$$Z_n = \sum_{i=1}^n Y_i / \sqrt{2n \log_2 n},$$

alors la suite $\{Z_n\}_{n \geq 3}$ est presque sûrement relativement compacte et a pour ensemble limite l'ellipsoïde $\mathcal{E}_\Sigma = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{y}^T \Sigma^{-1} \mathbf{y} \leq 1\}$.

La matrice de variance-covariance Σ étant inversible, nous pouvons normaliser les variables $\{Y_i : 1 \leq i \leq n\}$ en leur appliquant l'opérateur linéaire $\Sigma^{-1/2} : \mathbb{R}^d \rightarrow \mathbb{R}^d$. On pose

$X_i = \Sigma^{-1/2}Y_i$. Il s'ensuit $\mathbb{E}[X_1] = \mathbf{0}_d$ et $\mathbb{E}[X_1^T X_1] = \Sigma^{-1/2} \Sigma \Sigma^{-1/2} = \mathbf{I}_d$. Donc, nous obtenons,

$$\sum_{i=1}^n \{\Sigma^{-1/2}Y_i\} / \sqrt{2n \log_2 n} \stackrel{p.s.}{\rightsquigarrow} \mathcal{B}_d.$$

Il s'ensuit,

$$\sum_{i=1}^n Y_i / \sqrt{2n \log_2 n} \stackrel{p.s.}{\rightsquigarrow} \Sigma^{1/2}(\mathcal{B}_d).$$

Examinons l'ensemble limite $\Sigma^{1/2}(\mathcal{B}_d)$. Soit \mathbf{z} un vecteur de \mathbb{R}^d appartenant à $\Sigma^{1/2}(\mathcal{B}_d)$. Nous avons,

$$\begin{aligned} \mathbf{z} &= \Sigma^{1/2}\mathbf{y} \quad \text{pour un certain } \mathbf{y} \in \mathcal{B}_d \\ \mathbf{y} &= \Sigma^{-1/2}\mathbf{z} \\ \mathbf{y}^T \mathbf{y} &= (\Sigma^{-1/2}\mathbf{z})^T \Sigma^{-1/2}\mathbf{z} \\ &= \mathbf{z}^T \Sigma^{-1} \mathbf{z} \\ &= \mathbf{z}^T \Sigma^{-1} \mathbf{z}, \end{aligned}$$

ce qui clôt la démonstration. □

A.5 Continuité des fonctions $r_\psi(\cdot)$, $m_\psi(\cdot)$ et $\sigma_\psi^2(\cdot)$

Soit

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

la densité conditionnelle de Y sachant $X = x$. On rappelle les hypothèses classiques sur la distribution du couple (X, Y) :

(F.1) $f_{X,Y}(\cdot, \cdot)$ est continue sur $J \times \mathbb{R}$;

(F.2) $f_X(\cdot)$ est continue et strictement positive sur J ;

(F.3) $Y\mathbb{I}\{X \in J\}$ est bornée;

Fixons $x \in J$. Soit $\{x_n : n \geq 1\}$ une suite de points dans J telle que $x_n \rightarrow x$, lorsque $n \rightarrow \infty$. Les hypothèses (F.1) et (F.2) entraînent

$$\lim_{n \rightarrow \infty} f_{Y|X}(y|x_n) \rightarrow f_{Y|X}(y|x), \quad \text{pour tout } y \in \mathbb{R}.$$

Le théorème de Lebesgue implique alors,

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} \{f_{Y|X}(y|x_n) - f_{Y|X}(y|x)\} \mathbb{I}\{0 \leq f_{Y|X}(y|x_n) \leq f_{Y|X}(y|x)\} dy = 0.$$

Via le théorème de Scheffé, nous obtenons plus précisément, lorsque $z \rightarrow x$,

$$D(z, x) := \int_{\mathbf{R}} |f_{Y|X}(y|z) - f_{Y|X}(y|x)| dy \rightarrow 0.$$

A présent, en utilisant le fait que $\psi(\cdot)$ est bornée sur les intervalles compacts combiné à (F.3), nous obtenons, pour $p \in \mathbb{N}$,

$$I_p(z, x) := \left| \int_{\mathbb{R}} \psi^p(y) \{f_{Y|X}(y|z) - f_{Y|X}(y|x)\} dy \right| \leq c^p D(z, x) \rightarrow 0,$$

où c dénote une constante finie. Cette dernière inégalité appliquée lorsque $p = 1, 2$ entraîne la continuité sur J des fonctions $r_\psi(\cdot)$, $m_\psi(\cdot)$ et $\sigma_\psi^2(\cdot)$.

A.6 Construction des noyaux d'ordre élevés

Dans cette section, nous présentons une méthode élégante permettant la construction de noyaux d'ordre l et plus généralement des noyaux d'ordre (s, l) , lorsque $(s, l) \in \mathbb{N}^2$ tels que $0 < s \leq l - 2$.

