

RÉPUBLIQUE DU CAMEROUN

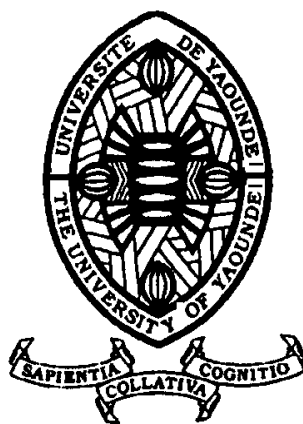
Paix-Travail-Patrie

UNIVERSITÉ DE YAOUNDÉ I

Facultés des Sciences

CENTRE DE RECHERCHE ET DE
FORMATION DOCTORALE EN
SCIENCES, TECHNOLOGIES ET
GEOSCIENCES

UNITE DE RECHERCHE ET DE
FORMATION DOCTORALE EN
MATHÉMATIQUES, INFORMATIQUE,
BIOINFORMATIQUE ET
APPLICATIONS



REPUBLIC OF CAMEROON

Peace-Work-Fatherland

UNIVERSITY OF YAOUNDE I

Faculty of Sciences

POSTGRADUATE SCHOOL FOR
SCIENCE, TECHNOLOGY AND
GEOSCIENCES

RESEARCH AND POSTGRADUATE
TRAINING UNIT FOR
MATHEMATICS, COMPUTER
SCIENCES AND APPLICATIONS

ESTIMATION STATISTIQUE DES MODÈLES DE PANEL NON LINÉAIRES À EFFETS INDIVIDUELS

Mémoire rédigé en vue de l'obtention du diplôme de **MASTER** en
Mathématiques

Par

ONANA ASSOUGA Florent

Matricule : 16Y554

Licence en Mathématiques

Sous la direction de :

Pr TAKAM SOH Patrice

Maitre de Conférences

Univesité de Yaoundé I

Année Académique : 2020-2021

ESTIMATION STATISTIQUE DES MODÈLES DE PANEL NON LINÉAIRES À EFFETS INDIVIDUELS

Mémoire rédigé et soutenu publiquement en vue de l'obtention
du Diplôme de MASTER en Mathématiques

Par

ONANA ASSOUGA Florent

Licence en Mathématiques

Matricule

16Y554

Sous l'encadrement de

Pr TAKAM SOH Patrice

Maitre de Conférences

Année académique : 2020-2021

Dédicace

Ce mémoire est dédié à ma grand-mère :

NDZIE ATANGANA Françoise

Remerciements

*P*arvenu au terme de ce travail, je remercie en premier l'Eternel Dieu tout puissant et lui rends grâce pour toute sa miséricorde et ses grâces pour m'avoir donné santé, courage et audace pour surpasser les difficultés.

*Q*u'il me soit permis d'exprimer ma gratitude à mon directeur de mémoire :

Pr TAKAM SOH Patrice pour avoir accepté de diriger ce travail, pour son soutien multiforme, pour sa disponibilité et son encadrement constant dans nos premiers pas de recherche.

*J*e remercie les enseignants de l'Université de Yaoundé I, en particulier ceux du département de Mathématiques et du laboratoire Probabilité-Statistique-Finance, pour leurs enseignements et leur accompagnement tout au long de notre formation académique et sociale.

*C*e travail est avant tout le fruit d'énormes sacrifices familiaux et je veux adresser des remerciements spéciaux à la famille et particulièrement à :

♡ Ma grand mère NDZIE ATANGANA Françoise pour l'éducation qu'elle a su nous donner, pour son amour sans faille et son support inconditionnel.

♡ Mes oncles MINFEGUE ASSOUGA Calvin, ONGUENE ASSOUGA Hilaire pour l'abnégation dont ils ont fait preuve pour nous garantir une scolarisation allant jusqu'à faire des sacrifices leur portant préjudice sur leurs propres objectifs ; aussi à OKOME OUMAROU Dominique pour le rôle de mère qu'elle a toujours su jouer à notre endroit.

♡ Mes frères et soeurs : ASSOUGA ELEBE Roland, NDZIE Emmanuel, FEUMENI Brayan, BELINGA ASSOUGA Armand, Rachelle, Marina NOUGA, Kengne Anges.

♡ Mes oncles et tantes : NGA ASSOUGA Marie Sylvie, NDZIE ASSOUGA Mirice, Marie Thérèse NOUGA, l'abbé Simon Pierre BITEGA.

♡ Mes amis : TCHOUPA Kareine, NYEMECK Stephane, SOBJIO Ludane, TSOGMO Cybèle, ABOULKADIRI SALI, Stéphane NGASSA, MENONO Landry, NYAMBONGO Arlette, Nicaise NDEL, Cabrel NDONTOSOP, FEZEU Clinton, FAISSAL MOUNCHINGAM, Ange MAFFO, Lidwine KEUFACK, TEDONMEGNE Landry.

Je dis merci de tout cœur à tous ceux qui, bien qu'ayant contribué à la réalisation de ce mémoire, ne sont pas mentionnés nommément ici.

Déclaration sur l'honneur

Le présent document est une œuvre originale du candidat et n'a été soumis nulle part ailleurs en partie ou en totalité, pour une autre évaluation. Les contributions externes ont été dûment mentionnées et recensées en bibliographie

Signature du candidat

ONANA ASSOUGA Florent

Résumé

Nous voulons proposer un modèle de régression liant une variable d'intérêt Y et un vecteur de covariables Z . Pour cela, nous nous proposons d'analyser un modèle non linéaire du type $Y = g(Z) + \varepsilon$ (le lien entre Y et Z étant inconnu). Nous nous intéressons à l'estimation de ce modèle en disposant des données longitudinales de la forme $\{Z_{it}\}_{i=1,t=1}^{N,T}$ où N représente le nombre d'individus et T le nombre de dates d'observation. Notre objectif est de proposer une estimation de $g(\cdot)$ et d'analyser ses propriétés (consistance, convergence uniforme presque sûre, normalité asymptotique) en prenant en compte l'effet individuel, soit $Y_{it} = g(Z_{it}) + \mu_i + \nu_{it}$, et de proposer un test non paramétrique sur la forme de la fonction de lien $g(\cdot)$.

Mots clés : modèles de panel non linéaires, effet individuel, convergence uniforme presque sûre, test non paramétrique

Abstract

We want to propose a regression model between one interest variable Y and one vector of covariables Z . For this, we will analyze the nonlinear regression of the form $Y = g(Z) + \varepsilon$. We are interest to the estimation of this model when we get longitudinal data of the form $\{Z_{it}\}_{i=1,t=1}^{N,T}$ where N represents the number of individuals and T the number of dates observation. Our aim is to propose the estimation of $g(\cdot)$ and to analyze its properties (consistancy, almost surely uniform convergence, asymptotic normality) by taking in count the individual effects, thus $Y_{it} = g(Z_{it}) + \mu_i + v_{it}$, and to propose a nonparameric test on the form of the link functon $g(\cdot)$.

Keywords : nonlinear panel data models, individual effects, almost surely uniform convergence, nonparametric test

Table des matières

Dédicaces	i
Remerciements	ii
Déclaration sur l'honneur	iv
Résumé	v
Abstract	vi
Table des figures	ix
Liste des tableaux	x
INTRODUCTION ET MOTIVATION	1
1 ESTIMATION NON PARAMÉTRIQUE PAR LA MÉTHODE DES NOYAUX	3
1.1 Estimation de la fonction de densité de probabilité d'une variable aléatoire unidimensionnelle	3
1.1.1 Construction de l'estimateur	3
1.1.2 Consistance de l'estimateur	6
1.2 Estimation de la fonction de densité de probabilité d'une variable aléatoire multidimensionnelle	9
1.2.1 Construction de l'Estimateur	9
1.2.2 Consistance et Normalité asymptotique de l'estimateur	10
1.2.3 Convergence uniforme presque sûre de l'estimateur	19
2 ESTIMATION DANS LES MODÈLES DE PANEL NON-LINÉAIRES	26
2.1 Structure de l'erreur	27

2.2	Construction de l'estimateur de $g(z)$ par la méthode des noyaux	28
2.3	Consistance de l'estimateur	31
2.4	Convergence uniforme presque sûre et normalité asymptotique de l'estimateur	35
2.4.1	Normalité asymptotique	35
2.4.2	Convergence uniforme presque sûre	37
2.5	Intervalle de confiance asymptotique de l'Estimateur $\hat{g}(z)$	39
2.6	Test de linéarité	40
3	APPLICATION ET SIMULATIONS	44
3.1	Simulation des fonctions de densités de probabilité	44
3.1.1	Une hypothèse pas valide	44
3.1.2	Une vérification	46
3.2	Simulations de Monté-Carlo sur un panel	47
3.3	Test de linéarité	50
	Conclusion et perspectives	51
	Bibliographie	51
	Annexe	54
A	Rappels	55
A.1	Inégalités et Modes de convergence	55
A.1.1	Inégalités	55
A.1.2	Modes de convergence	59
A.2	Loi faible des grands nombres et Théorème central limite	64
A.3	Estimation dans les modèles de régression linéaire : Estimateur des moindres carrés ordinaires (MCO) et Estimateur des moindres carrés généralisés (MCG)	66
A.4	Rappel d'algèbre	68
A.4.1	Projection	69
A.4.2	Produit de Kronecker des matrices	70
A.5	Codes R	70
B	Notations et Symboles	73

Table des figures

1.1	Les courbes de quelques fonctions noyaux usuelles	6
1.2	Tableau de variations de la fonction ϕ	22
3.1	Graphe de la densité f_1	45
3.2	Histogramme des données X	45
3.3	Le graphe des différentes densités estimées des données X	46
3.4	Le graphe de densité de la loi exponentielle de paramètre 0.5	47
3.5	Le graphe de densité des données Y	47
3.6	Le graphe de la fonction de lien exacte $g(z) = \sin 2z$	48
3.7	$N = 50, T = 3$	49
3.8	$N = 100, T = 3$	49
3.9	$N = 200, T = 3$	49

Liste des tableaux

3.1	Table des fréquences des données X	45
3.2	Valeurs des écarts types et paramètres de lissage	48
3.3	Estimation de la puissance du test J_N	50

Introduction et Motivation

Nous nous intéressons à la modélisation d'une variable d'intérêt Y (qui peut représenter par exemple le taux de production du cacao au Cameroun) en fonction d'un vecteur de covariables Z (qui peut représenter les variables climatiques). Nous nous plaçons dans le cas le plus réaliste où à partir des observations de Y et Z il n'est pas possible de suggérer un lien linéaire. Notre modèle d'intérêt est alors de la forme $Y = g(Z) + \varepsilon$ où ε représente l'erreur de mesure. Pour estimer ce modèle, nous nous plaçons dans le cas où l'on dispose des données longitudinales de la forme Y_{it}, Z_{it} $i = 1, \dots, N$ (N représentant le nombre d'individus), $t = 1, \dots, T$ (T représentant le nombre de dates d'observations) : ce sont les données de panel. Un panel est donc un échantillon formé à partir d'observations sur N sections d'unités pendant T -périodes consécutives pour donner lieu à un échantillon de la forme $\{X_{it}\}_{i=1, t=1}^{N, T}$ ou tout simplement, un panel est une répétition d'observations dans le temps pour les mêmes individus. Les individus ici peuvent être des travailleurs, les maisons, les firmes, les industries, les régions ou les pays, pour ne citer que ceux là. L'utilisation de ce type de données devient de plus en plus commun dans les champs de la médecine, de l'économétrie moderne. Ainsi un modèle de panel est un modèle de régression sur les données de panels de la forme

$$Y_{it} = g(Z_{it}) + u_{it}, \quad (1)$$

où l'on cherche à expliquer la variable Y à partir du vecteur de covariables Z sous la marge d'erreur u , donc la fonction $g(\cdot)$ est la fonction de lien explicative.

Ce type de modèle (modèle de panel ou panel data) est bien connu dans la littérature. En fonction de la nature de l'erreur et des hypothèses que l'on fait sur la fonction $g(\cdot)$, on obtient une classification des modèles de panels :

- Suivant la nature de l'erreur (voir [Baltagi 2021 \(4\)](#)) : selon que l'erreur se décompose par
 - $u_{it} = v_{it}$, on parle d'erreur simple
 - $u_{it} = \mu_i + v_{it}$ on parle d'erreur à structure one-way ou à effets individuels

- $u_{it} = \lambda_t + \mu_i + v_{it}$ on parle d'erreur à structure two-way où à effets mixtes
- Suivant la nature de la fonction $g(\cdot)$:
 - Si $g(x) = \beta x$ (c'est à dire linéaire), on parle de modèle de panel linéaire, et l'estimation dans le modèle passe par l'estimation du paramètre β .
 - Si la fonction $g(\cdot)$ est sémi-linéaire, c'est à dire de la forme $g(x, z) = \beta x + \theta(z)$ où $\theta(\cdot)$ est une fonction non linéaire, on parle de modèle de panel sémi-linéaire et on se retrouve dans l'estimation sémi-paramétrique.
 - Si il n'y a pas d'hypothèse de linéarité quelconque sur la fonction $g(\cdot)$, on parle de modèle de panel non linéaire et l'estimation dans ce cas devient non paramétrique

Rappelons également que des extensions de tels modèles ont été étudiées en prenant en compte l'aspect dynamique : $Y_{it} = \rho Y_{i,t-1} + u_{it}$. L'estimation non-paramétrique dans les modèles de panel devient de plus en plus populaire dans la littérature et de nombreuses applications ont été faites dans ces modèles à l'exemple de : [Engle et al \(1986\)](#) qui ont utilisé le modèle de panel sémi-linéaire pour estimer l'effet de la température z sur la demande d'électricité y ; [Stock \(1989\)](#) qui ont étudié l'effet de la proximité des déchets toxiques sur le prix des maisons à location; [Anglin and Gencay \(1996\)](#) qui a étudié l'effet de la qualité des maisons sur leurs prix de location, [Newey et al \(1990\)](#), [Melenberg and Soest \(1993\)](#), [Powell \(1993\)](#), [Mundra \(1997\)](#) [kniesner \(1994\)](#) ... Là sont quelques applications et motivations qui illustrent l'utilité de ces modèles.

Dans notre travail, nous voulons à partir des panel data estimer la fonction $g(\cdot)$ de façon non paramétrique en prenant en compte l'effet individuel sur les données soit

$$Y_{it} = g(Z_{it}) + \mu_i + v_{it}. \quad (2)$$

Dans un premier temps, nous allons nous intéresser à l'estimation non paramétrique par la méthode des noyaux d'une densité de probabilité dans le cas multidimensionnel et nous donnerons les propriétés statistiques en occurrence consistance, biais, variance, convergence uniforme presque sûre. Dans un deuxième temps nous estimerons la fonction $g(\cdot)$ de façon non paramétrique en nous inspirant de [Nadaraya-Watson \(17\)](#) et nous prouverons la consistance de l'estimateur obtenu (convergence en probabilité et convergence uniforme presque-sûre), nous ressortirons quelques propriétés de l'estimateur, notamment la normalité asymptotique et proposerons un intervalle de confiance ainsi qu'un test statistique sur la forme de la fonction $g(\cdot)$. Et enfin, nous appliquerons les résultats sur des données simulées.

ESTIMATION NON PARAMÉTRIQUE PAR LA MÉTHODE DES NOYAUX

Dans ce chapitre nous présentons la méthode d'estimation non paramétrique par la méthode des noyaux : notamment les fonctions de densité de probabilité. Nous démontrerons les propriétés de consistance, la normalité asymptotique et la convergence uniforme presque sûre de l'estimateur d'une fonction densité de probabilité par cette méthode.

Comment estimer la fonction de densité de probabilité $f_X(x) \equiv f(x)$ d'une variable aléatoire X continue à partir d'une suite d'observations indépendantes et identiquement distribuées i.i.d X_1, \dots, X_n ?

1.1 Estimation de la fonction de densité de probabilité d'une variable aléatoire unidimensionnelle

On suppose premièrement que la variable aléatoire X est réelle.

1.1.1 Construction de l'estimateur

Pour $x \in \mathbb{R}$, la fonction de répartition au point x , $F_X(x)$, de la variable aléatoire X est définie par :

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{E}(\mathbb{1}_{\{X \leq x\}}). \quad (1.1)$$

Et l'on peut estimer $F_X(x)$ à partir des observations X_1, \dots, X_n par la méthode des moments et obtenir :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}. \quad (1.2)$$

Ainsi, par la définition de la fonction de densité de probabilité $f_X(x)$ de la variable aléatoire X donnée par :

$$f_X(x) = \frac{d}{dx} F_X(x), \quad (1.3)$$

on déduit des égalités (1.2) et (1.3) qu'un estimateur de $f_X(x)$, connu sous le nom d'**Estimateur de Parzen-Rosenblatt(1962-1956)** est donné par :

$$f_X(x) \approx \frac{F_n(x+h) - F_n(x-h)}{2h}, \quad (1.4)$$

où h est un petit incrément positif appelé **paramètre de lissage**.

En effet en écrivant le développement de Taylor de la fonction de répartition $F_X(x)$ à l'ordre 1 aux points $x+h$ et $x-h$, on a

$$F_X(x+h) = F_X(x) + hf_X(x) + o(h), \quad (1.5)$$

et

$$F_X(x-h) = F_X(x) - hf_X(x) + o(h); \quad (1.6)$$

en retranchant l'égalité (1.6) à l'égalité (1.5) on obtient l'approximation de l'égalité (1.4).

En remplaçant l'égalité (1.2) dans l'égalité (1.4), on obtient :

$$\hat{f}_X(x) = \frac{1}{2nh} \sum_{i=1}^n \mathbb{1}_{\{x-h < X_i \leq x+h\}}. \quad (1.7)$$

Si l'on définit la fonction noyau uniforme par :

$$k_1(x) = \begin{cases} \frac{1}{2} & \text{si } -1 < x \leq 1 \\ 0 & \text{sinon} \end{cases}, \quad (1.8)$$

alors l'expression de $\hat{f}_X(x)$ donnée par (1.7) s'écrit encore :

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n k_1\left(\frac{X_i - x}{h}\right). \quad (1.9)$$

L'estimateur qu'on vient d'obtenir s'appelle **l'estimateur à noyau uniforme** car on a utilisé la fonction noyau uniforme donnée par l'égalité (1.8).

De façon générale, les estimateurs à noyaux sont une généralisation de l'estimateur obtenu à la formule (1.9). L'une des motivations est d'obtenir un estimateur assez lisse de la fonction de densité en remplaçant la fonction indicatrice k_1 par des fonctions continues.

Définition 1 (Fonction noyau d'ordre 2)

On appelle fonction noyau d'ordre 2, toute fonction positive bornée $k(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ intégrable et de carré intégrable satisfaisant :

$$\begin{aligned} (i) \quad & \int_{\mathbb{R}} k(u) du = 1 \\ (ii) \quad & k(u) = k(-u) \\ (iii) \quad & \int_{\mathbb{R}} u^2 k(u) du = \kappa_2 > 0 \end{aligned} \tag{1.10}$$

Remarque 1.1.1. La condition de symétrie (ii) implique que : $\int_{\mathbb{R}} uk(u) du = 0$, qui est une propriété propre aux fonctions noyaux.

Remarque 1.1.2. L'on peut définir de façon similaire une fonction noyau d'ordre supérieur $l \geq 2$ (avec $l \in \mathbb{N}$), cette fois la fonction $k(\cdot)$ satisfait :

$$\begin{aligned} (i) \quad & \int_{\mathbb{R}} k(u) du = 1 \\ (ii) \quad & \int_{\mathbb{R}} u^v k(u) du = 0, \quad (v = 1, \dots, l-1) \\ (iii) \quad & \int_{\mathbb{R}} u^l k(u) du = \kappa_l \neq 0 \end{aligned} \tag{1.11}$$

Les exemples de fonctions noyaux les plus utilisées sont :

- La fonction noyau uniforme : $k(z) = \frac{1}{2} \mathbb{1}_{\{-1 \leq z \leq 1\}}$
- La fonction noyau gaussienne : $k(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$
- La fonction noyau triangulaire : $k(z) = (1 - |z|) \mathbb{1}_{\{-1 \leq z \leq 1\}}$
- La fonction noyau d'EpanechniKov : $k(z) = \frac{3}{4\sqrt{5}} (1 - \frac{z^2}{5}) \mathbb{1}_{\{-\sqrt{5} \leq z \leq \sqrt{5}\}}$
- La fonction noyau cosinus : $k(z) = \frac{\pi}{4} \cos(\frac{\pi z}{2}) \mathbb{1}_{\{-1 \leq z \leq 1\}}$
- La fonction noyau Biweight : $k(z) = \frac{15}{16} (1 - z^2)^2 \mathbb{1}_{\{-1 \leq z \leq 1\}}$
- La fonction noyau Triweight : $k(z) = \frac{35}{32} (1 - z^2)^3 \mathbb{1}_{\{-1 \leq z \leq 1\}}$
- La fonction noyau de Laplace : $k(z) = \frac{1}{2} \exp(-|z|)$
- La fonction noyau Gaussienne tronquée : $k(z) = \frac{\phi(z)}{\int_{-1}^1 \phi(t) dt} \mathbb{1}_{\{-1 \leq z \leq 1\}}$ (où ϕ est la densité de la loi normale centrée réduite)

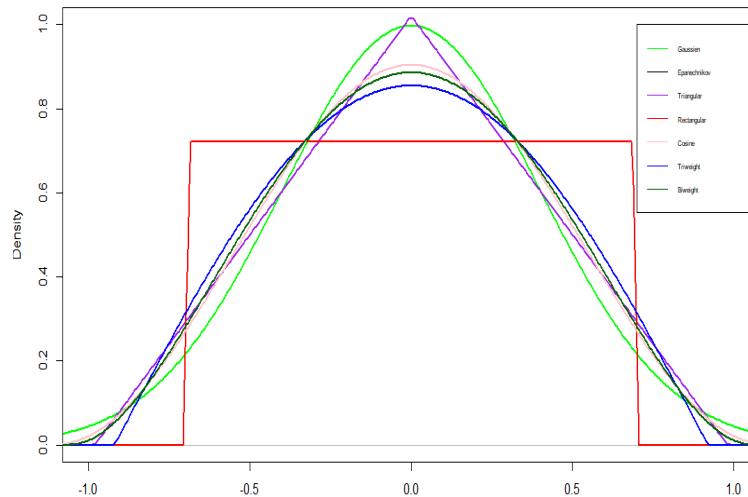


FIGURE 1.1 – Les courbes de quelques fonctions noyaux usuelles

Remarque 1.1.3. Dans toute la suite de notre mémoire, nous travaillons avec les fonctions noyaux d'ordre 2 et la fonction de densité de probabilité $f_X(\cdot)$ de la variable X sera simplement notée $f(\cdot)$.

Définition 2

L'estimateur à noyau $\hat{f}(x)$ de la fonction de densité de probabilité $f(\cdot)$ au point x , associé à la fonction noyau $k(\cdot)$ satisfaisant (1.10) est définie par :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{X_i - x}{h}\right). \quad (1.12)$$

Remarque 1.1.4. L'estimateur à noyau dépend donc du choix de la fonction noyau $k(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ et du paramètre de lissage $h > 0$.

1.1.2 Consistance de l'estimateur

Hypothèse 1.1.1. On suppose que X_1, \dots, X_n est une suite d'observations i.i.d ayant une fonction de densité de probabilité $f(x)$ trois fois différentiable.

Le Théorème suivant donne la consistance de l'estimateur à noyau $\hat{f}(x)$ de la densité $f(x)$.

Théorème 1

Soit x un point à l'intérieur du support de X et soit $\hat{f}(x)$ l'approximation définie par (1.12). On suppose que la fonction noyau $k(\cdot)$ vérifie les conditions (1.10), et est bornée. On suppose en outre que $nh \rightarrow \infty$ lorsque $n \rightarrow \infty$ et $h \rightarrow 0$; alors :

$$\begin{aligned} MSE(\hat{f}(x)) &= \frac{h^4}{4} [\kappa_2 f^{(2)}(x)]^2 + \frac{\kappa f(x)}{nh} + O(h^4 + (nh)^{-1}) \\ &= O(h^4 + (nh)^{-1}), \end{aligned} \quad (1.13)$$

où $\kappa_2 = \int_{\mathbb{R}} v^2 k(v) dv$ et $\kappa = \int_{\mathbb{R}} k^2(v) dv$.

Preuve. Par définition,

$$\begin{aligned} MSE(\hat{f}(x)) &\equiv \mathbb{E} \left\{ [\hat{f}(x) - f(x)]^2 \right\} \\ &\equiv \text{Var}(\hat{f}(x)) + [\text{biais}(\hat{f}(x))]^2. \end{aligned}$$

Nous évaluerons le biais($\hat{f}(x)$) et la $\text{Var}(\hat{f}(x))$ séparément.

On rappelle d'abord que si $g(x)$ est une fonction définie sur \mathbb{R} m -fois différentiable, la formule de Taylor s'écrit :

$$\begin{aligned} g(x) &= g(x_0) + \sum_{k=1}^{m-1} \frac{g^{(k)}(x_0)}{k!} (x - x_0)^k + \frac{g^{(m)}(\xi)}{m!} (x - x_0)^m \\ &= g(x_0) + \sum_{k=1}^{m-1} \frac{g^{(k)}(x_0)}{k!} (x - x_0)^k + o((x - x_0)^{m-1}) \\ &= g(x_0) + \sum_{k=1}^{m-1} \frac{g^{(k)}(x_0)}{k!} (x - x_0)^k + O((x - x_0)^m), \end{aligned}$$

où $\xi \in (x_0, x)$. On a d'une part

$$\begin{aligned} \text{biais}(\hat{f}(x)) &= \mathbb{E} \left\{ \frac{1}{nh} \sum_{i=1}^n k\left(\frac{X_i - x}{h}\right) \right\} - f(x) \\ &= \frac{1}{h} \mathbb{E} \left[k\left(\frac{X_1 - x}{h}\right) \right] - f(x) \text{ car les v.a sont i.i.d} \\ &= \frac{1}{h} \int_{\mathbb{R}} f(x_1) k\left(\frac{x_1 - x}{h}\right) dx_1 - f(x). \end{aligned}$$

En effectuant le changement de variables $v = \frac{x_1 - x}{h}$, on obtient

$$\text{biais}(\hat{f}(x)) = \frac{1}{h} \int_{\mathbb{R}} f(x + hv)k(v)h dv - f(x) = \int_{\mathbb{R}} f(x + hv)k(v)dv - f(x).$$

On écrit donc le développement de Taylor de f au voisinage de x à l'ordre 2 et on a :

$$\begin{aligned} \text{biais}(\hat{f}(x)) &= \int_{\mathbb{R}} \left\{ f(x) + f^{(1)}(x)hv + \frac{1}{2}f^{(2)}(x)h^2v^2 + \mathcal{O}(h^3) \right\} k(v)dv - f(x) \\ &= f(x) \underbrace{\int_{\mathbb{R}} k(v)dv}_{=1} + hf^{(1)}(x) \underbrace{\int_{\mathbb{R}} vk(v)dv}_{=0} + \frac{h^2}{2}f^{(2)}(x) \underbrace{\int_{\mathbb{R}} v^2k(v)dv}_{=\kappa_2} + \mathcal{O}(h^3) - f(x) \\ &= \frac{h^2}{2}f^{(2)}(x)\kappa_2 + \mathcal{O}(h^3) = \frac{h^2}{2}f^{(2)}(x)\kappa_2 + \mathcal{O}(h^2). \end{aligned} \quad (1.14)$$

Donc

$$[\text{biais}(\hat{f}(x))]^2 = \frac{h^4}{4} [\kappa_2 f^{(2)}(x)]^2 + \mathcal{O}(h^4). \quad (1.15)$$

Et d'autre part

$$\begin{aligned} \mathbb{V}ar(\hat{f}(x)) &= \mathbb{V}ar \left[\frac{1}{nh} \sum_{i=1}^n k\left(\frac{X_i - x}{h}\right) \right] \\ &= \frac{1}{n^2 h^2} \left\{ \sum_{i=1}^n \mathbb{V}ar \left[k\left(\frac{X_i - x}{h}\right) \right] \right\} \quad \text{car les v.a sont indépendantes} \\ &= \frac{1}{nh^2} \mathbb{V}ar \left(k\left(\frac{X_1 - x}{h}\right) \right) \quad \text{car les v.a sont identiquement distribuées} \\ &= \frac{1}{nh^2} \left\{ \mathbb{E} \left[k^2\left(\frac{X_1 - x}{h}\right) \right] - \left[\mathbb{E} \left(k\left(\frac{X_1 - x}{h}\right) \right) \right]^2 \right\} \\ &= \frac{1}{nh^2} \left\{ \int_{\mathbb{R}} f(x_1)k^2\left(\frac{x_1 - x}{h}\right)dx_1 - \left[\int_{\mathbb{R}} f(x_1)k\left(\frac{x_1 - x}{h}\right)dx_1 \right]^2 \right\}. \end{aligned}$$

En effectuant à nouveau le changement de variable $v = \frac{x_1 - x}{h}$, et en écrivant le développement Taylor de f au voisinage de $x \in \mathbb{R}$ à l'ordre 1, on obtient :

$$\begin{aligned} \mathbb{V}ar(\hat{f}_X(x)) &= \frac{1}{nh^2} \left\{ h \int_{\mathbb{R}} f(x + hv)k^2(v)dv - \left[h \int_{\mathbb{R}} f(x + hv)k(v)dv \right]^2 \right\} \\ &= \frac{1}{nh^2} \left\{ h \int_{\mathbb{R}} [f(x) + f^{(1)}(x)hv]k^2(v)dv - \mathcal{O}(h^2) \right\} \\ &= \frac{1}{nh} \left\{ f(x) \int_{\mathbb{R}} k^2(v)dv + \mathcal{O} \left(h \int_{\mathbb{R}} |v|k^2(v)dv \right) - \mathcal{O}(h^2) \right\} \\ &= \frac{1}{nh} \{ \kappa f(x) + \mathcal{O}(h) \} \end{aligned} \quad (1.16)$$

Les égalités (1.15) et (1.16) montrent donc que

$$MSE(\hat{f}(x)) = \frac{h^4}{4} [\kappa_2 f^{(2)}(x)]^2 + \frac{\kappa f_X(x)}{nh} + O(h^4 + (nh)^{-1}).$$

■

Corollaire 1.1. Lorsque $n \rightarrow \infty$, $h \rightarrow 0$, et $nh \rightarrow 0$, on a :

$$\hat{f}(x) - f(x) = O_p(h^2 + (nh)^{-1/2}) = o_p(1). \quad (1.17)$$

Preuve. En effet, on a l'égalité asymptotique $MSE(\hat{f}(x)) = O(h^4 + (nh)^{-1})$, ou encore $MSE(\hat{f}(x)) = \mathbb{E}\{|\hat{f}(x) - f(x)|^2\} = O(h^4 + (nh)^{-1})$.

On déduit donc du Théorème 3 de l'Annexe que

$$\hat{f}(x) - f(x) = O_p((h^4 + (nh)^{-1})^{1/2}) = O_p(h^2 + (nh)^{-1/2}).$$

Et comme $h^2 + (nh)^{-1/2} \rightarrow 0$, alors $O_p(h^2 + (nh)^{-1/2}) = O_p(o(1)) = o_p(1)$; d'où le résultat.

■

Remarque 1.1.5. Le corollaire précédent prouve la convergence en probabilité de $\hat{f}(x)$ vers $f(x)$, plus précisément traduite par l'égalité $\hat{f}(x) - f(x) = o_p(1)$.

1.2 Estimation de la fonction de densité de probabilité d'une variable aléatoire multidimensionnelle

Supposons que X_1, \dots, X_n constitue une suite i.i.d de variables q -dimensionnelles ($X_i \in \mathbb{R}^q$ avec $q > 1$) ayant une même fonction de densité $f(x) = f(x_1, \dots, x_q)$. Notons par X_{is} la s -ième composante du vecteur X_i ($s = 1, 2, \dots, q$).

1.2.1 Construction de l'Estimateur

L'estimateur se construit à partir des fonctions noyaux vues précédemment et la construction se calque à la construction de l'estimateur dans le cas unidimensionnel.

Définition 3

En utilisant le produit des fonctions noyaux construit à partir des fonctions noyaux unidimensionnelles, l'on peut estimer $f(x)$ par :

$$\hat{f}(x) = \frac{1}{nh_1 \cdots h_q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right). \quad (1.18)$$

où

$$K\left(\frac{X_i - x}{h}\right) = k\left(\frac{X_{i1} - x_1}{h_1}\right) \times \cdots \times k\left(\frac{X_{iq} - x_q}{h_q}\right) \quad (1.19)$$

avec $k(\cdot)$ une fonction noyau unidimensionnelle satisfaisant (1.10).

Remarque 1.2.1. Il faut noter que l'hypothèse d'indépendance est faite sur les variables X_i et non sur les variables X_{is} . Donc l'usage du produit dans l'équation (1.19) est une commodité. Aucune hypothèse d'indépendance n'est faite entre les X_{is} .

1.2.2 Consistance et Normalité asymptotique de l'estimateur

Hypothèse 1.2.1. On suppose que la fonction de densité de probabilité f est au moins trois fois différentiable et que $k(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction noyau d'ordre 2 satisfaisant :

$$H1) \int_{\mathbb{R}} k(v_s) dv_s = 1$$

$$H2) \int_{\mathbb{R}} v_s k(v_s) dv_s = 0$$

$$H3) \int_{\mathbb{R}} v_s^2 k(v_s) dv_s = \kappa_2 > 0.$$

Comme dans le cas unidimensionnel, on a le Théorème suivant :

Théorème 2

Sous l'**Hypothèse 1.2.1** et en supposant que $h_s \rightarrow 0$ pour tout $1 \leq s \leq q$ et $nh_1 \cdots h_q \rightarrow \infty$, l'erreur quadratique moyen de l'estimateur $\hat{f}(x)$ vérifie :

$$\begin{aligned} MSE(\hat{f}(x)) &= \frac{\kappa_2^2}{4} \left[\sum_{s=1}^q \partial_{ss} f(x) h_s^2 \right]^2 + \frac{\kappa_q^2 f(x)}{nh_1 \cdots h_q} + O\left(\left(\sum_{s=1}^q h_s^2 \right)^2 + (nh_1 \cdots h_q)^{-1} \right) \\ &= O\left(\left(\sum_{s=1}^q h_s^2 \right)^2 + (nh_1 \cdots h_q)^{-1} \right). \end{aligned} \quad (1.20)$$

Preuve. De la même façon, nous examinerons le biais($\hat{f}(x)$) et la $\mathbb{V}ar(\hat{f}(x))$.

Etape 1 : Calcul du biais de $\hat{f}(x)$.

On a

$$\begin{aligned} \text{biais}(\hat{f}(x)) &= \mathbb{E} \left\{ \frac{1}{nh_1 \cdots h_q} \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) \right\} - f(x) \\ &= \frac{1}{h_1 \cdots h_q} \mathbb{E} \left[K \left(\frac{X_1 - x}{h} \right) \right] - f(x) \quad \text{car les v.a sont i.d} \\ &= \frac{1}{h_1 \cdots h_q} \int_{\mathbb{R}^q} f(x_1) K \left(\frac{x_1 - x}{h} \right) dx_1 - f(x). \quad \text{où } dx_1 = dx_{11} \cdots dx_{1q} \end{aligned}$$

Considérons le difféomorphisme

$$\begin{aligned} \varphi : \mathbb{R}^q &\longrightarrow \mathbb{R}^q \\ x_1 = (x_{11}, \dots, x_{1q}) &\mapsto v = \frac{x_1 - x}{h} \equiv \left(\frac{x_{11} - x_1}{h_1}, \dots, \frac{x_{1q} - x_q}{h_q} \right), \end{aligned} \quad (1.21)$$

φ est inversible et sa réciproque est donnée par $\varphi^{-1}(v) = x + hv \equiv (x_1 + h_1 v_1, \dots, x_q + h_q v_q)^T$ avec $v = (v_1, \dots, v_q)^T$. La matrice Jacobienne de φ^{-1} en $v \in \mathbb{R}^q$ est

$$J_{\varphi^{-1}(v)} = \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & & \vdots \\ & 0 & \ddots & \ddots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & 0 & h_q \end{pmatrix} = \text{diag}(h_1, \dots, h_q).$$

Donc $\det J_{\varphi^{-1}(v)} = \prod_{j=1}^q h_j$ et $|\det J_{\varphi^{-1}(v)}| = \prod_{j=1}^q h_j$ car $h_j > 0$ pour tout $1 \leq j \leq q$ par définition.