Définition A.6.1 Une fonction mesurable $L : \mathbb{R} \rightarrow \mathbb{R}$ est appelée **noyau d'ordre** (s, l) , si elle vérifie :

$$\int_{\mathbb{R}} x^j L(x) dx = \begin{cases} 0 & \text{pour } j \in \{0, \dots, s-1, s+1, \dots, l-1\}, \\ (-1)^{(s)} s! & \text{pour } j = s, \\ c_l \neq 0 & \text{pour } j = l. \end{cases} \quad (\text{A.24})$$

Nous rappelons qu'un noyau d'ordre (s, l) est approprié pour l'estimation de dérivées s -ièmes de fonctions appartenant à $C^l(\mathbb{R})$.

Lemme A.6.1 Si un noyau K est d'ordre l et s -fois différentiable, alors $K^{(s)}$ est un noyau d'ordre (s, l) .

On a

$$\int_{\mathbb{R}} K(u) du = 1 \Rightarrow \frac{(-1)^{(s)}}{s!} \int_{\mathbb{R}} u^s K^{(s)}(u) du = 1,$$

□

Pour illustrer l'utilité des noyaux d'ordre (s, l) , on considère l'exemple de l'estimation de la dérivée s -ième de la densité $f_X^{(s)}$,

$$\hat{f}_{X;n}^{(s)}(x) = \frac{1}{nh^{s+1}} \sum_{i=1}^n K^{(s)}\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}.$$

On suppose le noyau K à support compact. Le biais de l'estimateur à noyau de la dérivée d'ordre s de la fonction de densité s'écrit, après intégrations par parties,

$$\mathbb{E}[\hat{f}_{X;n}^{(s)}(x)] - f_X^{(s)}(x) = \int_{\mathbb{R}} h^{-s} K^{(s)}(u) f_X(x - hu) du - f_X^{(s)}(x)$$

$$= \int_{\mathbf{R}} \{f_X^{(s)}(x - hu) - f_X^{(s)}(x)\} K(u) du.$$

Si la densité $f_X(x)$ admet des dérivées jusqu'à l'ordre l bornées dans un voisinage du point x , via le développement de Taylor on obtient,

$$\mathbb{E}[\hat{f}_{X;n}^{(s)}(x)] = \sum_{j=1}^{l-1} h^{j-s} \frac{(-1)^j}{j!} f_X^{(j)}(x) \int u^j K^{(s)}(u) du + O(h^{l-s}). \quad (\text{A.25})$$

D'après (A.25) et (A.24), l'utilisation d'un noyau $K^{(s)}$ d'ordre (s, l) permet de réduire le biais de $\hat{f}_{X;n}^{(s)}(x)$ à l'ordre $O(h^{l-s})$.

Construction d'un noyau d'ordre (s, l)

On peut construire des noyaux d'ordre élevé en utilisant des polynômes par morceaux ([127], [102] et [54]) ou via une transformation de Fourier ([30], Hall and Marron (1987)). Dans cette section, on s'intéresse à la théorie développée par Berlinet ([8]), qui permet de classer les différents noyaux à partir d'une belle hiérarchisation fondée sur la théorie des espaces autoreproduisants.

En suivant l'article de Berlinet, il est possible de formuler une caractérisation des noyaux d'ordre élevés équivalente à (A.24). On pose $l = r + 1$. Tout au long de cette section, on dénote par V_r l'espace des polynômes de degré au plus r .

Définition A.6.2 Une fonction mesurable K est appelée noyau d'ordre $(s, r + 1)$ si et seulement si

$$\begin{cases} \forall P \in V_r, & \int_{\mathbf{R}} P(x) K(x) dx = P^{(s)}(0) \\ \text{et} & \int_{\mathbf{R}} x^{r+1} K(x) dx = c_{r+1} \neq 0. \end{cases}$$

En d'autres termes, si K est un noyau d'ordre $(s, r + 1)$, la forme linéaire sur V_r

$$P \rightarrow \int_{\mathbf{R}} P(x) K(x) dx \quad (\text{A.26})$$

est l'évaluation de $P^{(s)}$ au point zéro, d'après A.6.2. Ceci nous conduit à introduire la notion de sous-espace de Hilbert à noyau autoreproduisant de l'espace L^2 et plus particulièrement des espaces de polynômes car sur de tels espaces, les applications telles (A.26) ont une représentation agréable en termes de bases orthogonales. On utilisera comme dénomination leur abréviation anglo-saxonne RKHS correspondant à *Reproducing Kernel Hilbert Subspaces*. La construction d'une hiérarchie entre noyaux d'ordre élevé est établie à partir de la théorie des espaces autoreproduisants, via une succession de noyaux autoreproduisants appliqués à un noyau de base.