En appliquant la formule du changement de variables, on obtient donc :

$$\begin{aligned} \text{biais}(\hat{f}(x)) &= \frac{1}{h_1 \cdots h_q} \int_{\mathbb{R}^q} f(\varphi^{-1}(v)) K(v) h_1 \cdots h_q dv - f(x) \\ &= \int_{\mathbb{R}^q} f(x + hv) K(v) dv - f(x) \quad \text{avec } dv = dv_1 \cdots dv_q \end{aligned}$$

La formule de Taylor de f au voisinage de $x \in \mathbb{R}^q$ à l'ordre 2 s'écrit

$$f(x + l) = f(x) + \nabla f(x).l + \frac{1}{2} l^T \text{Hess} f(x) l + O(\|l\|^3).$$

Pour $l = hv \equiv (h_1 v_1, \dots, h_q v_q)$, on a

$$f(x + hv) = f(x) + \sum_{s=1}^q \partial_s f(x) h_s v_s + \frac{1}{2} \sum_{s=1}^q \sum_{s'=1}^q \partial_{ss'} f(x) h_s v_s h_{s'} v_{s'} + O(\|hv\|^3).$$

Ecriture simplifiée de $O(\|hv\|^3)$

On commence par remarquer que

$$\|hv\|^3 = \left(\sum_{s=1}^q h_s^2 v_s^2 \right)^{3/2} \leq C_1 \left(\sum_{s=1}^q h_s^2 \right)^{3/2} \quad \text{où } C_1 = \left(\max_{1 \leq s \leq q} v_s^2 \right)^{3/2}.$$

Par convexité de la fonction $u \mapsto u^3$ dans \mathbb{R}_+^* , on a

$$\left(\sum_{s=1}^q h_s^2 \right)^3 \leq q^2 \left(\sum_{s=1}^q h_s^6 \right).$$

Par la sous additivité de la fonction $u \mapsto \sqrt{u}$ dans \mathbb{R}_+^* , on a

$$\left(\sum_{s=1}^q h_s^6 \right)^{1/2} \leq \left(\sum_{s=1}^q h_s^3 \right).$$

D'où

$$\|hv\|^3 \leq C \left(\sum_{s=1}^q h_s^3 \right) \quad \text{où } C = C_1 q.$$

On déduit que $\|hv\|^3 = O\left(\sum_{s=1}^q h_s^3\right)$ et donc

$$O(\|hv\|^3) = O\left(O\left(\sum_{s=1}^q h_s^3\right)\right) = O\left(\sum_{s=1}^q h_s^3\right).$$

Ainsi

$$\begin{aligned} \text{biais}(\hat{f}(x)) &= \int_{\mathbb{R}^q} f(x) K(v) dv + \sum_{s=1}^q \partial_s f(x) h_s \int_{\mathbb{R}^q} v_s K(v) dv + \frac{1}{2} \sum_{s=1}^q \sum_{s'=1}^q \partial_{ss'} f(x) h_s h_{s'} \int_{\mathbb{R}^q} v_s v_{s'} K(v) dv \\ &\quad + \int_{\mathbb{R}^q} \left\{ O\left(\sum_{s=1}^q h_s^3\right) \right\} K(v) dv - f(x). \end{aligned}$$

Analysons les termes $A = \int_{\mathbb{R}^q} K(v) dv$, $B_s = \int_{\mathbb{R}^q} v_s K(v) dv$ et $C_{s,s'} = \int_{\mathbb{R}^q} v_s v_{s'} K(v) dv$ avec $dv = dv_1 \cdots dv_q$.

$$A = \int_{\mathbb{R}^q} K(v)dv = \int_{\mathbb{R}^q} \prod_{s=1}^q k(v_s)dv_1 \cdots dv_q = \prod_{s=1}^q \left(\underbrace{\int_{\mathbb{R}^q} k(v_s)dv_s}_{=1} \right) = 1.$$

$$\begin{aligned} B_s &= \int_{\mathbb{R}^q} v_s K(v_s)dv = \int_{\mathbb{R}^q} v_s \prod_{l=1}^q k(v_l)dv_1 \cdots dv_q \\ &= \left(\prod_{l \neq s} \underbrace{\int_{\mathbb{R}} k(v_l)dv_l}_{=1} \right) \left(\underbrace{\int_{\mathbb{R}} v_s k(v_s)dv_s}_{=0} \right) = 0. \end{aligned}$$

Expression $C_{s,s'}$ pour $s \neq s'$

$$\begin{aligned} C_{s,s'} &= \int_{\mathbb{R}^q} v_s v_{s'} K(v)dv \\ &= \left(\prod_{l \neq s, l \neq s'} \int_{\mathbb{R}} k(v_l)dv_l \right) \left(\int_{\mathbb{R}} v_s k(v_s)dv_s \right) \left(\int_{\mathbb{R}} v_{s'} k(v_{s'})dv_{s'} \right) \\ &= 1 \times 0 \times 0 = 0 \end{aligned}$$

Expression $C_{s,s'}$ pour $s = s'$

$$C_{s,s} = \int_{\mathbb{R}^q} v_s^2 K(v)dv = \left(\prod_{l \neq s} \int_{\mathbb{R}} k(v_l)dv_l \right) \left(\int_{\mathbb{R}} v_s^2 k(v_s)dv_s \right) = 1 \times \kappa_2 = \kappa_2.$$

D'où $C_{s,s'} = \kappa_2 \mathbb{1}_{\{s=s'\}}$.

Ainsi

$$\text{biais}(\hat{f}(x)) = f(x) + 0 + \frac{\kappa_2}{2} \sum_{s=1}^q \partial_{ss} f(x) h_s^2 + \mathcal{O} \left(\sum_{s=1}^q h_s^3 \right) - f(x).$$

D'où

$$\text{biais}(\hat{f}(x)) = \frac{\kappa_2}{2} \sum_{s=1}^q \partial_{ss} f(x) h_s^2 + \mathcal{O} \left(\sum_{s=1}^q h_s^3 \right). \quad (1.22)$$

Etape 2 : Calcul de $\mathbb{V}ar(\hat{f}(x))$

En utilisant le fait que les v.a sont indépendantes et identiquement distribuées, on aboutit à :

$$\mathbb{V}ar(\hat{f}(x)) = \frac{1}{nh_1^2 \cdots h_q^2} \mathbb{V}ar \left(K \left(\frac{X_1 - x}{h} \right) \right) = \frac{1}{nh_1^2 \cdots h_q^2} \left\{ \underbrace{\mathbb{E} \left(K^2 \left(\frac{X_1 - x}{h} \right) \right)}_A - \underbrace{\left(\mathbb{E} \left(K \left(\frac{X_1 - x}{h} \right) \right) \right)^2}_B \right\}.$$

En effectuant à nouveau la formule de changement de variables grâce au difféomorphisme φ , on a

$$A = \int_{\mathbb{R}^q} f(x_1) K^2\left(\frac{x_1 - x}{h}\right) dx_1 = h_1 \cdots h_q \int_{\mathbb{R}^q} f(x + hv) K^2(v) dv.$$

La formule de Taylor de f au voisinage de $x \in \mathbb{R}^q$ à l'ordre 1 donne

$$\exists \eta \in (x, x + hv) \text{ [au sens où } \forall i = 1, \dots, q \quad \eta_i \in (x_i, x_i + h_i v_i)]$$

tel que

$$f(x + hv) = f(x) + \nabla f(\eta) \cdot hv = f(x) + \sum_{s=1}^q \partial_s f(\eta) h_s v_s \quad \text{avec } hv \equiv (h_1 v_1, \dots, h_q v_q).$$

Ainsi

$$A = h_1 \cdots h_q \left(\int_{\mathbb{R}^q} f(x) K^2(v) dv + \sum_{s=1}^q \partial_s f(\eta) h_s \int_{\mathbb{R}^q} v_s K^2(v) dv \right).$$

On observe que

$$\begin{aligned} \int_{\mathbb{R}^q} K^2(v) dv &= \prod_{s=1}^q \left(\underbrace{\int_{\mathbb{R}} k^2(v_s) dv_s}_{\kappa} \right) = \kappa^q, \\ \int_{\mathbb{R}^q} v_s K^2(v) dv &= \left(\prod_{l \neq s} \int_{\mathbb{R}} k(v_l) dv_l \right) \left(\underbrace{\int_{\mathbb{R}} v_s k^2(v_s) dv_s}_{\kappa_1} \right) = \kappa^{q-1} \kappa_1, \end{aligned}$$

on déduit que

$$A = h_1 \cdots h_q \left(\kappa^q f(x) + \sum_{s=1}^q \partial_s f(\eta) h_s \kappa^{q-1} \kappa_1 \right).$$

$$\begin{aligned} B &= h_1 \cdots h_q \int_{\mathbb{R}^q} K(v) f(x + hv) dv \\ &= h_1 \cdots h_q \int_{\mathbb{R}^q} K(v) \{f(x) + \mathcal{O}(\|hv\|)\} dv. \end{aligned}$$

$$\text{Donc } B^2 = h_1^2 \cdots h_q^2 (f^2(x) + \mathcal{O}(\|hv\|)).$$

Ainsi

$$\mathbb{V}ar(\hat{f}(x)) = \frac{1}{nh_1 \cdots h_q} \left\{ \kappa^q f(x) + \sum_{s=1}^q \partial_s f(\eta) h_s \kappa^{q-1} \kappa_1 - h_1 \cdots h_q (f^2(x) + \mathcal{O}(\|hv\|)) \right\} \quad (1.23)$$

Expression simplifiée de $\sum_{s=1}^q \partial_s f(\eta) h_s \kappa^{q-1} \kappa_1$

$$\sum_{s=1}^q \partial_s f(\eta) h_s \kappa^{q-1} \kappa_1 \leq \underbrace{\kappa^{q-1} \kappa_1 \max_{1 \leq s \leq q} \partial_s f(\eta)}_C \sum_{s=1}^q h_s \leq C \sum_{s=1}^q |h_s| \leq C' \|h\|.$$

D'où

$$\sum_{s=1}^q \partial_s f(\eta) h_s \kappa^{q-1} \kappa_1 = O(\|h\|).$$

Expression simplifiée de $h_1 \cdots h_q f^2(x)$

$$h_1 \cdots h_q f^2(x) \leq |h_1| \cdots |h_q| f^2(x) \leq \|h\|^q f^2(x).$$

D'où

$$h_1 \cdots h_q f^2(x) = O(\|h\|^q).$$

Expression simplifiée de $h_1 \cdots h_q O(\|h\|)$

$$h_1 \cdots h_q O(\|h\|) = O(\|h\|^q) O(\|h\|) = O(\|h\|^{q+1}).$$

On obtient donc que

$$\mathbb{V}ar(\hat{f}(x)) = \frac{1}{nh_1 \cdots h_q} \{ \kappa^q f(x) + O(\|h\|) \}. \quad (1.24)$$

Etape 3 : Dédution du MSE $(\hat{f}(x))$

Les égalités (1.22) et (1.24) montrent que

$$\begin{aligned} MSE(\hat{f}(x)) &= \left\{ \frac{\kappa_2}{2} \sum_{s=1}^q \partial_{ss} f(x) h_s^2 + O\left(\sum_{s=1}^q h_s^3 \right) \right\}^2 + \frac{1}{nh_1 \cdots h_q} \{ \kappa^q f(x) + O(\|h\|) \} \\ &= \frac{\kappa_2^2}{4} \left[\sum_{s=1}^q \partial_{ss} f(x) h_s^2 \right]^2 + \frac{\kappa^q f(x)}{nh_1 \cdots h_q} + O\left(\left(\sum_{s=1}^q h_s^3 \right)^2 + \frac{\|h\|}{nh_1 \cdots h_q} \right) \\ &= \frac{\kappa_2^2}{4} \left[\sum_{s=1}^q \partial_{ss} f(x) h_s^2 \right]^2 + \frac{\kappa^q f(x)}{nh_1 \cdots h_q} + O\left(\left(\sum_{s=1}^q h_s^3 \right)^2 + (nh_1 \cdots h_q)^{-1} \right). \end{aligned}$$

Remarque : Si on avait écrit le reste du biais sous forme de $O\left(\sum_{s=1}^q h_s^2 \right)$, le résultat précédent

deviendrait

$$MSE(\hat{f}(x)) = \frac{\kappa_2^2}{4} \left[\sum_{s=1}^q \partial_{ss} f(x) h_s^2 \right]^2 + \frac{\kappa^q f(x)}{nh_1 \cdots h_q} + O \left(\left(\sum_{s=1}^q h_s^2 \right)^2 + (nh_1 \cdots h_q)^{-1} \right).$$

En observant que lorsque $h_s \rightarrow 0 \forall s$ et $nh_1 \cdots h_q \rightarrow \infty$, on a l'équivalence

$$MSE(\hat{f}(x)) \sim O \left(\left(\sum_{s=1}^q h_s^2 \right)^2 + (nh_1 \cdots h_q)^{-1} \right)$$

ou encore l'égalité asymptotique suivante

$$MSE(\hat{f}(x)) = O \left(\left(\sum_{s=1}^q h_s^2 \right)^2 + (nh_1 \cdots h_q)^{-1} \right).$$

■

Corollaire 1.2. Lorsque $n \rightarrow \infty$, $h_s \rightarrow 0$ pour tout $1 \leq s \leq q$ et $nh_1 \cdots h_q \rightarrow \infty$, alors $\hat{f}(x)$ converge en moyenne quadratique vers $f(x)$, donc :

$$\hat{f}(x) - f(x) = O_p \left(\sum_{s=1}^q h_s^2 + (nh_1 \cdots h_q)^{-1/2} \right) = o_p(1). \quad (1.25)$$

Preuve. Découle du Théorème 3 de l'Annexe. ■

Théorème 3

Soit X_1, \dots, X_n une suite i.i.d de vecteurs aléatoires de dimension q ayant une fonction de densité de probabilité $f(\cdot)$ trois fois différentiable dont les dérivées sont bornées et continues. Soit x un point à l'intérieur du support de X . Si, $n \rightarrow \infty$, $h_s \rightarrow 0$ pour tout $s = 1, \dots, q$, $nh_1 \cdots h_q \rightarrow \infty$, et $(nh_1 \cdots h_q) \sum_{s=1}^q h_s^3 \rightarrow 0$, alors

$$\sqrt{nh_1 \cdots h_q} \left[\hat{f}(x) - f(x) - \frac{\kappa_2}{2} \sum_{s=1}^q h_s^2 \partial_{ss} f(x) \right] \xrightarrow{d} \mathcal{N}(0, \kappa^q f(x)). \quad (1.26)$$

Preuve. Les équations (1.22) et (1.24) montrent que la moyenne asymptotique et la variance asymptotique de $\mathcal{A} \equiv \sqrt{nh_1 \cdots h_q} \left[\hat{f}(x) - f(x) - \frac{\kappa_2}{2} \sum_{s=1}^q h_s^2 \partial_{ss} f(x) \right]$ sont respectivement 0 et

$\kappa^q f(x)$; et on a :

$$\begin{aligned}
 \mathcal{A} &= \sqrt{nh_1 \cdots h_q} [\hat{f}(x) - \mathbb{E}(\hat{f}(x))] + \sqrt{nh_1 \cdots h_q} \left[\mathbb{E}(\hat{f}(x)) - f(x) - \frac{\kappa_2}{2} \sum_{s=1}^q h_s^2 \partial_{ss} f(x) \right] \\
 &= \sqrt{nh_1 \cdots h_q} [\hat{f}(x) - \mathbb{E}(\hat{f}(x))] + \mathcal{O} \left(\sqrt{nh_1 \cdots h_q} \sum_{s=1}^q h_s^3 \right) \text{ par l'équation (1.22)} \\
 &= \frac{1}{\sqrt{nh_1 \cdots h_q}} \sum_{i=1}^n \left[K \left(\frac{X_i - x}{h} \right) - \mathbb{E} \left(K \left(\frac{X_i - x}{h} \right) \right) \right] + \mathcal{O}(\circ(1)) \\
 &\equiv \sum_{i=1}^n [Z_{n,i} - \mathbb{E}(Z_{n,i})] + \mathcal{O}(1)
 \end{aligned} \tag{1.27}$$

où

$$Z_{n,i} = \frac{1}{\sqrt{nh_1 \cdots h_q}} K \left(\frac{X_i - x}{h} \right) \tag{1.28}$$

A présent vérifions les hypothèses du Théorème Central Limite de Lyapunov (Lemme 9 de l'Annexe).

D'une part, puisque les variables aléatoires X_i sont indépendantes, alors les variables aléatoires $Z_{n,i} = \frac{1}{\sqrt{nh_1 \cdots h_q}} K \left(\frac{X_i - x}{h} \right)$ sont également indépendantes dont on note la moyenne et la variance respectivement par $\mathbb{E}(Z_{n,i}) = \mu_{n,i}$ et $\mathbb{V}ar(Z_{n,i}) = \sigma_{n,i}^2$.

On a

$$\begin{aligned}
 \sigma_{n,i}^2 &= \frac{1}{nh_1 \cdots h_q} \mathbb{V}ar \left(K \left(\frac{X_i - x}{h} \right) \right) \\
 &= \frac{1}{nh_1 \cdots h_q} \left\{ \mathbb{E} \left(K^2 \left(\frac{X_i - x}{h} \right) \right) - \left(\mathbb{E} \left(K \left(\frac{X_i - x}{h} \right) \right) \right)^2 \right\} \\
 &= \frac{1}{nh_1 \cdots h_q} \left\{ h_1 \cdots h_q (\kappa^q f(x) + \mathcal{O}(\|h\|)) + h_1^2 \cdots h_q^2 (f^2(x) + \mathcal{O}(\|h\|)) \right\} \\
 &= \frac{1}{n} \left(\kappa^q f(x) + \mathcal{O}(\|h\|) + h_1 \cdots h_q f^2(x) + \mathcal{O}(\|h\|^{q+1}) \right) \\
 &= \frac{1}{n} \kappa^q f(x) + \mathcal{O}(\|h\|) = \frac{1}{n} \kappa^q f(x) + \mathcal{O}(1).
 \end{aligned}$$

Donc

$$\sigma_n^2 = \sum_{i=1}^n \sigma_{n,i}^2 = \kappa^q f(x) + \mathcal{O}(1). \tag{1.29}$$

Pour $\delta = 2$, puisque K est bornée, i.e, $\exists M > 0$ tel que $|K(v)| \leq M \quad \forall v \in \mathbb{R}^q$, alors

$$\begin{aligned} \mathbb{E} \left[|Z_{n,i}|^{2+2} \right] &= \mathbb{E} \left[\frac{1}{\sqrt{nh_1 \dots h_q}^4} \left| K \left(\frac{X_i - x}{h} \right) \right|^4 \right] \\ &= \frac{1}{n^2 h_1^2 \dots h_q^2} \int f(x_1) \left| K \left(\frac{x_1 - x}{h} \right) \right|^4 dx_1 \\ &\leq \frac{1}{n^2 h_1^2 \dots h_q^2} M^4 \int f(x_1) dx_1 = \frac{M^4}{n^2 h_1^2 \dots h_q^2} < \infty. \end{aligned} \quad (1.30)$$

Il reste enfin à montrer que

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E} \left(|Z_{n,i} - \mu_{n,i}|^{2+2} \right) = 0.$$

On a

$$|Z_{n,i} - \mu_{n,i}| = \frac{1}{\sqrt{nh_1 \dots h_q}} \left| K \left(\frac{X_i - x}{h} \right) - \mathbb{E} \left(K \left(\frac{X_i - x}{h} \right) \right) \right| \leq \frac{2M}{\sqrt{nh_1 \dots h_q}},$$

donc

$$|Z_{n,i} - \mu_{n,i}|^4 \leq \frac{4M^2}{n^2 h_1^2 \dots h_q^2}.$$

D'où

$$\sum_{i=1}^n \mathbb{E} \left(|Z_{n,i} - \mu_{n,i}|^{2+2} \right) \leq \frac{4M^2}{nh_1^2 \dots h_q^2} \xrightarrow{n \rightarrow \infty} 0. \quad (1.31)$$

Grâce aux égalités (1.29), (1.30), (1.31) établies, on peut appliquer le Théorème central limite de Lyapunov et déduire que

$$\mathcal{A} = \sum_{i=1}^n [Z_{n,i} - \mathbb{E}(Z_{n,i})] \xrightarrow{d} \mathcal{N}(0; \kappa^q f(x)).$$

■

Il est à noter que nous avons montré la convergence et la consistance de l'estimateur à noyaux point par point, pour les points à l'intérieur du support de X (donc pas sur le bord ou la frontière). L'on montre que si x est un point situé sur la frontière du support de X , alors $\hat{f}(x)$ n'est plus un estimateur consistant de $f(x)$.

Contre Exemple 1.2.1. Supposons que X soit à valeurs dans $[0, 1]$ et que $f(0) > 0$. L'on montre que $\mathbb{E} \left[\hat{f}(0) \right] = \frac{f(0)}{2} + \mathcal{O}(h)$; donc $\hat{f}(0)$ n'est pas un estimateur non biaisé même asymptotiquement. En effet

$$\hat{f}(0) = \frac{1}{nh} \sum_{i=1}^n k \left(\frac{X_i - 0}{h} \right),$$

et

$$\begin{aligned}\mathbb{E}[\hat{f}(0)] &= \frac{1}{h} \mathbb{E}\left[k\left(\frac{X_1}{h}\right)\right] = \frac{1}{h} \int_0^1 f(x_1) k\left(\frac{x_1}{h}\right) dx_1 \\ &= \int_0^{1/h} f(hv) k(v) dv \xrightarrow{h \rightarrow 0} f(0) \int_0^\infty k(v) dv = \frac{f(0)}{2}.\end{aligned}\quad (1.32)$$

On remarque donc que pour 0 étant sur la frontière du support de X l'estimateur $\hat{f}(0)$ est biaisé et donc non consistant.

Dans la suite, nous allons démontrer la consistance presque sûre uniformément de l'estimateur non paramétrique de f obtenu par la méthode des noyaux et ainsi déduire sa convergence uniforme presque sûre.

1.2.3 Convergence uniforme presque sûre de l'estimateur

Nous établissons donc la convergence uniforme presque sûre de $\hat{f}(x) - f(x)$ pour $x \in \mathcal{S}$, où \mathcal{S} est une partie compacte ne contenant pas les points frontières du support de X .

Hypothèse 1.2.2. H1) La fonction $K(\cdot)$ est bornée et intégrable, notamment

$$|K(v)| \leq M \quad \forall v \in \mathbb{R}^q, \text{ et } \int_{\mathbb{R}^q} K(v) dv = 1.$$

H2) La fonction $K(\cdot)$ est lipschitzienne, ie, $\exists C_1 > 0, |K(u) - K(v)| \leq C_1 \|u - v\|$

Théorème 4 (Convergence uniforme presque sûre)

Sous certaines conditions de régularité sur $f(\cdot)$ données dans Masry(1996) (16), et en supposant que $\inf_{x \in \mathcal{S}} f(x) \geq \delta > 0$, on a :

$$\sup_{x \in \mathcal{S}} |\hat{f}(x) - f(x)| = \mathcal{O}\left(\frac{(\ln(n))^{1/2}}{(nh_1 \cdots h_q)^{1/2}} + \sum_{i=1}^n h_s^2\right) \quad p.s \quad (1.33)$$

Preuve. Dans les méthodes de choix des paramètres de lissage optimaux, on montre que l'ordre du biais de $\hat{f}(x)$ est petit pour $h_1 = \cdots = h_q = h$ avec $h > 0$. Donc nous supposons $h_1 = \cdots = h_q = h$ pour raison de simplicité et l'estimateur $\hat{f}(x)$ s'écrit donc

$$\hat{f}(x) = \frac{1}{nh^q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right).$$

Les étapes de la preuve

Notons par $W_n \equiv W_n(x)$ une suite de variables aléatoires. D'après le Lemme de Borel-Cantelli, pour montrer que W_n est de l'ordre de (η) , ie, $W_n = O(\eta)$ p.s, avec $\eta > 0$, il suffit de montrer que $\sum_{n=1}^{\infty} \mathbb{P}(|W_n| > \eta)$ est finie.

La présence du \sup dans l'égalité (1.33) complique la preuve car \mathcal{S} est un ensemble infini.

Si L_n est un ensemble fini, alors on peut écrire

$$\mathbb{P}\left(\max_{x \in L_n} W_n(x) > \eta\right) \leq \text{card}(L_n) \max_{x \in L_n} \mathbb{P}(W_n(x) > \eta). \quad (1.34)$$

En effet, $\left\{\max_{x \in L_n} W_n(x) > \eta\right\} = \{\exists x \in L_n, W_n(x) > \eta\}$, donc

$$\mathbb{P}\left(\max_{x \in L_n} W_n(x) > \eta\right) = \mathbb{P}\left(\bigcup_{x \in L_n} W_n(x) > \eta\right) \leq \sum_{x \in L_n} \mathbb{P}(W_n(x) > \eta) \leq \text{card}(L_n) \max_{x \in L_n} \mathbb{P}(W_n(x) > \eta).$$

La partie \mathcal{S} n'étant pas finie, on ne peut pas encore se servir de l'égalité (1.34) pour borner $\mathbb{P}\left(\sup_{x \in \mathcal{S}} |\hat{f}(x) - f(x)| > \eta\right)$.

Néanmoins, comme \mathcal{S} est compacte, on peut le partitionner par un nombre fini de sous ensembles de \mathbb{R}^q de volume fini et se ramener à un problème du type $\mathbb{P}\left(\max_{x \in L_n} W_n(x) > \eta\right)$ et pouvoir utiliser l'inégalité (1.34). En utilisant donc cette idée, on écrit

$$\begin{aligned} |\hat{f}(x) - f(x)| &= |\hat{f}(x) - \mathbb{E}(\hat{f}(x)) + \mathbb{E}(\hat{f}(x)) - f(x)| \\ &\leq |\hat{f}(x) - \mathbb{E}(\hat{f}(x))| + |\mathbb{E}(\hat{f}(x)) - f(x)|. \end{aligned}$$

Notre preuve se déroulera en deux grandes étapes :

Etape 1 : On montrera que

$$\sup_{x \in \mathcal{S}} |\mathbb{E}(\hat{f}(x)) - f(x)| = O(h^2). \quad (1.35)$$

Etape 2 : On montrera que

$$\sup_{x \in \mathcal{S}} |\hat{f}(x) - \mathbb{E}(\hat{f}(x))| = O\left(\frac{(\ln n)^{1/2}}{(nh^q)^{1/2}}\right) \text{ p.s.} \quad (1.36)$$

Étape 1 : Établissons l'égalité (1.35)

Puisque $\mathcal{S} \subset \overset{\circ}{\mathcal{S}}$, on peut appliquer la formule de changement de variables pour $x \in \mathcal{S}$ et écrire

$$\begin{aligned}\mathbb{E}(\hat{f}(x)) - f(x) &= \int_{\mathbb{R}^q} f(x + hv)K(v)dv - f(x) \\ &= \int_{\mathbb{R}^q} \left\{ f(x) + \sum_{s=1}^q \partial_s f(x) h_s v_s + \frac{1}{2} \sum_{s=1}^q \sum_{s'=1}^q \partial_{ss'} f(\tilde{x}) h_s v_s h_{s'} v_{s'} \right\} K(v)dv - f(x) \\ &= \frac{1}{2} \sum_{s=1}^q \partial_{s,s} f(\tilde{x}) h_s^2 \kappa_2.\end{aligned}$$

Et comme $h_s = h \forall s$, on a

$$\mathbb{E}(\hat{f}(x)) - f(x) = h^2 \left(\frac{\kappa_2}{2} \sum_{s=1}^q \partial_{s,s} f(\tilde{x}) \right) = \mathcal{O}(h^2)$$

qui ne dépend pas de $x \in \mathcal{S}$, donc en passant au sup à gauche on a

$$\sup_{x \in \mathcal{S}} |\mathbb{E}(\hat{f}(x)) - f(x)| = \mathcal{O}(h^2).$$

Étape 2 : Établissons l'égalité (1.36)

On commence par observer qu'on a toujours $\mathcal{S} \subset \bigcup_{x \in \mathcal{S}} B(x, r)$ où $r > 0$ est quelconque ; ceci signifie qu'on peut toujours recouvrir \mathcal{S} par une famille quelconque de boules indicées par leurs centres . Puisque \mathcal{S} est compact (fermée et bornée), on peut le recouvrir par un nombre fini $J_n = J(n)$ de boules $I_j = I_{j,n}$ de centre $x_{j,n}$ et de rayon $r = r(n)$.

On écrit :

$$\begin{aligned}\sup_{x \in \mathcal{S}} |\hat{f}(x) - \mathbb{E}(\hat{f}(x))| &= \max_{1 \leq j \leq J(n)} \sup_{x \in \mathcal{S} \cap I_j} |\hat{f}(x) - \mathbb{E}(\hat{f}(x))| \\ &= \max_{1 \leq j \leq J(n)} \sup_{x \in \mathcal{S} \cap I_j} |\hat{f}(x) - \hat{f}(x_{j,n})| + \max_{1 \leq j \leq J(n)} \sup_{x \in \mathcal{S} \cap I_j} |\hat{f}(x_{j,n}) - \mathbb{E}(\hat{f}(x_{j,n}))| \\ &\quad + \max_{1 \leq j \leq J(n)} \sup_{x \in \mathcal{S} \cap I_j} |\mathbb{E}(\hat{f}(x_{j,n})) - \mathbb{E}(\hat{f}(x))| \\ &\equiv Q_1^n + Q_2^n + Q_3^n.\end{aligned}$$

Étape 2.a : Expression de $Q_2^n = \max_{1 \leq k \leq L(n)} \sup_{x \in S \cap I_k} |\hat{f}(x_{k,n}) - \mathbb{E}(\hat{f}(x_{k,n}))|$

Posons $W_n(x) = \hat{f}(x) - \mathbb{E}(\hat{f}(x))$. On a :

$$W_n(x) = \sum_{i=1}^n \underbrace{\frac{1}{nh^q} \left\{ K\left(\frac{X_i - x}{h}\right) - \mathbb{E}\left[K\left(\frac{X_i - x}{h}\right)\right] \right\}}_{Z_{n,i}} = \sum_{i=1}^n Z_{n,i}.$$

Soit $\eta > 0$, on a :

$$\begin{aligned} \mathbb{P}(Q_2^n > \eta) &= \mathbb{P}\left(\max_{1 \leq k \leq J(n)} |W_n(x_{j,n})| > \eta\right) \\ &\leq \mathbb{P}\left[\bigcup_{j=1}^{J(n)} \{|W_n(x_{j,n})| > \eta\}\right] \leq \sum_{j=1}^{J(n)} \mathbb{P}(|W_n(x_{j,n})| > \eta) \\ &\leq J(n) \sup_{x \in S} \mathbb{P}(|W_n(x)| > \eta). \end{aligned} \quad (1.37)$$

Puisque $K(\cdot)$ est borné, on peut poser $A_1 = \sup_{x \in S} K(x)$ et on a

$$|Z_{n,i}| \leq \frac{2A_1}{nh^q} \quad \forall i = 1 \dots n.$$

Définissons λ_n tel que $\lambda_n |Z_{n,i}| \leq \frac{1}{2} \quad \forall i = 1, \dots, n$.

(N'importe quel choix de λ_n tel que $\lambda_n \leq \frac{nh^q}{4A_1}$ est judicieux, mais on montrera que le choix $\lambda_n = \sqrt{nh^q \ln n}$ est optimal.)

Montrons que $e^x \leq 1 + x + x^2$ pour $|x| \leq \frac{1}{2}$

On pose $\phi(x) = e^x - 1 - x - x^2$ pour $|x| \leq \frac{1}{2}$ et on a le tableau de variations suivant :

x	$-\frac{1}{2}$	0	$\frac{1}{2}$
$\phi''(x)$	-		+
$\phi'(x)$	$e^{-\frac{1}{2}}$	$-\alpha$	
$\phi(x)$	$\phi(-\frac{1}{2})$	0	$\phi(\frac{1}{2})$

FIGURE 1.2 – Tableau de variations de la fonction ϕ

où $\alpha > 0$. Donc $\phi(x) \leq 0 \quad \forall |x| \leq \frac{1}{2}$, ie $e^x \leq 1 + x + x^2$.

Ainsi

$$e^{\lambda_n Z_{n,i}} \leq 1 + \lambda_n Z_{n,i} + \lambda_n^2 Z_{n,i}^2 \text{ et } e^{-\lambda_n Z_{n,i}} \leq 1 - \lambda_n Z_{n,i} + \lambda_n^2 Z_{n,i}^2.$$

Donc comme $\mathbb{E}(Z_{n,i}) = 0$, alors :

$$\mathbb{E}(e^{\lambda_n Z_{n,i}}) \leq 1 + \lambda_n^2 \mathbb{E}(Z_{n,i}^2) \text{ et } \mathbb{E}(e^{-\lambda_n Z_{n,i}}) \leq 1 + \lambda_n^2 \mathbb{E}(Z_{n,i}^2).$$

On rappelle en outre que $1 + v \leq e^v$ pour tout $v \geq 0$.

On déduit donc que

$$\mathbb{E}(e^{\lambda_n Z_{n,i}}) \leq e^{\mathbb{E}(\lambda_n^2 Z_{n,i}^2)} \text{ et } \mathbb{E}(e^{-\lambda_n Z_{n,i}}) \leq e^{\mathbb{E}(\lambda_n^2 Z_{n,i}^2)}. \quad (1.38)$$

En utilisant l'inégalité de Markov (A.8) que l'on rappelle encore ici, $\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}[\exp(aX)]}{\exp(ac)}$, on peut écrire que

$$\begin{aligned} \mathbb{P}(|W_n(x)| > \eta) &= \mathbb{P}\left(\left|\sum_{i=1}^n Z_{n,i}\right| > \eta\right) = \mathbb{P}\left(\sum_{i=1}^n Z_{n,i} > \eta\right) + \mathbb{P}\left(-\sum_{i=1}^n Z_{n,i} > \eta\right) \\ &\leq \frac{\mathbb{E}\left(\exp\left\{\lambda_n \sum_{i=1}^n Z_{n,i}\right\}\right) + \mathbb{E}\left(\exp\left\{-\lambda_n \sum_{i=1}^n Z_{n,i}\right\}\right)}{e^{\lambda_n \eta}} \quad D'après (A.8) \\ &\leq 2e^{-\lambda_n \eta} \exp\left(\lambda_n^2 \sum_{i=1}^n \mathbb{E}(Z_{n,i}^2)\right). \quad D'après (1.38) \end{aligned}$$

$$\text{Or } Z_{n,i} = \frac{1}{nh^q} \left\{ K\left(\frac{X_i - x}{h}\right) - \mathbb{E}\left(K\left(\frac{X_i - x}{h}\right)\right) \right\} \leq \frac{1}{nh^q} K\left(\frac{X_i - x}{h}\right).$$

Donc

$$\begin{aligned} \mathbb{E}(Z_{n,i}^2) &\leq \frac{1}{(nh^q)^2} \mathbb{E}\left(K^2\left(\frac{X_i - x}{h}\right)\right) = \frac{1}{(nh^q)^2} h^q \int K^2(v) f(x + hv) dv \\ &\leq \frac{1}{n^2 h^q} \int K^2(v) \{f(x) + o(1)\} dv = A_2 (n^2 h^q)^{-1} (1 + o(1)); \end{aligned}$$

$$\text{où } A_2 = \int K^2(v) dv.$$

On déduit que

$$\mathbb{P}(|W_n(x)| > \eta) \leq 2e^{-\lambda_n \eta} e^{A_2 \frac{\lambda_n^2}{nh^q}} (1 + o(1)). \quad (1.39)$$

Comme le membre droit de l'inégalité (1.39) ne dépend pas de x , en passant au sup, on a

$$\begin{aligned} \sup_{x \in S} \mathbb{P}(|W_n(x)| > \eta) &\leq 2e^{-\lambda_n \eta + \frac{A_2 \lambda_n^2}{nh^q}} (1 + o(1)) = 2e^{-\lambda_n \eta + \frac{A_2 \lambda_n^2}{nh^q} + o(1)} \\ &= 2e^{-\lambda_n \eta + \frac{A_2 \lambda_n^2}{nh^q}} e^{o(1)} = 2e^{-\lambda_n \eta + \frac{A_2 \lambda_n^2}{nh^q}}. \quad (1.40) \end{aligned}$$

Nous aimerons avoir $\eta \rightarrow 0$ le plus rapidement possible et en même temps que $\lambda_n \eta \rightarrow \infty$, ce qui assurerait que le membre droit de l'inégalité (1.40) soit sommable (On rappelle qu'une série $\{a_n\}_{n=1}^{\infty}$ est sommable si $|\sum_{n=1}^{\infty} a_n| < \infty$).