Soit K_0 dénotant une fonction de densité (notre noyau de base) et soit V le RKHS de $L^2(K_0)$. L'espace de fonctions V muni du produit scalaire $(\varphi, \psi) = \int \varphi(x)\psi(x)K_0(x)dx$

est un espace de Hilbert de fonctions à valeurs réelles et il existe une fonction $\mathcal{K}(x, y)$ (appelée le noyau autoreproduisant) telle que

$$\begin{cases} \forall x \in \mathbb{R}, \mathcal{K}(x, \cdot) \in V; \\ \forall \varphi \in V, \forall x \in \mathbb{R}, \int_{\mathbb{R}} \mathcal{K}(x, u)\varphi(u)K_0(u)du = \varphi(x). \end{cases}$$

La dernière égalité est la propriété de reproduction. L'existence du noyau autoreproduisant \mathcal{K} est équivalent à la continuité sur V de toutes les évaluations de la forme $f \rightarrow f(x)$. Si la suite $\{\varphi_i\}_{i \in I \subseteq \mathbb{N}}$ est une base orthonormale dans V , il s'ensuit la décomposition ou écriture suivante :

$$\forall x \in \mathbb{R}, \quad \mathcal{K}(x, \cdot) = \sum_{i \in I} \varphi_i(x)\varphi_i(\cdot).$$

Si K_0 a ses moments finis jusqu'à l'ordre $2r$, alors V_r , l'espace des polynômes de degré au plus r , est un RKHS de $L^2(K_0)$ comme n'importe quel sous-espace de fonctions de dimension finie. Soit $(P_i)_{0 \leq i \leq r}$ la suite des $(r+1)$ premiers polynômes orthonormaux dans $L^2(K_0)$. Pour $s \in \mathbb{N}$, nous posons

$$\mathcal{K}_r^{(s)}(x, y) = \sum_{i=0}^r P_i^{(s)}(y)P_i(x) = \sum_{i=s}^r P_i^{(s)}(y)P_i(x),$$

car les polynômes P_i sont exactement de degré i . La fonction $\mathcal{K}_r^{(s)}(x, y)$ représente la dérivation d'ordre s .

Lemme A.6.2 *Pour tout $\varphi \in L^2(K_0)$,*

$$\int_{\mathbb{R}} \varphi(x)\mathcal{K}_r^{(s)}(x, y)K_0(x)dx = \frac{d^s(\Pi_r(\varphi))}{dx^s}(y),$$

où Π_r dénote la projection de $L^2(K_0)$ sur V_r .

Soit $Q(x) = \sum_{i=0}^r \alpha_i P_i(x)$ un polynôme de degré au plus r . Nous avons

$$\int_{\mathbb{R}} \mathcal{K}_r^{(s)}(x, y)Q(x)K_0(x)dx = \sum_{i=0}^r \alpha_i P_i^{(s)}(y) = Q^{(s)}(y).$$

Il s'ensuit, pour $\varphi \in L^2(K_0)$,

$$\int_{\mathbb{R}} \mathcal{K}_r^{(s)}(x, y)\varphi(x)K_0(x)dx = \int_{\mathbb{R}} \mathcal{K}_r^{(s)}(x, y)\Pi_r(\varphi)(x)K_0(x)dx = \frac{d^s(\Pi_r(\varphi))}{dx^s}(y).$$

□

D'après (A.26) et le lemme A.6.2, le produit $\mathcal{K}_r^{(s)}(\cdot, 0)K_0(\cdot)$ désigne exactement un noyau d'ordre $(s, r+1)$ (confer également les théorèmes 73 et 78, p. 138 et 159, [9]).

Théorème A.6.1 Soient P un polynôme de degré au plus r , K_0 une densité dont les premiers moments sont finis jusqu'à l'ordre $(2r + 1)$ et \mathcal{K}_r le noyau autoreproduisant de l'espace V_r dans $L^2(K_0)$. Alors, $P(x)K_0(x)$ est un noyau d'ordre $(s, r + 1)$ si et seulement si

$$\begin{cases} \forall x \in \mathbb{R}, & P(x) = \mathcal{K}_r^{(s)}(x, 0) \\ \int_{\mathbb{R}} x^{r+1} P(x) K_0(x) dx = c_{r+1} \neq 0. \end{cases}$$

Soit $R(x) \in V_r$ un polynôme de degré au plus r qui admet donc une décomposition dans la base $\{1, x, x^2, \dots, x^r\}$. Nous avons

$$\int_{\mathbb{R}} R(x) P(x) K_0(x) dx = \int_{\mathbb{R}} R(x) \mathcal{K}_r^{(s)}(x, 0) K_0(x) dx = R^{(s)}(0).$$