Pour avoir le terme d'une série sommable, on cherche à écrire $\bigwedge = e^{-\lambda_n \eta + \frac{A_2 \lambda_n^2}{nh^q}}$ sous la forme $\frac{C'}{n^\alpha}$ où C' et $\alpha > 1$ sont à déterminer.

Idée intuitive

On observe que $\bigwedge = (e^\lambda)^{-\eta} (e^{(nh^q)^{-1} \lambda_n^2})^{A_2}$.

Il vient que η et λ_n peuvent être choisis sous les formes respectives

$$\eta = C \frac{\ln n}{\lambda_n} \text{ et } \lambda_n^2 = nh^q \ln n.$$

Pour ces choix de η et λ_n , on a

$$\bigwedge = \frac{1}{n^C} \times \frac{1}{n^{-A_2}} = \frac{1}{n^{C-A_2}} \equiv \frac{1}{n^\alpha}.$$

On a donc

$$\mathbb{P}(Q_2^n > \eta) \leq \frac{2J(n)}{n^\alpha}, \quad (1.41)$$

avec $\alpha = C - A_2$. En choisissant donc C suffisamment grand (prendre par exemple $C > 1 + A_2$), on peut obtenir que $\frac{2J(n)}{n^\alpha}$ est sommable en choisissant proprement l'ordre $J(n)$, ie,

$$\sum_{n=1}^{\infty} \mathbb{P}(Q_2^n > \eta) \leq 2 \sum_{n=1}^{\infty} \frac{J(n)}{n^\alpha} < \infty.$$

Par le Lemme de Borel-Cantelli, on déduit que

$$Q_2^n \leq \eta \text{ p.s.}$$

ou encore $Q_2^n = \mathcal{O}(\eta)$ p.s. Or $\eta = C \frac{\ln n}{\lambda_n} = C \frac{\sqrt{\ln n}}{\sqrt{nh^q}}$;

d'où

$$Q_2^n = \mathcal{O}\left(\frac{\sqrt{\ln n}}{(nh^q)^{1/2}}\right) \text{ p.s.} \quad (1.42)$$

Étape 2.b : Expressions de $Q_1^n = \max_{1 \leq j \leq J(n)} \sup_{x \in S \cap I_j} |\hat{f}(x) - \hat{f}(x_{j,n})|$ et de

$$Q_3^n = \max_{1 \leq j \leq J(n)} \sup_{x \in S \cap I_j} |\mathbb{E}(\hat{f}(x_{j,n})) - \mathbb{E}(\hat{f}(x))|$$

Expression de Q_1^n

Par la condition de Lipschitz de la fonction $K(\cdot)$, on a

$$\begin{aligned} \sup_{x \in S \cap I_j} |\hat{f}(x) - \hat{f}(x_{j,n})| &= \frac{1}{nh^q} \sup_{x \in S \cap I_j} \left| K\left(\frac{X_i - x}{h}\right) - K\left(\frac{X_i - x_{j,n}}{h}\right) \right| \\ &\leq \frac{1}{nh^q} C_1 \sup_{x \in S \cap I_j} \left\| \frac{x - x_{j,n}}{h} \right\| \leq \frac{r(n)C_1}{nh^{q+1}}. \end{aligned} \quad (1.43)$$

Étant donné que $r(n) > 0$ est quelconque, on peut le choisir de telle sorte que l'expression $\frac{r(n)C_1}{nh^{q+1}}$ soit égale à une constante multiplicative près au terme $\frac{\sqrt{\ln n}}{(nh^q)^{1/2}}$.

On a alors $\frac{r(n)C_1}{nh^{q+1}} = C' \frac{\sqrt{\ln n}}{(nh^q)^{1/2}}$,

ce qui implique que $r(n) = C' \frac{\sqrt{\ln n}}{\sqrt{n}} h^{\frac{q+2}{2}}$; donc on peut prendre par exemple

$$r(n) = \frac{\sqrt{\ln n}}{\sqrt{n}} h^{\frac{q+2}{2}}.$$

Et pour ce $r(n)$, on a

$$Q_1^n = O\left(\frac{\sqrt{\ln n}}{(nh^q)^{1/2}}\right). \quad (1.44)$$

Expression Q_3^n

Par le même argument que précédemment en remarquant que

$$\mathbb{E}(\hat{f}(x_{j,n})) - \mathbb{E}(\hat{f}(x)) = \mathbb{E}(\hat{f}(x_{j,n}) - \hat{f}(x)),$$

On aboutit également à

$$Q_3^n = O\left(\frac{\sqrt{\ln n}}{(nh^q)^{1/2}}\right). \quad (1.45)$$

Les égalités (1.42), (1.44) et (1.45) montrent donc que

$$\sup_{x \in S} |\hat{f}(x) - \mathbb{E}(\hat{f}(x))| = O\left(\frac{(\ln n)^{1/2}}{(nh^q)^{1/2}}\right) \quad p.s.$$

Les égalités (1.35) et (1.36) ainsi établies achèvent la preuve, car comme $h_s = h \quad \forall s$, alors

$$O\left(\sum_{s=1}^q h_s^2\right) = O(qh^2) = O(h^2).$$

■

Remarque 1.2.2. Puisque la convergence uniforme presque sûre implique la convergence en probabilité, alors sous les mêmes conditions on a :

$$\sup_{x \in S} |\hat{f}(x) - f(x)| = O_p\left(\frac{(\ln(n))^{1/2}}{(nh_1 \cdots h_q)^{1/2}} + \sum_{s=1}^n h_s^2\right)$$

ESTIMATION DANS LES MODÈLES DE PANEL NON-LINÉAIRES

Le cadre

Un panel est une collection d'observations faites sur N individus sur T périodes. Dans ce chapitre, $i = 1, 2, \dots, N$ représente l'indice de l'individu et $t = 1, 2, \dots, T$ est l'indice du temps. T est relativement petit par rapport à N .

Nous considérons le modèle de panel non paramétrique suivant :

$$Y_{it} = g(Z_{it}) + u_{it}, \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T, \quad (2.1)$$

où Z_{it} est un vecteur aléatoire continu de dimension q , nous noterons

$$Z_{it} = (Z_{it,1}, Z_{it,2}, \dots, Z_{it,q}). \quad (2.2)$$

On suppose que les observations sont indépendantes par individus, Z_{it} est strictement exogène, i.e, $\mathbb{E}(u_{it} | Z_{i1}, Z_{i2}, \dots, Z_{iT}) = 0 \quad \forall t$. Nous supposons en plus que l'erreur prend en compte l'effet individuel, soit :

$$u_{it} = \mu_i + v_{it}, \quad (2.3)$$

où μ_i est une suite de variables i.i.d de moyenne nulle et de variance σ_μ^2

v_{it} est un processus de moyenne nulle et de variance σ_v^2 .

La fonction $g(\cdot)$ est inconnue et l'on veut l'estimer de façon non paramétrique par la méthode des noyaux.

Hypothèse 2.0.1. *H1) La fonction $g(\cdot)$ est bornée et au moins deux fois différentiable*

H2) $\mathbb{E}(u_{it} | Z_{it}) = 0$

2.1 Structure de l'erreur

Il est question d'écrire le vecteur erreur u sous forme matricielle, de déterminer sa matrice de variance covariance et de déduire son inverse. Ils servent notamment dans l'estimation de la composante linéaire dans les panels semi-linéaires de la forme

$$y_{it} = x_{it}^T \beta + \theta(Z_{it}) + u_{it}. \quad (2.4)$$

On a $u_{it} = \mu_i + v_{it}$ et on note $u = (u_{11}, \dots, u_{1T}, u_{21}, \dots, u_{2T}, \dots, u_{N1}, \dots, u_{NT})^T$.

Proposition 1

i) Le vecteur u s'écrit encore sous la forme

$$u = (I_N \otimes e_T) \mu + v, \quad (2.5)$$

avec $\mu = (\mu_1, \dots, \mu_N)^T$, $v = (v_{11}, \dots, v_{1T}, \dots, v_{N1}, \dots, v_{NT})^T$ et $e_T = \underbrace{(1, \dots, 1)^T}_{T\text{-fois}}$.

ii) La matrice de variance-covariance du vecteur u est

$$\Sigma = \mathbb{E}(uu^T) = \sigma_\mu^2 I_N \otimes J_T + \sigma_v I_{NT} = I_N \otimes [\sigma_\mu^2 \bar{J}_T + \sigma_v^2 E_T] \equiv I_N \otimes \Omega, \quad (2.6)$$

où $J_T = e_T e_T^T$, $\bar{J}_T = \frac{J_T}{T}$, $E_T = I_T - \bar{J}_T$ et $\sigma_\mu^2 = T\sigma_\mu^2 + \sigma_v^2$.

iii) L'inverse de la matrice de variance covariance de u est

$$\Sigma^{-1} = I_N \otimes \left[\frac{1}{\sigma_\mu^2} \bar{J}_T + \frac{1}{\sigma_v^2} E_T \right] \equiv I_N \otimes \Omega^{-1}. \quad (2.7)$$

Preuve. i) Découle de la définition même du produit de Kronecker des matrices.

ii) On a

$$\begin{aligned} uu^T &= ((I_N \otimes e_T) \mu + v) ((I_N \otimes e_T) \mu + v)^T \\ &= ((I_N \otimes e_T) \mu + v) (\mu^T (I_N \otimes e_T^T) + v^T) \\ &= (I_N \otimes e_T) \mu \mu^T (I_N \otimes e_T^T) + (I_N \otimes e_T) \mu v^T + v \mu^T (I_N \otimes e_T^T) + v v^T. \end{aligned}$$

Puisque μ et ν sont indépendants et que $\mathbb{E}(\mu) = \mathbb{E}(\nu) = 0$, on déduit que

$$\begin{aligned}\Sigma = \mathbb{E}(uu^T) &= (I_N \otimes e_T)\sigma_\mu^2(I_N \otimes e_T^T) + 0 + 0 + \sigma_\nu^2 I_{NT} = \sigma_\mu^2(I_N \otimes e_T e_T^T) + \sigma_\nu^2 I_{NT} \\ &= \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2 I_{NT}.\end{aligned}$$

D'où la première égalité.

La seconde égalité n'est qu'une réécriture de l'égalité précédente car $I_{NT} = I_N \otimes I_T$.

iii) On remarque que $e_T^T e_T = T$, donc $\frac{1}{T} = (e_T^T e_T)^{-1}$; on déduit donc que

$$\bar{J}_T = \frac{J_T}{T} = e_T(e_T^T e_T)^{-1} e_T^T. \quad (2.8)$$

Donc \bar{J}_T est la matrice de la projection orthogonale sur le sous espace vectoriel $\langle e_T \rangle$ engendré par le vecteur e_T : donc elle est idempotente. On déduit donc que $E_T = I_T - \bar{J}_T$ est la matrice de la projection orthogonale sur le sous espace vectoriel $\langle e_T \rangle^\perp$: donc E_T est idempotente et est orthogonale à la matrice \bar{J}_T . Sous ces remarques, on montre aisément que l'inverse de la matrice $\Omega = \sigma_1^2 \bar{J}_T + \sigma_\nu^2 E_T$ est $\Omega^{-1} = \frac{1}{\sigma_1^2} \bar{J}_T + \frac{1}{\sigma_\nu^2} E_T$.

Remarque 2.1.1. Cette décomposition est inspirée de [Wansbet \(1962\)](#).

■

2.2 Construction de l'estimateur de $g(z)$ par la méthode des noyaux

Par construction du modèle, on a

$$Y_{it} = \underbrace{\mathbb{E}(Y_{it} | Z_{it})}_{g(Z_{it})} + \underbrace{Y_{it} - \mathbb{E}(Y_{it} | Z_{it})}_{u_{it}} = g(Z_{it}) + u_{it},$$

et $\mathbb{E}(u_{it}) = 0$ ou $\mathbb{E}(u_{it} | Z_{it}) = 0$.

Donc nous interpréterons $g(z)$ pour $z \in \mathbb{R}^q$ comme étant l'espérance conditionnelle de Y_{it} sachant que $Z_{it} = z$, ie, $g(z) = \mathbb{E}(Y_{it} | Z_{it} = z) = \mathbb{E}(Y | Z = z)$, et le Théorème suivant prouve davantage cette construction optimal :

Théorème 1

Notons par \mathcal{G} la classe des fonctions boréliennes mesurables ayant des moments d'ordre 2 finis. Supposons que $g(z) \equiv \mathbb{E}(Y | Z = z)$ appartient à \mathcal{G} et que $\mathbb{E}(Y^2) < \infty$. Alors $\mathbb{E}(Y | Z)$ est le meilleur optimal de Y sachant Z au sens suivant :

$$\mathbb{E}\{[Y - r(Z)]^2\} \geq \mathbb{E}\{[Y - \mathbb{E}(Y | Z)]^2\} \quad \forall r(\cdot) \in \mathcal{G}. \quad (2.9)$$

Preuve. En introduisant $\mathbb{E}(Y | Z)$ dans le membre de gauche de l'inégalité (2.9), on a le résultat. En effet :

$$\begin{aligned} \mathbb{E}\{[Y - r(Z)]^2\} &= \mathbb{E}\{[Y - \mathbb{E}(Y | Z) + \mathbb{E}(Y | Z) - r(Z)]^2\} \\ &= \mathbb{E}\{[Y - \mathbb{E}(Y | Z)]^2\} + \mathbb{E}\{[\mathbb{E}(Y | Z) - r(Z)]^2\} \\ &\quad + \underbrace{2\mathbb{E}\{(Y - \mathbb{E}(Y | Z))(\mathbb{E}(Y | Z) - r(Z))\}}_A \\ &= \mathbb{E}\{[Y - \mathbb{E}(Y | Z)]^2\} + \mathbb{E}\{[\mathbb{E}(Y | Z) - r(Z)]^2\} \\ &\geq \mathbb{E}\{[Y - \mathbb{E}(Y | Z)]^2\}. \end{aligned}$$

Car en utilisant les propriétés de l'espérance conditionnelle pour avoir $A = 0$.

■

On a donc la proposition suivante :

Proposition 2 (Estimateur de Nadaraya Watson)

Le prédicteur optimal de la fonction $g(\cdot)$ dans la classe \mathcal{G} des fonctions boréliennes mesurables de moments d'ordre 2 est défini par :

$$\hat{g}(z) = \frac{\int y \hat{f}_{y,z}(z, y) dy}{\hat{f}(z)} = \frac{\sum_{i=1}^N \sum_{t=1}^T Y_{it} K_{h,it,z}}{\sum_{i=1}^N \sum_{t=1}^T K_{h,it,z}}; \quad (2.10)$$

avec

$$K_{h,it,z} = \prod_{l=1}^q \frac{1}{h_l} k\left(\frac{Z_{it,l} - z_l}{h_l}\right). \quad (2.11)$$

Preuve. En notant $f_{y,z}(z, y)$ et $f(z)$ les densités de probabilité de (Z, Y) et Z respectivement,

on peut écrire que :

$$g(z) = \mathbb{E}(Y | Z = z) = \frac{\int y f_{y,z}(z, y) dy}{f(z)}. \quad (2.12)$$

Il suffit donc d'estimer $f_{y,z}(z, y)$ et $f(z)$ par la méthode des noyaux vu au **Chapitre 1**, par :

$$\hat{f}_{y,z}(z, y) = \frac{1}{NT h_0 h_1 \cdots h_q} \sum_{i=1}^N \sum_{t=1}^T K\left(\frac{Z_{it} - z}{h}\right) k\left(\frac{y - Y_{it}}{h_0}\right) \quad (2.13)$$

et

$$\hat{f}(z) = \frac{1}{NT h_1 \cdots h_q} \sum_{i=1}^N \sum_{t=1}^T K\left(\frac{Z_{it} - z}{h}\right) \quad (2.14)$$

où

$$K\left(\frac{Z_{it} - z}{h}\right) = \prod_{l=1}^q k\left(\frac{Z_{it,l} - z_l}{h_l}\right) \quad (2.15)$$

et h_0 est le paramètre de lissage associé à Y et $h_1 \cdots h_q$ celui associé à Z . On a :

$$\int y f_{y,z}(z, y) dy = \frac{1}{NT h_0 h_1 \cdots h_q} \sum_{i=1}^N \sum_{t=1}^T K\left(\frac{Z_{it} - z}{h}\right) \int y k\left(\frac{y - Y_{it}}{h_0}\right) dy.$$

En effectuant le changement de variables $u = \frac{y - Y_{it}}{h_0}$, et en utilisant le fait que $\int k(u) du = 1$ et $\int u k(u) du = 0$ on aboutit à

$$\int y k\left(\frac{y - Y_{it}}{h_0}\right) dy = \int (Y_{it} + u h_0) k(u) h_0 du = Y_{it} h_0.$$

Donc

$$\begin{aligned} \int y \hat{f}_{y,z}(z, y) dy &= \frac{1}{NT h_0 h_1 \cdots h_q} \sum_{i=1}^N \sum_{t=1}^T K\left(\frac{Z_{it} - z}{h}\right) Y_{it} h_0. \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Y_{it} \underbrace{\frac{1}{h_1 \cdots h_q} K\left(\frac{Z_{it} - z}{h}\right)}_{K_{h,it,z}} \equiv \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Y_{js} K_{h,it,z}. \end{aligned} \quad (2.16)$$

De même

$$\hat{f}(z) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T K_{h,it,z}, \quad (2.17)$$

où

$$K_{h,it,z} = \prod_{l=1}^q \frac{1}{h_l} k\left(\frac{Z_{it,l} - z_l}{h_l}\right).$$

Grâce aux égalités (2.16) et (2.17) on aboutit à l'estimateur :

$$\hat{g}(z) = \frac{\int y \hat{f}_{y,z}(z, y) dy}{\hat{f}(z)} = \frac{\sum_{i=1}^N \sum_{t=1}^T Y_{it} K_{h,it,z}}{\sum_{i=1}^N \sum_{t=1}^T K_{h,it,z}}.$$

L'estimateur ainsi obtenu est appelé *Estimateur de Nadaraya Watson (1964)* (17).