□

D'après le théorème A.6.1, la hiérarchie de noyaux associée à K_0 est la famille de noyaux :

$$\mathcal{K}_r^{(s)}(x, 0) K_0(x) = \sum_{i=s}^r P_i^{(s)}(0) P_i(x) K_0(x), \quad (r, s) \in I^2, \quad r \geq s.$$

Chaque noyau $\mathcal{K}_r^{(s)}(x, 0) K_0(x)$ avec un moment d'ordre $(r + 1)$ fini et non nul est bien un noyau d'ordre $(s, r + 1)$. En somme, les noyaux peuvent être regroupés dans différentes hiérarchies possédants la propriété suivante : chaque classe ou hiérarchie de noyaux est identifiée par une fonction de densité (ou noyau de base) K_0 , qui appartient à cette classe, et contient des noyaux d'ordre 2, 3, 4, ..., produits de polynômes avec K_0 .

Cette méthodologie développée par Berlinet, (1990) [7], (1993) [8], nous permet de retrouver la hiérarchie des noyaux de MSE ou MISE (cf. [54]) asymptotiquement minimale. La famille des noyaux d'ordre $(s, r + 1)$ de AMISE minimale est identique à la hiérarchie associée au noyau d'Epanechnikov. On rappelle la définition du noyau d'Epanechnikov [44],

$$K^E(x) := \frac{3}{4}(1 - x^2)_+, \quad (\text{A.27})$$

et minimise la AMSE et la AMISE parmi tous les noyaux d'ordre 2, comme nous l'avons remarqué dans la section 1.4. Les noyaux K d'ordre $(s, r + 1)$ de support $[-1, 1]$ et de MISE minimale sont solutions du problème suivant :

$$\text{Minimiser } \begin{cases} T(K) =: \left\{ \int_{-1}^1 K^2(x) dx \right\}^{r+1-s} \left| \int_{-1}^1 x^{r+1} K(x) dx \right|^{2s+1}, \\ \text{avec } K \text{ tel que } \forall P \in V_r \quad \int_{-1}^1 P(x) K(x) dx = P^{(s)}(0). \end{cases} \quad (\text{A.28})$$

On obtient le théorème suivant :

Théorème A.6.2 Le polynôme solution de (A.28) et de support $(-1, 1)$ est donné par

$$K_r^{(s)}(x) = \sum_{i=s}^r P_i^{(s)}(0) P_i(x) \times \frac{3}{4}(1 - x^2)_+,$$

où les P_i sont les polynômes orthogonaux dans $L^2(K^E)$.

A.7 Remarque sur le terme de centrage

L'objectif de cette dernière section est de montrer que la différence entre l'espérance de $\hat{m}_{\psi;n}^{(k)}(x)$ et l'approximation $\tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)]$ est asymptotiquement négligeable. Nous avons,

$$\hat{m}_{\psi;n}^{(k)}(x) = \sum_{j=0}^k \binom{k}{j} \hat{r}_{\psi;n}^{(j)}(x) \{\hat{f}_{X;n}(x)^{-1}\}^{(k-j)},$$

d'après le développement de Leibniz.

Proposition A.7.1 *Lorsque Y est bornée et $nh^{1+2k} \rightarrow \infty$, nous obtenons,*

$$\mathbb{E}[\hat{m}_{\psi;n}^{(k)}(x)] = \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)] + O((nh^{1+2k})^{-1}). \quad (\text{A.29})$$

La démonstration est similaire à celle de la proposition 1.3.3. \square

Cette dernière proposition nous permet de remplacer, dans nos lois uniformes du logarithme, l'approximation $\tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)]$ par le terme exact de centrage. Il suffit de constater que, via (A.29),

$$\sup_{x \in I} \left\{ \mathbb{E}[\hat{m}_{\psi;n}^{(k)}(x)] - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(k)}(x)] \right\} = o\left(\left\{ \frac{\log(h_n^{-1})}{nh_n^{2k+1}} \right\}^{1/2}\right).$$

Il s'ensuit le théorème suivant, en reprenant les hypothèses et notations de la section 2.2.

Théorème A.7.1 *Supposons (F.1-3), (H.1-3), (K.1-4). Lorsque $n \rightarrow \infty$, nous avons,*

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_{\psi;n}^{(k)}(x) - \mathbb{E}[\hat{m}_{\psi;n}^{(k)}(x)] \} - \sigma_m(I) \right| \stackrel{\mathbf{P}}{=} o(1).$$

Supposons (F.1-3), (H.3-5), (K.1-4). Alors, nous obtenons, lorsque $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{m}_{\psi;n}^{(k)}(x) - \mathbb{E}[\hat{m}_{\psi;n}^{(k)}(x)] \} - \sigma_m(I) \right| \stackrel{p.s.}{=} o(1).$$