■

2.3 Consistance de l'estimateur

Proposition 3 (Convergence en probabilité de l'estimateur $\hat{g}(z)$)

Sous l'hypothèse que z est un point intérieur, $g(z)$ et $f(z)$ sont au moins trois fois continûment différentiables, et $f(z) > 0$, alors lorsque $N \rightarrow \infty$, $h_s \rightarrow 0$ (pour tout s) et $NTH_q \rightarrow \infty$ (avec $H_q = h_1 \cdots h_q$), nous avons

$$\begin{aligned} \hat{g}(z) - g(z) &= O_p \left(\sum_{s=1}^q h_s^2 + (NTh_1 \cdots h_q)^{-1/2} \right) \\ &= O_p \left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2} \right). \end{aligned} \quad (2.18)$$

Preuve. On écrit :

$$\hat{g}(z) - g(z) = \frac{(\hat{g}(z) - g(z))\hat{f}(z)}{\hat{f}(z)} \equiv \frac{\hat{m}(z)}{\hat{f}(z)}, \quad (2.19)$$

où $\hat{m}(z) = (\hat{g}(z) - g(z))\hat{f}(z)$. En utilisant le fait que $Y_{it} = g(Z_{it}) + u_{it}$, on a :

$$\begin{aligned} \hat{m}(z) &= \hat{g}(z)\hat{f}(z) - g(z)\hat{f}(z) \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Y_{it} K_{h,it,z} - \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T g(z) K_{h,it,z} \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (g(Z_{it}) + u_{it}) K_{h,it,z} - \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T g(z) K_{h,it,z} \\ &= \underbrace{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (g(Z_{it}) - g(z)) K_{h,it,z}}_{\hat{m}_1(z)} + \underbrace{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T u_{it} K_{h,it,z}}_{\hat{m}_2(z)} \end{aligned}$$

En posant $H_q = h_1 \cdots h_q$, on peut encore écrire

$$\hat{m}_1(z) = \frac{1}{NTH_q} \sum_{i=1}^N \sum_{t=1}^T (g(Z_{it}) - g(z)) K\left(\frac{Z_{it} - z}{h}\right) \text{ et } \hat{m}_2(z) = \frac{1}{NTH_q} \sum_{i=1}^N \sum_{t=1}^T u_{it} K\left(\frac{Z_{it} - z}{h}\right).$$

Etape 1 : Expression de $\hat{m}_1(z)$

En utilisant la formule du changement de variables via le difféomorphisme (1.21), on a d'une part :

$$\begin{aligned} \mathbb{E} [\hat{m}_1(z)] &= H_q^{-1} \mathbb{E} \left[(g(Z_{11}) - g(z)) K\left(\frac{Z_{11} - z}{h}\right) \right] \\ &= H_q^{-1} \int f(z_{11}) [g(z_{11}) - g(z)] K\left(\frac{z_{11} - z}{h}\right) dz_{11} \\ &= \int_{\mathbb{R}^q} f(z + hv) [g(z + hv) - g(z)] K(v) dv \\ &= \int_{\mathbb{R}^q} \left\{ f(z) + \sum_{s=1}^q \partial_s f(z) h_s v_s + \frac{1}{2} \sum_{s,s'=1}^q \partial_{ss'} f(z) h_s v_s h_{s'} v_{s'} + \mathcal{O}\left(\sum_{s=1}^q h_s^2\right) \right\} \\ &\quad \times \left[\sum_{s=1}^q \partial_s g(z) h_s v_s + \frac{1}{2} \sum_{s,s'=1}^q \partial_{ss'} g(z) h_s v_s h_{s'} v_{s'} + \mathcal{O}\left(\sum_{s=1}^q h_s^2\right) \right] K(v) dv \\ &= \frac{\kappa_2}{2} \sum_{s=1}^q h_s^2 \{ 2\partial_s f(z) \partial_s g(z) + f(z) \partial_{ss} g(z) \} + \mathcal{O}\left(\sum_{s=1}^q h_s^2\right) \\ &= f(z) \sum_{s=1}^q h_s^2 B_s(z) + \mathcal{O}\left(\sum_{s=1}^q h_s^2\right) \end{aligned} \quad (2.20)$$

avec

$$B_s(z) = \frac{\kappa_2}{2} \frac{2\partial_s f(z) \partial_s g(z) + f(z) \partial_{ss} g(z)}{f(z)}. \quad (2.21)$$

Et d'autre part

$$\begin{aligned}
\mathbb{V}ar[\hat{m}_1(z)] &= \frac{1}{NTH_q^2} \mathbb{V}ar \left[(g(Z_{11}) - g(z)) K \left(\frac{Z_{11} - z}{h} \right) \right] \\
&= (NTH_q^2)^{-1} \left(\mathbb{E} \left\{ \left[(g(Z_{11}) - g(z)) K \left(\frac{Z_{11} - z}{h} \right) \right]^2 \right\} - \left\{ \mathbb{E} \left[(g(Z_{11}) - g(z)) K \left(\frac{Z_{11} - z}{h} \right) \right] \right\}^2 \right) \\
&\leq (NTH_q^2)^{-1} \left(H_q \int f(z + hv) [g(z + hv) - g(z)]^2 K^2(v) dv \right) \\
&= (NTH_q)^{-1} \int_{\mathbb{R}^q} \left\{ \sum_{s,s'} \partial_s g(z) \partial_{s'} g(z) h_s v_s h_{s'} v_{s'} + O(\|h\|^2) \right\} \\
&\times \left\{ f(z) + \sum_{s=1}^q \partial_s f(z) h_s v_s + \frac{1}{2} (hv)^T \text{Hess} f(z) (hv) + O(\|h\|^2) \right\} K^2(v) dv \\
&= O \left((NTH_q)^{-1} \sum_{s=1}^q h_s^2 \right) \tag{2.22}
\end{aligned}$$

On remarque que

$$\mathbb{E} \left[\left| \hat{m}_1(z) - f(z) \sum_{s=1}^q h_s^2 B_s(z) \right|^2 \right] = \left\{ \mathbb{E} \left[\hat{m}_1(z) - f(z) \sum_{s=1}^q h_s^2 B_s(z) \right] \right\}^2 + \mathbb{V}ar(\hat{m}_1(z)), \tag{2.23}$$

et les égalités (2.20) et (2.22) montrent que

$$\mathbb{E} \left[\left| \hat{m}_1(z) - f(z) \sum_{s=1}^q h_s^2 B_s(z) \right|^2 \right] = O \left(\left(\sum_{s=1}^q h_s^2 \right)^2 + (NTH_q)^{-1} \sum_{s=1}^q h_s^2 \right).$$

En appliquant le Théorème 3 du Chapitre 1, on déduit que

$$\hat{m}_1(z) = f(z) \sum_{s=1}^q h_s^2 B_s(z) + O_p \left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2} \left(\sum_{s=1}^q h_s^2 \right)^{1/2} \right). \tag{2.24}$$

Etape 2 : Expression de $\hat{m}_2(z)$

On remarque que $\hat{m}_2(z)$ est d'espérance nulle ; en effet

$$\begin{aligned}
\mathbb{E}(\hat{m}_2(z)) &= \frac{1}{NTH_q} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left(u_{it} K \left(\frac{Z_{it} - z}{h} \right) \right) \\
&= \frac{1}{NTH_q} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left[\mathbb{E} \left(u_{it} K \left(\frac{Z_{it} - z}{h} \right) \mid Z_{it} \right) \right] \\
&= \frac{1}{NTH_q} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left[K \left(\frac{Z_{it} - z}{h} \right) \underbrace{\mathbb{E}(u_{it} \mid Z_{it})}_{=0} \right] \\
&= 0. \tag{2.25}
\end{aligned}$$

Donc

$$\begin{aligned}
\mathbb{E} \{ |\hat{m}_2(z)|^2 \} &= \text{Var} [\hat{m}_2(z)] = (NTH_q^2)^{-1} \mathbb{E} \left[u_{11}^2 K^2 \left(\frac{Z_{11} - z}{h} \right) \right] \\
&= (NTH_q^2)^{-1} \mathbb{E} \left[\mathbb{E} \left(u_{11}^2 K^2 \left(\frac{Z_{11} - z}{h} \right) \mid Z_{11} \right) \right] = (NTH_q^2)^{-1} \mathbb{E} \left[\sigma^2(Z_{11}) K^2 \left(\frac{Z_{11} - z}{h} \right) \right] \\
&= (NTH_q^2)^{-1} \int_{\mathbb{R}^q} \sigma^2(z_{11}) K^2 \left(\frac{z_{11} - z}{h} \right) f(z_{11}) dz_{11} \\
&= (NTH_q)^{-1} \int_{\mathbb{R}^q} \sigma^2(z + hv) K^2(v) f(z + hv) dv \\
&= (NTH_q)^{-1} \left\{ f(z) \sigma^2(z) \int_{\mathbb{R}^q} K^2(v) dv + \mathcal{O} \left(\sum_{s=1}^q h_s^2 \right) \right\} \\
&= (NTH_q)^{-1} \Omega(z) + \mathcal{O} \left((NTH_q)^{-1} \sum_{s=1}^q h_s^2 \right); \tag{2.26}
\end{aligned}$$

avec $\Omega(z) = \kappa^q f(z) \sigma^2(z)$, $\kappa = \int k^2(v) dv$ et $\sigma^2(z) = \mathbb{E} (u_{it}^2 \mid Z_{it} = z)$.

Donc

$$\mathbb{E} \{ [\hat{m}_2(z)]^2 \} = \mathcal{O} \left((NTH_q)^{-1} \left(1 + \sum_{s=1}^q h_s^2 \right) \right).$$

On déduit donc que

$$\hat{m}_2(z) = \mathcal{O}_p \left((NTH_q)^{-1/2} + (NTH_q)^{-1/2} \left(\sum_{s=1}^q h_s^2 \right)^{1/2} \right). \tag{2.27}$$

Etape 3 : Expression de $\hat{m}(z)$

Les égalités (2.24) et (2.27) montrent que

$$\hat{m}(z) = \hat{m}_1(z) + \hat{m}_2(z) = \mathcal{O}_p \left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2} \left(\sum_{s=1}^q h_s^2 \right)^{1/2} + (NTH_q)^{-1/2} \right).$$

Donc

$$\hat{m}(z) = \hat{m}_1(z) + \hat{m}_2(z) = \mathcal{O}_p \left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2} \right). \tag{2.28}$$

Etape 4 : Déduction

Comme $\hat{g}(z) - g(z) = \frac{\hat{m}(z)}{\hat{f}(z)}$ et que $\hat{f}(z) = f(z) + \mathcal{O}_p(1)$, alors

$$\hat{g}(z) - g(z) = \mathcal{O}_p \left(\frac{\hat{m}(z)}{f(z) + \mathcal{O}_p(1)} \right) = \mathcal{O}_p \left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2} \right). \tag{2.29}$$

On a ainsi montré que $\hat{g}(z) - g(z) = \mathcal{O}_p \left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2} \right) = \mathcal{O}_p(1)$.

D'où le résultat.

■

2.4 Convergence uniforme presque sûre et normalité asymptotique de l'estimateur

2.4.1 Normalité asymptotique

Théorème 2

Sous l'hypothèse que z est un point intérieur, $g(z)$ et $f(z)$ sont trois fois continûment différentiables, et $f(z) > 0$, alors lorsque $N \rightarrow \infty$, $h_s \rightarrow 0$ (pour tout s), $NTH_q \rightarrow \infty$ et $\sqrt{NTH_q} \sum_{s=1}^q h_s^2 \rightarrow 0$, nous avons

$$\sqrt{NTH_q} \left(\hat{g}(z) - g(z) - \sum_{s=1}^q h_s^2 B_s(z) \right) \xrightarrow{d} \mathcal{N} \left(0, \kappa^q \frac{\sigma^2(z)}{f(z)} \right), \quad (2.30)$$

où

$$B_s(z) = \frac{\kappa_2}{2} \frac{2\partial_s f(z) \partial_s g(z) + f(z) \partial_{ss} g(z)}{f(z)} \text{ et } \sigma^2(z) = \mathbb{E} \left(u_{it}^2 \mid Z_{it} = z \right). \quad (2.31)$$

Preuve. L'égalité (2.26) montre que

$$\text{Var} \left(\sqrt{NTH_q} \hat{m}_2(z) \right) = \Omega(z) + \mathcal{O} \left(\sum_{s=1}^q h_s^2 \right), \quad (2.32)$$

où on rappelle que $\Omega(z) = \kappa^q f(z) \sigma^2(z)$.

Etape 1 : En utilisant le Théorème central limite de Lyapunov comme au Chapitre 1, montrons que :

$$\sqrt{NTH_q} \hat{m}_2(z) \xrightarrow{d} \mathcal{N} (0, \Omega(z)). \quad (2.33)$$

On a

$$\sqrt{NTH_q} \hat{m}_2(z) = \frac{1}{\sqrt{NTH_q}} \sum_{i=1}^N \sum_{t=1}^T \left[u_{it} K \left(\frac{Z_{it} - z}{h} \right) - 0 \right] = \sum_{i=1}^N \sum_{t=1}^T [Z_{NT,it} - \mathbb{E}(Z_{NT,it})],$$

où $Z_{NT,it} = \frac{1}{\sqrt{NTH_q}} u_{it} K \left(\frac{Z_{it} - z}{h} \right)$ est une suite de v.a indépendantes et $\mathbb{E}(Z_{NT,it}) = 0$.

On a

$$\begin{aligned}\mathbb{V}ar(Z_{NT,it}) &= \frac{1}{NTH_q} \mathbb{E} \left(u_{it}^2 K^2 \left(\frac{Z_{it} - z}{h} \right) \right) \\ &= \frac{1}{NTH_q} \mathbb{E} \left(\sigma^2(Z_{it}) K^2 \left(\frac{Z_{it} - z}{h} \right) \right) = \frac{1}{NTH_q} H_q \left(\Omega(z) + \mathcal{O} \left(\sum_{s=1}^q h_s^2 \right) \right) \\ &= \frac{1}{NT} \left(\Omega(z) + \mathcal{O} \left(\sum_{s=1}^q h_s^2 \right) \right).\end{aligned}$$

Donc

$$\sum_{i=1}^N \sum_{t=1}^T \mathbb{V}ar(Z_{NT,it}) = \Omega(z) + \mathcal{O} \left(\sum_{s=1}^q h_s^2 \right) = \Omega(z) + o(1). \quad (2.34)$$

Aussi,

$$\mathbb{E} \left(|Z_{NT,it}|^{2+1} \right) = \frac{1}{(\sqrt{NTH_q})^3} \mathbb{E}(u_{it}^3) M^3 < \infty; \quad (2.35)$$

et

$$\sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left(|Z_{NT,it}|^{2+1} \right) \leq \frac{1}{\sqrt{NT}(H_q)^{3/2}} \mathbb{E}(u_{it}^3) M^3 \xrightarrow{NT \rightarrow \infty} 0. \quad (2.36)$$

Grâce aux égalités (2.34), (2.35) et (2.36), les hypothèses du TCL de Lyapunov sont vérifiées, et on déduit que

$$\sqrt{NTH_q} \hat{m}_2(z) = \sum_{i=1}^N \sum_{t=1}^T [Z_{NT,it}] \xrightarrow{d} \mathcal{N}(0, \Omega(z)).$$

Etape 2 : Trouvons la loi asymptotique de $\hat{m}(z)$

L'égalité (2.24) et l'hypothèse $\sqrt{NTH_q} \sum_{s=1}^q h_s^2 \rightarrow 0$ montrent que

$$\sqrt{NTH_q} \left(\hat{m}_1(z) - f(z) \sum_{s=1}^q h_s^2 B_s(z) \right) = \mathcal{O}_p \left(\sqrt{NTH_q} \sum_{s=1}^q h_s^2 + \|h\| \right) = o_p(1).$$

De la relation $\hat{f}(z) = f(z) + o_p(1)$, on déduit que

$$\sqrt{NTH_q} \left(\hat{m}_1(z) - \hat{f}(z) \sum_{s=1}^q h_s^2 B_s(z) \right) = \mathcal{O}_p \left(\sqrt{NTH_q} \sum_{s=1}^q h_s^2 + \|h\| \right) = o_p(1). \quad (2.37)$$

Ainsi les égalités (2.33) et (2.37) montrent que

$$\begin{aligned} \sqrt{NTH_q} \left(\hat{m}(z) - f(z) \sum_{s=1}^q h_s^2 B_s(z) \right) &= \underbrace{\sqrt{NTH_q} \hat{m}_2(z) + \sqrt{NTH_q} \left(\hat{m}_1(z) - \hat{f}(z) \sum_{s=1}^q h_s^2 B_s(z) \right)}_{o_p(1)} \\ &= \sqrt{NTH_q} \hat{m}_2(z) + o_p(1) \\ &\xrightarrow{d} \mathcal{N}(0, \Omega(z)); \end{aligned} \quad (2.38)$$

Etape 3 : Dédution

On peut à présent écrire que :

$$\begin{aligned} \sqrt{NTH_q} \left(\hat{g}(z) - g(z) - \sum_{s=1}^q h_s^2 B_s(z) \right) &= \frac{\sqrt{NTH_q} \left(\hat{g}(z) - g(z) - \sum_{s=1}^q h_s^2 B_s(z) \right) \hat{f}(z)}{\hat{f}(z)} \\ &= \frac{\sqrt{NTH_q} \left(\hat{m}(z) - \hat{f}(z) \sum_{s=1}^q h_s^2 B_s(z) \right)}{\hat{f}(z)}. \end{aligned}$$

En utilisant le Théorème SLUTSKY, le fait que $\hat{f}(z) = f(z) + o(1)$ et l'égalité (2.38), on déduit que

$$\sqrt{NTH_q} \left(\hat{g}(z) - g(z) - \sum_{s=1}^q h_s^2 B_s(z) \right) \xrightarrow{d} \frac{1}{f(z)} \mathcal{N}(0, \Omega(z)) = \mathcal{N} \left(0, \frac{\Omega(z)}{(f(z))^2} \right).$$

Or

$$\frac{\Omega(z)}{(f(z))^2} = \frac{\kappa^q f(z) \sigma^2(z)}{(f(z))^2} = \kappa^q \frac{\sigma^2(z)}{f(z)};$$

et on a le résultat voulu.