Bibliographie

- [1] Abramson, I. (1982). Arbitrariness of the pilot estimator in adaptive kernel methods. *J. Multivariate Analysis*, **12**, 562-567.
- [2] Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm for empirical processes. *Ann. Probab.*, **12**, 1041-1067.
- [3] Arcones, M. (2003). The large deviation principle of stochastic processes. I. *Theory Probab. Appl.*, **47**, 567-583.
- [4] Arcones, M. (2004). The large deviation principle of stochastic processes. II. *Theory Probab. Appl.*, **48**, 19-44.
- [5] Arcones, M. (2004). Convergence of the optimal M -estimator over a parametric family of M -estimators. A paraître dans *Test*. Preprint.
- [6] Bartlett, M. S. (1963). Statistical estimation of density functions. *Sankhyā*, Ser. A, **25**, 245-254.
- [7] Berline, A. (1990). Reproducing kernels and finite order kernels. In : Rousas, G. (ed.) *Nonparametric functional estimation and related topics*, p. 3-18. Kluwer, London New York.
- [8] Berline, A. (1993). Hierarchies of higher order kernels. *Proba. Theor. Relat. Fields*, **94**, 489-504.
- [9] Berline, A. et Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer.
- [10] Biau, G. (1999). Estimateurs à noyaux itérés : synthèse bibliographique. *Journal de la S.F.D.S.*, **140.1**.
- [11] Blondin, D. (2004). Estimation nonparamétrique multidimensionnelle des dérivées de la régression. *C. R. Acad. Sci. Paris*, **339**, 713-716.
- [12] Borovkov, A. (1987). *Statistique Mathématique*. Mir, Moscou.
- [13] Bosq, D. et Lecoutre, J. P. (1987). *Théorie de l'Estimation Fonctionnelle*. Economica, Paris.
- [14] Cheng, P.E. (1995). A note on strong convergence rates in nonparametric regression. *Stat. and Prob. letters*, **24**, 357-364.
- [15] Choi, E., Hall, P. et Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *Ann. Statist.*, **28.5**, 1339-1355.

- [16] Chu, C.-K. et Marron, J. S. (1991). Choosing a kernel regression estimator. *Statistical Science*, **6.4**, 404-436.
- [17] Collomb, G. (1977). Quelques propriétés de la méthode du noyau pour l'estimation non-paramétrique de la régression en un point fixé. *C. R. Acad. Sci. Paris*, **285 A**, 289-292.
- [18] Collomb, G. (1979). Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. *C. R. Acad. Sci. Paris*, **288**, 161-163.
- [19] Collomb, G. (1981). Estimation non-paramétrique de la régression : revue bibliographique. *Internat. Statist. Rev.*, **49**, 75-93.
- [20] Deheuvels, P. (1977). Estimation non-paramétrique de la densité par histogrammes généralisés. *Rev. Stat. Appl.*, **25**, 5-42.
- [21] Deheuvels, P. (1992). Functional laws of the iterated logarithm for large increments of empirical and quantile processes. *Stoch. Proc. and their Applications*, **43**, 133-163.
- [22] Deheuvels, P. (2000). Limit laws for kernel density estimators for kernels with unbounded supports. *Asymptotics in Statistics and Probability*. M. L. Puri (Ed.) 117-132. VSP. International Science Publishers, Amsterdam.
- [23] Deheuvels, P. (2000). Strong approximation of quantile processes by iterated Kiefer processes. *Ann. Probab.*, **28.2**, 909-945.
- [24] Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. In *Recent Advance in Reliability Theory : Methodology, Practice and Inference*, Birkhauser, Boston, p. 477-492.
- [25] Deheuvels P. et Einmahl J. H. J. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. *Ann. Probab.*, **28**, 1301-1335.
- [26] Deheuvels, P. et Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.*, **20**, 1248-1287.
- [27] Deheuvels, P. et Mason, D. M. (1994). Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann. Probab.*, **22**, 1619-1661.
- [28] Deheuvels, P. et Mason, D. M. (1995). Nonstandard local empirical processes indexed by sets. *J. Statist. Plan. Inf.*, **45**, 91-112.
- [29] Deheuvels, P. et Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Infer. Stoc. Processes*, **7.3**, 225-277.
- [30] Devroye, L. (1987). *A Course in Density Estimation*. Birkhauser, Boston.
- [31] Devroye, L. (1989). The double kernel method in density estimation. *Ann. Inst. Henri Poincaré*, **25.4**, 553-580.
- [32] Devroye, L. et Györfi, L. (1985). *Nonparametric Density Estimation : The L_1 view*. Wiley, New York.