■

2.4.2 Convergence uniforme presque sûre

De façon similaire lorsque l'on établissait la convergence uniforme presque sûre de l'estimateur de la fonction de densité par la méthode des noyaux, l'on peut également établir cette convergence pour l'estimateur \hat{g} de la fonction de régression g sur un sous ensemble compact \mathcal{S} de \mathbb{R}^q ne contenant pas la frontière du support de Z .

Hypothèse 2.4.1. On suppose que :

- (i) $f(z)$ est différentiable et $g(z)$ est deux fois différentiable et dont les fonctions dérivées satisfont les conditions de Lipschitz $|m(z) - m(v)| \leq C \|z - v\|$ avec $C > 0$ ($m(\cdot) = \partial_{ss} g(\cdot)$)

ou $\partial_s f(\cdot)$.

(ii) $\sigma^2(z) = \mathbb{E}(u_{it} | Z_{it} = z)$ est une fonction continue, et $\inf_{z \in S} f(z) \geq \delta > 0$.

(iii) La fonction noyau $k(\cdot)$ est symétrique et a un support compact. On définit $H_l(v) = \|v\|^l K(v)$ et on suppose que $|H_l(v) - H_l(u)| \leq C\|u - v\|$ pour tout $0 \leq l \leq 3$.

Théorème 3

Sous l'Hypothèse 2.4.1, nous avons

$$\sup_{z \in S} |\hat{g}(z) - g(z)| = O\left(\frac{(\ln NT)^{1/2}}{(NTH_q)^{1/2}} + \sum_{s=1}^q h_s^2\right) \quad p.s. \quad (2.39)$$

Preuve. En maintenant les mêmes notations que dans les sections précédentes, notamment $\hat{g}(z) - g(z) = \frac{\hat{m}(z)}{\hat{f}(z)}$ où $\hat{m}(z) = (\hat{g}(z) - g(z))\hat{f}(z)$. On montre que

$$\sup_{z \in S} |\hat{m}(z)| = O\left(\sum_{s=1}^q h_s^2 + \frac{(\ln NT)^{1/2}}{(NTH_q)^{1/2}}\right) \quad p.s. \quad (2.40)$$

Puisque que la fonction de densité est bornée et qu'on a une convergence presque-sûre de l'estimateur $\hat{f}(z)$ vers $f(z)$ donnée par le **Théorème 4** du **Chapitre 1**, on déduit qu'il existe $\delta' > 0$ tel que

$$\inf_{z \in S} \hat{f}(z) \geq \delta' > 0 \quad p.s.$$

Ainsi

$$\begin{aligned} \sup_{z \in S} |\hat{g}(z) - g(z)| &= \sup_{z \in S} \left| \frac{\hat{m}(z)}{\hat{f}(z)} \right| \leq \frac{\sup_{z \in S} |\hat{m}(z)|}{\inf_{z \in S} \hat{f}(z)} \\ &= O\left(\sum_{s=1}^q h_s^2 + \frac{(\ln NT)^{1/2}}{(NTH_q)^{1/2}}\right) \quad p.s. \end{aligned} \quad (2.41)$$

■

2.5 Intervalle de confiance asymptotique de l'Estimateur $\hat{g}(z)$

On se fixe $\alpha \in [0; 1]$, et on veut construire un intervalle de confiance $[A_n, B_n]$ de niveau $1 - \alpha$ pour $g(z)$ avec $z \in \mathbb{R}^q$, ie

$$\mathbb{P}(g(z) \in [A_n, B_n]) = 1 - \alpha.$$

Etape 1 : Détermination d'un estimateur ponctuel de $g(z)$

L'estimateur ponctuel de $g(z)$ est celui de Nadaraya Watson déjà construit

$$\hat{g}(z) = \frac{\sum_{i=1}^N \sum_{t=1}^T Y_{it} K_{h,it,z}}{\sum_{i=1}^N \sum_{t=1}^T K_{h,it,z}}.$$

Etape 2 : Définissons une fonction asymptotiquement pivotale ie une fonction qui dépend de l'échantillon et de $g(z)$ dont la loi asymptotique ne dépend pas de $g(z)$.

Grâce au **Théorème 2**, la fonction asymptotiquement pivotale que nous considérons est

$$U = \sqrt{NTH_q f(z)} \left(\frac{\hat{g}(z) - g(z) - \sum_{s=1}^q h_s^2 B_s(z)}{\sigma(z) \kappa^{q/2}} \right) \xrightarrow{d} \mathcal{N}(0, 1) \quad (2.42)$$

Etape 3 : Dédution de l'intervalle de confiance

Puisque la fonction pivotale suit la loi normale centrée réduite qui est symétrique, on déduit

$$\mathbb{P}\left(-U_{1-\frac{\alpha}{2}} < U < U_{1-\frac{\alpha}{2}}\right) = 1 - \alpha; \quad (2.43)$$

où $U_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite.

En résolvant l'inéquation $-U_{1-\frac{\alpha}{2}} < U < U_{1-\frac{\alpha}{2}}$ pour obtenir un encadrement de $g(z)$, on aboutit à

$$\hat{g}(z) - \sum_{s=1}^q h_s^2 B_s(z) - \frac{\sigma(z) \kappa^{q/2}}{\sqrt{NTH_q f(z)}} U_{1-\frac{\alpha}{2}} < g(z) < \hat{g}(z) - \sum_{s=1}^q h_s^2 B_s(z) + \frac{\sigma(z) \kappa^{q/2}}{\sqrt{NTH_q f(z)}} U_{1-\frac{\alpha}{2}}. \quad (2.44)$$

Donc un intervalle de confiance de $g(z)$ est

$$IC^\infty(g(z)) = \left[\hat{g}(z) - \sum_{s=1}^q h_s^2 \hat{B}_s(z) - \frac{\sigma(z) \kappa^{q/2}}{\sqrt{NTH_q \hat{f}(z)}} U_{1-\frac{\alpha}{2}}; \hat{g}(z) - \sum_{s=1}^q h_s^2 \hat{B}_s(z) + \frac{\sigma(z) \kappa^{q/2}}{\sqrt{NTH_q \hat{f}(z)}} U_{1-\frac{\alpha}{2}} \right]. \quad (2.45)$$

2.6 Test de linéarité

Dans cette section, nous nous proposons de construire un test de spécification sur la forme de la fonction de lien $g(\cdot)$ comme suggéré par [Henderson, Carroll and Li \(2008\)](#) (11). Plus précisément, nous présentons un test pour tester le modèle paramétrique contre le modèle non paramétrique. Nous considérons les deux cas de spécification suivants, où la fonction $g : \mathbb{R}^{d+q} \rightarrow \mathbb{R}$ est complètement non linéaire et inconnue

$$y_{it} = x'_{it}\beta + z'_{it}\gamma + u_{it}, \quad (2.46)$$

$$y_{it} = g(x_{it}, z_{it}) + u_{it}. \quad (2.47)$$

On suppose que u_{it} suit un one-way, c'est à dire $u_{it} = \mu_i + v_{it}$.

Les hypothèses à tester sont

$$H_0 : (2.46) \text{ versus } H_1 : (2.47);$$

où " $H_0 : (2.46)$ " signifie que le modèle linéaire dans l'égalité (2.46) représente le vrai modèle, c'est l'hypothèse nulle.

" $H_1 : (2.47)$ " signifie que le modèle non linéaire dans l'égalité (2.47) représente le vrai modèle, c'est l'hypothèse alternative.

Nous estimerons le modèle sous l'hypothèse nulle et l'hypothèse alternative et évaluerons la distance entre les modèles estimés. Notamment, pour le modèle (2.46), nous estimons $\tilde{\beta}_G$ et $\tilde{\gamma}_G$ par les MCG, et pour le modèle (2.47), nous estimons $\hat{g}(x, z)$ par la méthode des noyaux ; et la statistique de test est

$$J_N = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[x'_{it} \tilde{\beta}_G + z'_{it} \tilde{\gamma}_G - \hat{g}(x_{it}, z_{it}) \right]^2. \quad (2.48)$$

On rappelle l'expression de $\hat{g}(x, z)$:

$$\hat{g}(x, z) = \frac{\sum_{i=1}^N \sum_{t=1}^T Y_{it} K\left(\frac{Z_{it} - z}{h}\right) K\left(\frac{X_{it} - x}{h'}\right)}{\sum_{i=1}^N \sum_{t=1}^T K\left(\frac{Z_{it} - z}{h}\right) K\left(\frac{X_{it} - x}{h'}\right)} \quad (2.49)$$

où $H_q = h_1 \cdots h_q$ est le paramètre de lissage associé à Z_{it} et $H'_d = h'_1 \cdots h'_d$ est le paramètre de lissage associé à X_{it} .

Trouvons les expressions de $\tilde{\beta}_G$ et $\tilde{\gamma}_G$ par les MCG

On écrit le modèle (2.46) sous forme matricielle et on obtient

$$Y = X\beta + Z\gamma + u \quad (2.50)$$

avec $Y, u \in \mathbb{R}^{NT}$ sont des colonnes, X est une matrice de taille $NT \times d$, Z est une matrice de taille $NT \times q$ et $\beta \in \mathbb{R}^d$ et $\gamma \in \mathbb{R}^q$ sont les vecteurs des paramètres.

En multipliant l'égalité (2.50) par $M = I_{NT} - Z(Z^T Z)^{-1} Z^T$, on remarque que $MZ = 0$, et on obtient

$$MY = MX\beta + Mu. \quad (2.51)$$

A présent la matrice de variance covariance de Mu est donnée par

$$V = \text{Var}(Mu) = M\Sigma M \quad (2.52)$$

où Σ est donné par l'équation (2.6).

Et l'inverse généralisé de V est donné par

$$V^+ = \Sigma^{-1} - \Sigma^{-1} Z (Z^T \Sigma^{-1} Z)^{-1} Z^T \Sigma^{-1}. \quad (2.53)$$

On déduit donc que l'estimateur des moindres carrés généralisés de β est donné par

$$\tilde{\beta}_G = (X^T M V^+ M X)^{-1} X^T M V^+ M Y = (X^T V^+ X)^{-1} X^T V^+ Y = \beta + (X^T V^+ X)^{-1} X^T V^+ u; \quad (2.54)$$

car $M V^+ M = V^+$.

De la même façon on trouve $\tilde{\gamma}_G$, cette fois en multipliant l'équation (2.50) par $Q = I_{NT} - X(X^T X)^{-1} X^T$ pour obtenir le modèle de régression linéaire $QY = QZ\gamma + Qu$, et $U = \text{Var}(Qu) = Q\Sigma Q$ dont l'inverse généralisé est donné par $U^+ = \Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}$

On déduit que $\tilde{\gamma}_G$ est donné par

$$\tilde{\gamma}_G = (Z^T Q U^+ Q Z)^{-1} Z^T Q U^+ Q Y = (Z^T U^+ Z)^{-1} Z^T U^+ Y = \gamma + (Z^T U^+ Z)^{-1} Z^T U^+ u. \quad (2.55)$$

Sous H_0 , J_N converge en probabilité vers 0, et sous H_1 , J_N converge vers une constante positive. Donc J_N peut être utilisé pour détecter si H_0 est vraie ou pas. Pour construire le test, on a besoin de la loi normale asymptotique de la statistique J_N , mais il a été prouvé que la théorie asymptotique ne procure pas de bonnes approximations pour les tests non paramétriques (voir [Härdle and Mammen, 1993](#) ; [Lee and Ullah, 2000](#) ; [Li and Wang, 1998](#) ; [Whang and Andrews, 1993](#)).

Sans trouver la distribution asymptotique de la statistique de test J_N , [Henderson, Carroll and Li \(2008\)](#) (11) ont proposé la méthode du bootstrap pour obtenir la zone critique de rejet et ont démontré à l'aide des simulations que le test proposé marche très bien pour des échantillons finis.

Idée intuitive de la construction du test par Bootstrap

Si α désigne l'erreur de première espèce, on aura tendance à rejeter H_0 lorsque

$$\begin{aligned} \alpha &= \mathbb{P}_{H_0}(\text{re jeter } H_0) = \mathbb{P}(J_N > C) \\ &\approx \frac{1}{M} \sum_{i=1}^M \mathbb{1}_{\{J_N^i > C\}} \end{aligned} \quad (2.56)$$

Donc l'idée serait de calculer les J_N^i à l'aide d'échantillons provenant de l'échantillon de départ et de trouver une constante C qui permettra de conclure sur le test.

Les étapes de la procédure

- On détermine $\tilde{\beta}_G$, $\tilde{\gamma}_G$, et $\hat{g}(x, z)$ grace aux formules (2.54), (2.55) et (2.49) respectivement.
- On calcule

$$\hat{u}_{it} = y_{it} - x_{it}^T \tilde{\beta}_G - z_{it}^T \tilde{\gamma}_G \quad i = 1, \dots, N \quad t = 1, \dots, T.$$

et on pose

$$\hat{u}_i = (\hat{u}_{i1}, \dots, \hat{u}_{iT})^T \quad i = 1, \dots, N.$$

- On effectue un tirage avec répétition dans les $\{\hat{u}_j\}_{j=1}^N$ pour obtenir un nouveau vecteur d'erreurs qu'on note u_i^* $i = 1, \dots, N$.
- On génère donc $y_{it}^* = x_{it}^T \tilde{\beta}_G + z_{it}^T \tilde{\gamma}_G + u_{it}^*$ et l'échantillon $\{x_{it}, z_{it}, y_{it}^*\}$ est appelé **échantillon bootstrap**.

- On utilise donc l'échantillon bootstrap pour estimer β et γ du modèle (2.46) à nouveau par les MCG qu'on note $\tilde{\beta}_G^*$ et $\tilde{\gamma}_G^*$, et pour estimer la fonction $g(\cdot)$ du modèle (2.47) par la méthode des noyaux qu'on note $\hat{g}^*(\cdot)$.
- On détermine alors J_N^* à partir de J_N où l'on remplace $\tilde{\beta}_G$, $\tilde{\gamma}_G$ et $\hat{g}(\cdot)$ respectivement par $\tilde{\beta}_G^*$, $\tilde{\gamma}_G^*$ et $\hat{g}^*(\cdot)$.
- On répète ainsi les étapes précédentes à partir du troisième point M -fois pour obtenir J_N^i $i = 1, \dots, M$ et ainsi trouver la constante C de l'équation (2.56), où les valeurs admissibles de C peuvent être les J_N^i .

APPLICATION ET SIMULATIONS

Dans ce chapitre, il est question de montrer l'applicabilité de tous ce qui a été démontré précédemment, notamment l'estimation des densités des probabilités à partir d'observations, et l'estimation dans les modèles de panel.

3.1 Simulation des fonctions de densités de probabilité

Dans cette section, nous démontrons l'utilité de l'estimation non paramétrique des densités de probabilité par la méthode des noyaux plutôt que par la méthode paramétrique.