-
- [33] Devroye, L. et Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer, New York.
- [34] Eggermont, P. P. B. et LaRiccia, V. N. (2001). *Maximum Penalized Likelihood Estimation*. Springer, New-York.
- [35] Einmahl, J. H. J. (1987). *Multivariate Empirical Processes. CWI Tract 32*. Centrum Wisk. Inform., Amsterdam.
- [36] Einmahl, U. (1993). Toward a general law of the iterated logarithm in Banach space. *Ann. Probab.*, **21.4**, 2012-2045.
- [37] Einmahl, U. (1995). On the cluster set problem for the generalized law of the iterated logarithm in Euclidean space. *Ann. Probab.*, **23.2**, 817-851.
- [38] Einmahl, U. et Kuelbs, J. (2001). Cluster sets for a generalized law of the iterated logarithm in Banach spaces. *Ann. Probab.*, **29.4**, 1451-1475.
- [39] Einmahl, U. et Mason, D.M. (1996). Some universal results on the behavior of increments of partial sums. *Ann. Probab.* **24.3** 1388-1407.
- [40] Einmahl, U. et Mason, D.M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory and Related Fields*, **107**, 283-311.
- [41] Einmahl, U. et Mason, D.M. (1998). Strong approximations to the local empirical process. Birkhäuser Verlag, Basel. *Progress in Probability*, **43**, 75-92.
- [42] Einmahl, U. et Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability*, **13.1**, 1-37.
- [43] Einmahl, U. et Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type functions estimators. *Ann. Statist.*, (à paraître).
- [44] Epanechnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.*, **14**, 153-158.
- [45] Fan, J. (1992). Design-adaptative nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998-1004.
- [46] Fan, J. (1992). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, **21**, 196-216.
- [47] Fan, F., Gasser, T., Gijbels, I., Brockmann, M., Engel, J. (1995). On nonparametric estimation via local polynomial regression. *Discussion Paper # 9511*, Institute of Statistics, Catholic University of Louvain, Louvain-la-Neuve, Belgium.
- [48] Fan, F. et Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting : variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society, Ser.B*, **57**, 371-394.
- [49] Fan, F. et Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Monographs on Statistics and Applied Probability, **66**. Chapman & Hall, London.

- [50] Finkelstein, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.*, **42**, 607-615.
- [51] Gaenssler, P. et Stute, W. (1979). Empirical processes. A survey of results for independent and identically distributed random variables. *Ann. Probab.*, **7.2**, 193-243.
- [52] Gasser, T. et Müller, H. G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, **757**, 23-68. Springer Verlag, Berlin.
- [53] Gasser, T. et Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171-185.
- [54] Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. B*, **47**, 238-252.
- [55] Giné, E. et Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data : Rates holding uniformly over adaptative intervals. *Ann. Inst. H. Poincaré Probab. Statist.*, **37**, 503-522.
- [56] Giné, E. et Guillou, A. (2002). Rates on strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, **38.6**, 907-921.
- [57] Giné, E., Mason, D. M. et Zaitsev, A. Yu. (2003). The L_1 -norm density estimators process. *Ann. Probab.*, **31.2**, 719-768.
- [58] Giné, E. et Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.*, **12**, 929-989.
- [59] Granovsky, B. L., Müller, H.-G. et Pfeifer, C. (1995). Some remarks on optimal kernel functions. *Statist. & Decisions* **13**, 101-116.
- [60] Hall, P. (1981). Laws of the iterated logarithm for nonparametric density estimators. *Z. Wahrsch. Verw. Gebiete*, **56**, 47-61.
- [61] Hall, P. (1984). Asymptotic properties of integrated square error and cross validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Gebiete*, **67**, 175-196.
- [62] Hall, P. (1991). On iterated logarithm laws for linear arrays and nonparametric regression estimators. *Ann. Prob.*, **19.2**, 740-757.
- [63] Hall, P. et Marron, J. S. (1987). Choice of kernel order in density estimation. *Ann. Statist.*, **12**, 766-774.
- [64] Hall, P. et Müller, H.-G. (2003). Order-preserving nonparametric regression with applications to conditional distribution and quantile function estimation. *J. Amer. Statist. Assoc.* **98** 598-608.
- [65] Härdle, W. (1984). Robust regression function estimation. *J. Multivariate Anal.*, **14**, 169-180.
- [66] Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.

-
- [67] Härdle, W. and Gasser, T. (1985). On robust kernel estimation of derivatives regression functions. *Scand. J. Statist.*, **12**, 233-240.
- [68] Härdle, W., Janssen, P. and Serfing, R. (1988). Strong uniform consistency rates of estimators of conditional functionals. *Ann. Statist.*, **16.4**, 1428-1449.
- [69] Härdle, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum. *J. Amer. Statist. Assoc.*, **83**, 86-95.
- [70] Härdle, W., Hall, P. and Marron, J. S. (1992). Regression smoothing parameters that are not far from their optimum. *J. Amer. Statist. Assoc.*, **87**, 227-233.
- [71] Härdle, W. et Kelly, G. (1987). Nonparametric kernel regression estimation - optimal choice of bandwidth. *Statistics*, **18.1**, 21-35.
- [72] Härdle, W. et Luckaus, S. (1984). Uniform consistency rates of a class of regression function estimators. *Ann. Statist.*, **12.2**, 612-623.
- [73] Härdle, W. et Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, **13.4**, 1465-1481.
- [74] Härdle, W. et Tsybakov, A. B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.*, **16.1**, 120-135.
- [75] Hastie, T. et Loader, C. (1993). Local regression : automatic kernel carpentry. *Statistical Science*, **8.2**, 120-143.
- [76] Hobson, E. W. (1927). *The Theory of Functions of a Real Variable and the Theory of Fourier Series*. vol. 1, 3rd. ed. Cambridge Univ. Press.
- [77] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Mathem. Statist.*, **42**, 1540-1552.
- [78] Huber, P. (1974). *Robust Statistics*. Wiley, New York.
- [79] Jain, M. C. et Marcus, M. B. (1978). Continuity of sub-Gaussian processes. Dekker, New York. *Advances in Probability*, **4**, 81-196.
- [80] Jennen-Steinmetz, C. et Gasser, T. (1988). A unifying approach to nonparametric regression estimation. *J. Amer. Statist. Assoc.*, **83**, No. 404, 1084-1088.
- [81] Johnston, G. (1979). *Smooth nonparametric regression analysis*. Inst. of Stat. Mimeo Series, **1253**, Ph.D. dissertation, Univ. of No. Carolina et Chapell Hill.
- [82] Johnston, G. (1982). Probabilities of maximal deviations for nonparametric regression function estimation. *J. Mult. Analysis*, **12**, 402-414.
- [83] Jones, M. C., Marron, J. S. et Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, **19**, 401-407.
- [84] Klass, M. (1976). Toward a universal law of the iterated logarithm, I. *Z. Wahrsch. Verw. Gebiete*, **36**, 165-178.
- [85] Komlós, I., Major, P., Tusnády, G. (1975). An approximation of partials sums of independent random variable and the sample distribution function. *Z. Wahrsch. Verw. Gebiete*, **32**, 111-131.

- [86] Konakov, V. D. et Piterbag, V. I. (1984). On the convergence rate of maximal deviation distribution for kernel regression estimates. *J. Mult. Analysis*, **15**, 279-294.
- [87] Krieger, A. M. et Pickands, J. (1981). Weak convergence and efficient density estimation at a point. *Ann. Statist.*, **9**, 1066-1078.
- [88] Ledoux, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM : Prob. Statist.*, **1**, 63-87.
- [89] Ledoux, M. et Talagrand, M. (1991). *Probability in Banach spaces : Isoperimetry and Processes*. Springer-Verlag, Berlin.
- [90] Lejeune, M. (1985). Estimation non-paramétrique par noyaux : régression polynomiale mobile. *Revue de Statist. Appliq.*, **33**, 43-68.
- [91] Lenze, B. (1990). On constructive one-sided approximation of multivariate functions of bounded variation. *Numer. Funct. Anal. Optim.*, **11**, 55-83.
- [92] Leung, D. H. Y. et Marriott, F. H. C. (1991). Finding extrema and zeros in nonparametric regression when the data contains outliers. *Nonparam. Statist.*, **1**, 69-82.
- [93] Linton, O. et Nielsen, J. P. (1994). A multiplicative bias reduction method for nonparametric regression. *Stat. & Probab. Letters* **19** 181-187.
- [94] Mack, Y. P. et Müller, H.-G. (1987). Adaptive nonparametric estimation of a multivariate regression function. *J. of Mult. Analysis*, **23**, 169-182.
- [95] Mack, Y. P. et Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405-415.
- [96] Mammen, E., et Marron, J. S. (1997). Mass centred kernel smoothers. *Biometrika*, **84**, 765-777.
- [97] Mammitzsch, V. (2001). On optimal standard kernels. *Statist. & Decisions*, **19**, 1-8.
- [98] Mason, D. M. (2003). A uniform functional law of the logarithm for a local empirical process. Birkhäuser Verlag Basel. *Progress in Probability*, **55**, 135-151.
- [99] Mason, D. M. (2004). A uniform functional law of the logarithm for the local empirical process. *Ann. Probab.*, **32.2**, 1391-1418.
- [100] Mason, D. M., Shorack, G. R. et Wellner, J. A. (1983). Strong limit theorems for the oscillation moduli of the uniform empirical process. *Z. Warsch. verw. Gebiete*, **65**, 83-97.
- [101] S. J. Montgomery-Smith (1993). Comparison of sums of independent identically distributed random vectors. *Probab. Math. Statist.*, **14**, 281-285.
- [102] Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Statist.*, **12**, 766-774.
- [103] Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.*, **82**, 231-238.
- [104] Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Lecture Notes in Statistics, **46**. Springer-Verlag, Berlin.

-
- [105] Müller, H.-G. (1997). Density adjusted kernel smoothers for random design nonparametric regression. *Statist. & Proba. Letters*, **19**, 181-187.
- [106] Müller, H.-G. et Prewitt K. A. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *J. Mult. Analysis*, **47**, 1-21.
- [107] Müller, H.-G. et Song, K.-S. (1993). Identity reproducing multivariate nonparametric regression. *J. of Mult. Analysis*, **46**, 237-253.
- [108] Nadaraya, E. A. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141-142.
- [109] Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer, Dordrecht.
- [110] Natanson, I. P. (1955). *Theory of Functions of a Real Variable*, **1**. Ungar, New York.
- [111] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065-1076.
- [112] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.
- [113] Priestley, M.B. et Chao, M.T. (1972). Nonparametric function fitting. *J. Royal Statist. Soc. B* **23** 395-436.
- [114] Rodriguez-Poo, J., Sperlich, S., Vieu, P. (2001). Normalité asymptotique d'estimateurs de maximum de vraisemblance pour modèles non-paramétriques de régression multidimensionnelle. *C. R. Acad. Sci. Paris* **333** Série 1, 61-64.
- [115] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, **27**, 470-472.
- [116] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832-837.
- [117] Ruppert, D., Sheather, S.J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257-1270.
- [118] Ruppert, D. et Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.*, **22**, 1346-1370.
- [119] Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *Annals of Mathematical Statistics*, **40.4**, 1187-1195.
- [120] Schuster, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of points. *Annals of Mathematical Statistics*, **43.1**, 84-88.
- [121] Schuster, E. F. et Yakowitz, S. (1979). Contributions to the theory of nonparametric regression with application to system identification. *Ann. Statist.*, **7**, 139-149.
- [122] Scott, D. W. (1992). *Multivariate Density Estimation - Theory, Practice and Visualization*. Wiley, New York.

- [123] Serfling, R. (1980). *Approximations theorems of mathematical statistics*. Wiley, New York.
- [124] Shorack, G. R. et Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.
- [125] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, **6**, 177-189.
- [126] Singh, R.S. (1977). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *Ann. Statist.*, **5**, 394-399.
- [127] Singh, R.S. (1979). Mean squared errors of estimates of a density and its derivatives. *Biometrika*, **66**, 177-180.
- [128] Staniswalis, J.G. (1989). The kernel estimate of a regression function in likelihood-based models. *J. Amer. Stat. Assoc.*, **84**, No. 405, 276-283.
- [129] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5.4**, 595-645.
- [130] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.
- [131] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Warsch. Verw. Gebiete*, **3**, 211-226.
- [132] Stute, W. (1982a). The oscillation behavior of empirical processes. *Ann. Probab.*, **10.1**, 86-107.
- [133] Stute, W. (1982b). A law of the iterated logarithm for kernel density estimators. *Ann. Probab.*, **10.2**, 414-422.
- [134] Stute, W. (1984). The oscillation behavior of empirical processes : the multivariate case. *Ann. Probab.*, **12.2**, 361-379.
- [135] Stute, W. (1986a). Conditional empirical processes. *Ann. Statist.*, **14**, 638-647.
- [136] Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.*, **14**, 891-901.
- [137] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, **22.1**, 28-76.
- [138] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.*, **126**, 505-563.
- [139] Tenreiro, C. (1997). Asymptotic normality of local polynomial estimators of regression function and its derivatives for time series. *Nonparametric Statistics*, **8**, 365-378.
- [140] Tibshirani, R. et Hastie, T. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.*, **82**, No. 398, 559-567.
- [141] Truong, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.*, **17.2**, 606-617.
- [142] Tsybakov, A. B. (2003). *Introduction à l'estimation non-paramétrique*. Mathématiques & Applications, **41**. Springer.

-
- [143] Tusnády, G. (1977). A remark on the approximation of the sample df in the multivariate case. *Period. Math. Hungar.* **8** 53-55.
- [144] Van de Geer, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge series in statistical and probabilistic mathematics.
- [145] Van der Vaart, A. W. et Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.
- [146] Van der Vaart, A. W. (1998). *Asymptotics Statistics*. Cambridge University Press, New York.
- [147] Wand, M. P. et Gutierrez, R. G. (1997). Exact risk approaches to smoothing parameter selection. *Nonparametric Statistics*, **8**, 337-354.
- [148] Wand, M. P. et Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- [149] Watson, G. S. (1964). Smooth regression analysis. *Sankhyà Ser. A*, **26**, 359-372.
- [150] Wu, T.-J. et Tsai, M.-H. (2004). Root n bandwidths selectors in multivariate density estimation. *Probab. Theory Relat. Fields*, **129**, 537-558.
- [151] Zhao, P.-L. (1994). Asymptotics of kernel estimators based on local maximum likelihood. *Nonparametric Statist.*, **4**, 79-90.