3.1.1 Une hypothèse pas valide

Premièrement, nous disposons de la série de données suivante dont nous **ne connaissons pas à priori la loi** :

$$X = \{-0.57, 0.25, -0.08, 1.4, -1.05, -1, 0.37, -1.15, 0.73, 1.59\}.$$

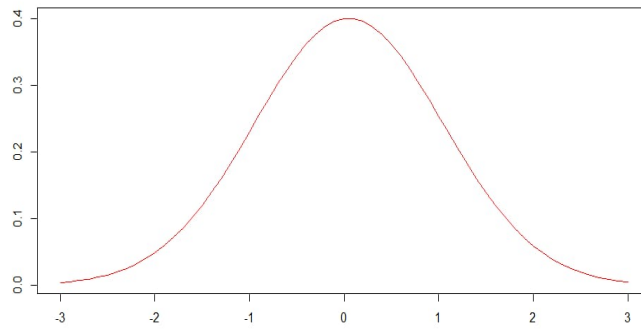
Nous faisons donc l'hypothèse (qui n'est pas fondée) que cette distribution suit une loi normale. Donc une estimation de la fonction de densité de probabilité de cette série passe par l'estimation paramétrique de la moyenne des données et de la variance. On a donc

$$\hat{\mu} = \frac{1}{10} \sum_{i=1}^{10} x_i = 0.049 \text{ et } \hat{\sigma} = \frac{1}{10} \sum_{i=1}^{10} (x_i - \hat{\mu})^2 = 0.99736;$$

Donc on espère que la densité de probabilité de ces observations soit

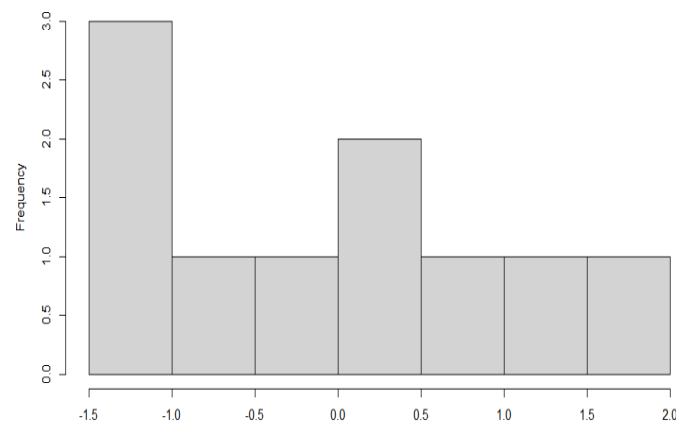
$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}; \quad (3.1)$$

et donc le graphe est donné par :

FIGURE 3.1 – Graphe de la densité f_1

Traçons l'histogramme de cette série de données

classes	$[-1.5, -1[$	$[-1, -0.5[$	$[-0.5, 0[$	$[0, 0.5[$	$[0.5, 1[$	$[1, 1.5[$	$[1.5, 2[$
Fréquence	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

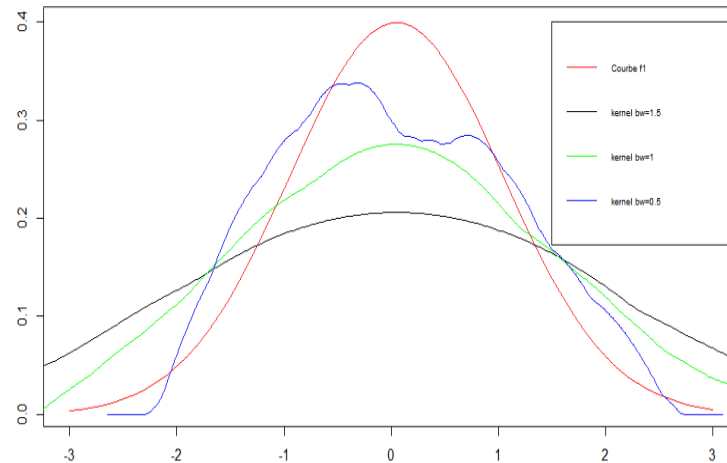
TABLE 3.1 – Table des fréquences des données X FIGURE 3.2 – Histogramme des données X

On remarque que cet histogramme n'a pas l'allure de celle d'une distribution normale. Donc l'hypothèse de normalité faite sur les données semble à priori erronée.

Nous estimerons la densité de probabilité de ces données par la méthode de noyau étudié en utilisant la fonction noyau d'Epanechnikov dont on rappelle l'expression : $k(z) = \frac{3}{4\sqrt{5}}(1 - \frac{z^2}{5})\mathbb{1}_{\{-\sqrt{5} \leq z \leq \sqrt{5}\}}$. La fonction de densité de probabilité estimée s'écrit alors

$$f_2(x) = \frac{1}{10h} \sum_{i=1}^{10} k\left(\frac{x_i - x}{h}\right).$$

On trace le graphe de la fonction f_2 pour le paramètre de lissage $h = 0.5$, $h = 1$ et 1.5 .

FIGURE 3.3 – Le graphe des différentes densités estimées des données X

On se rend compte de l'influence importante du choix du paramètre de lissage h . C'est la courbe en bleu ($h=0.5$) qui semble mettre en exergue les deux pics que possède l'histogramme de ces données. En ce qui concerne l'ordre du choix du paramètre de lissage, des méthodes statistiques ont été développées, mais elles ne constituaient pas notre centre d'intérêt.

3.1.2 Une vérification

Deuxièmement, nous simulons une série de données Y qui suit une loi exponentielle de paramètre 0.5 :

```
#Simulations d'un échantillon de données de taille 1000 et de loi exp(0.5)
Y=rexp(1000,0.5).
```

Donc la vraie densité de ces données Y est la fonction f définie sur \mathbb{R}^+ par

$$f(x) = 0.5 \exp(-0.5x),$$

dont le graphe est donné par

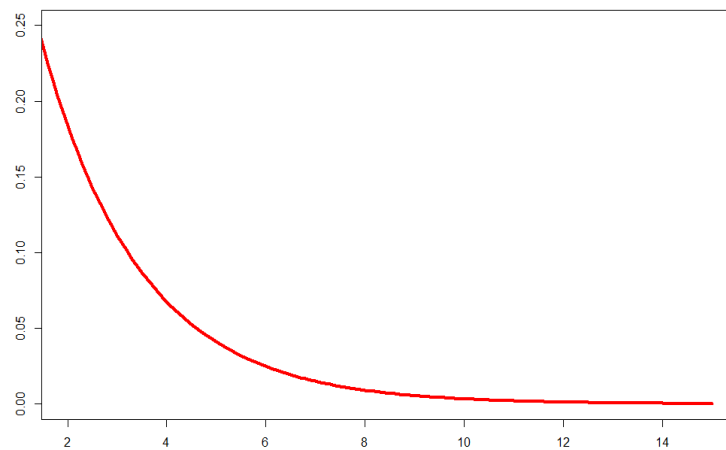
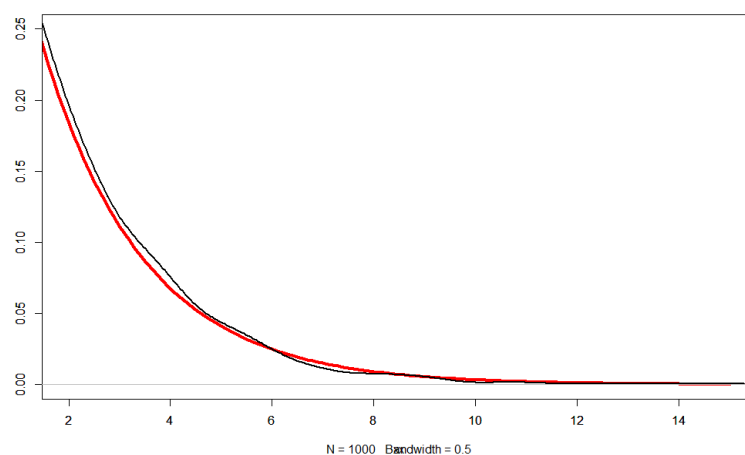


FIGURE 3.4 – Le graphe de densité de la loi exponentielle de paramètre 0.5

Le graphe suivant présente la densité des données Y estimée par la méthode des noyaux (la courbe en noir) et le graphe de la densité exacte précédente :

FIGURE 3.5 – Le graphe de densité des données Y

Les deux graphes semblent avoir la même allure, et même si on ne connaissait pas la loi des données Y au départ, on serait à mesure de dire qu'elles suivent une loi exponentielle grâce au graphe de la densité trouvée.

3.2 Simulations de Monté-Carlo sur un panel

Pour illustrer l'implémentation de l'estimation de $g(z)$ obtenue au **Chapitre 3**, nous ferons une petite simulation sur un ensemble de données finies afin d'en évaluer la performance de l'estimateur.

Dans cette section, nous travaillons avec le modèle suivant proposé par Wang(2003)

$$Y_{it} = \sin(2Z_{it}) + \mu_i + v_{it} \quad (3.2)$$

$$Z_{it} \text{ i.i.d} \sim \mathcal{U}([-1, 1]) \quad (3.3)$$

$$v_{it} \text{ i.i.d} \sim \mathcal{N}(0, 1) \quad (3.4)$$

$$\mu_i \text{ i.i.d} \sim \mathcal{U}([-1, 1]) \quad (3.5)$$

La fonction de lien exacte est $g : [0, 1] \rightarrow \mathbb{R}$ définie par $g(z) = \sin 2z$ pour $z \in [0, 1]$.

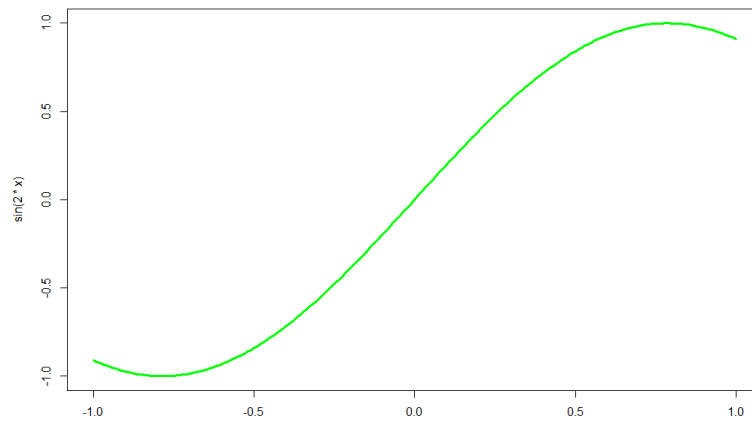


FIGURE 3.6 – Le graphe de la fonction de lien exacte $g(z) = \sin 2z$

Nous utiliserons la fonction noyau Gaussienne $k(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$ et le paramètre de lissage h associé à Z_{it} vaut $h = \hat{\sigma}_z(NT)^{-1/5}$ où $\hat{\sigma}_z$ est l'écart type des observations $\{Z_{it}\}_{i,t=1}^{N,T}$. Les résultats sont présentés pour $T = 3$ et $N = 50, 100, 200$.

On simule les données Z_{it} pour $N = 50, 100, 200$

```
# Simulation des données Z_it
Z_it=runif(3N, -1, 1)
```

Et on a la table des différents écarts types et paramètres de lissage suivant

N	50	100	200
$\hat{\sigma}_z$	0.5455888	0.5815814	0.5890999
h	0.2002844	0.1858602	0.1638924

TABLE 3.2 – Valeurs des écarts types et paramètres de lissage

La fonction $\hat{g}(\cdot)$ estimée par la méthode des noyaux est

$$\hat{g}(z) = \frac{\sum_{i=1}^N \sum_{t=1}^T Y_{it} k\left(\frac{Z_{it} - z}{h}\right)}{\sum_{i=1}^N \sum_{t=1}^T k\left(\frac{Z_{it} - z}{h}\right)} \quad (3.6)$$

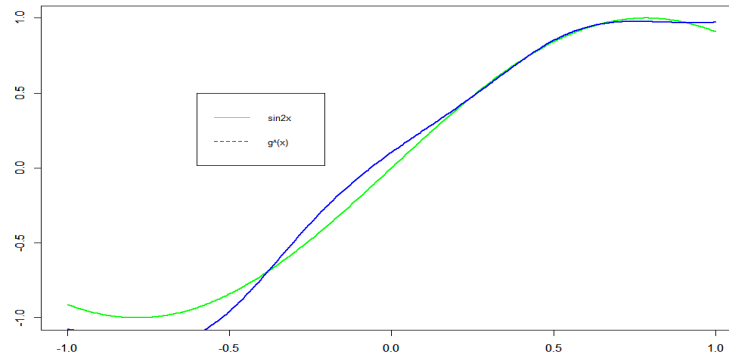


FIGURE 3.7 – $N = 50, T = 3$

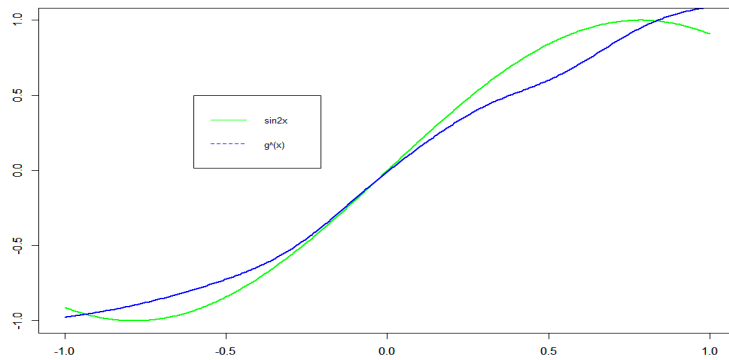


FIGURE 3.8 – $N = 100, T = 3$

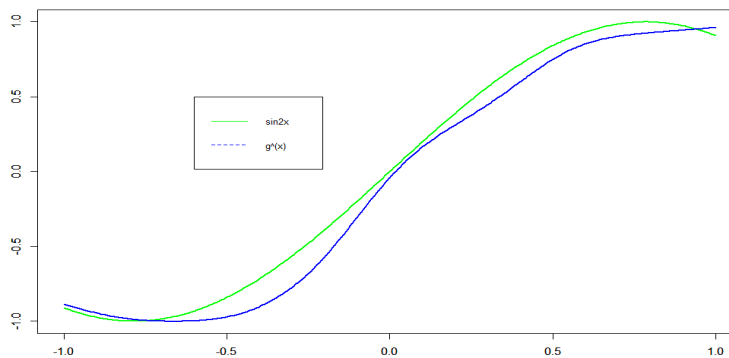


FIGURE 3.9 – $N = 200, T = 3$

Les graphes précédents montrent comment la courbe de la fonction estimée $\hat{g}(\cdot)$ s'apparente davantage à la vraie courbe de la fonction $g(\cdot)$ lorsque N prend des valeurs de plus en plus grandes.

3.3 Test de linéarité

Nous présentons ici la performance du test sur la forme de la fonction de lien dans un modèle de panel. Dans la construction du test, nous considérons les données comme utilisées dans [Henderson, Carroll and Li \(2008\)](#) (11) :

$$Y_{it} = X_{it}\beta + Z_{it}\gamma + u_{it} \quad (3.7)$$

$$Y_{it} = X_{it}^2\beta + Z_{it}^2\gamma + u_{it} \quad (3.8)$$

Dans chacun de ces modèles, X_{it} et Z_{it} sont des variables aléatoires i.i.d réelles de lois respectives $\mathcal{U}([-1, 1])$ et $\mathcal{U}([2, 4])$.

$\beta = 5$ et $\gamma = 2$ et u_{it} est généré comme précédemment.

On utilise toujours la fonction noyau Gaussienne et le paramètre de lissage associé à Z_{it} vaut $h_z = \hat{\sigma}_z(NT)^{-\frac{1}{4+q}}$ et celui associé à X_{it} vaut $h_x = \hat{\sigma}_x(NT)^{-\frac{1}{4+d}}$ où $\hat{\sigma}_z$ et $\hat{\sigma}_x$ représentent les écarts types des observations (Remarque : $q=d=1$).

Dans les simulations reportées dans la table suivante, le nombre de bootstrap est égale à 400. On remarque que la puissance du test dans les trois cas est assez impressionnante et

	$N = 50$			$N=100$			$N=200$		
	1%	5%	10 %	1%	5%	10%	1 %	5%	10%
biais du test	0.09	0.037	0.092	0.011	0.054	0.108	0.010	0.052	0.093
Puissance du test	0.972	0.986	1.000	0.985	1.000	1.000	1.000	1.000	1.000

TABLE 3.3 – Estimation de la puissance du test J_N

tend à être égale à 1 lorsque la taille de l'échantillon N croît. Aussi, ces résultats montrent bien la performance du test pour les données de panel lorsque N est assez large et T fixe.

Conclusion et perspectives

A l'issue de ce travail, nous avons proposé un estimateur non paramétrique de la fonction de lien dans les modèles de panel non linéaires et nous nous sommes ensuite intéressés à ses propriétés statistiques. Nous avons ainsi montré que l'estimateur $\hat{g}(\cdot)$ est consistant de l'ordre de $\mathcal{O}_p\left(\sum_{s=1}^q h_s^2 + (NTH_q)^{-1/2}\right)$ et que $\hat{g}(\cdot)$ converge uniformément presque sûrement vers $g(\cdot)$ avec une vitesse de convergence de l'ordre de $\mathcal{O}\left(\frac{(\ln NT)^{1/2}}{(NTH_q)^{1/2}} + \sum_{s=1}^q h_s^2\right)$.

L'originalité de ce travail repose beaucoup plus sur les détails apportés sur la compréhension et la construction du test statistique de linéarité proposé pour tester la forme des fonctions de lien. En effet, dans [Nonparametric Econometrics Theory and Practice](#) de Qi Li and Jeffrey Scott Racine (2007) (13), la plupart des résultats ont été démontrés sous les mêmes hypothèses à la seule différence que, plutôt que de supposer $(nh_1 \cdots h_q) \sum_{s=1}^q h_s^6 \rightarrow 0$, on a supposé que $(nh_1 \cdots h_q) \sum_{s=1}^q h_s^3 \rightarrow 0$. De plus, il a été présenté ici un intervalle de confiance de l'estimateur de $g(\cdot)$ ainsi qu'un test sur sa forme. Ce test a des portées pratiques sur plusieurs domaines. En effet, il y'a des domaines dans lesquels les formes de la fonction de lien sont bien connues, et il serait judicieux lorsqu'on dispose des données de tester effectivement que cette fonction s'ajuste sur les données. L'application du modèle sur les données simulées a permis de montrer l'applicabilité et de donner un cadre plus facile pour les applications.

Le modèle analysé dans ce travail prend en compte l'effet individuel mais pas l'effet temps ; il serait intéressant de regarder si en prenant en compte l'effet temps dans le modèle les estimations sont améliorées ou pas.

Bibliographie

- [1] Chunrong Ai and Xiaohong Chen. The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, 170(2) :442–457, 2012.
- [2] Takeshi Amemiya. The estimation of the variances in a variance-components model. *International Economic Review*, pages 1–13, 1971.
- [3] Donald WK Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica : Journal of the Econometric Society*, pages 307–345, 1991.
- [4] Badi H Baltagi. *Econometric analysis of panel data*. Springer Nature, 2021.
- [5] Badi H Baltagi, Dong Li, et al. Series estimation of partially linear panel data models with fixed effects. *Annals of economics and finance*, 3(1) :103–116, 2002.
- [6] A Colin Cameron and Pravin K Trivedi. *Microeconometrics : methods and applications*. Cambridge university press, 2005.
- [7] Luc P Devroye. The uniform convergence of the nadaraya-watson regression function estimate. *Canadian Journal of Statistics*, 6(2) :179–191, 1978.
- [8] Boris Vladimirovich Gnedenko. *The theory of probability*. CUP Archive, 1962.
- [9] Bruce E Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3) :726–748, 2008.
- [10] Nancy E Heckman. Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society : Series B (Methodological)*, 48(2) :244–248, 1986.

- [11] Daniel J Henderson, Raymond J Carroll, and Qi Li. Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1) :257–275, 2008.
- [12] Pascal Lavergne and Alban Thomas. Semiparametric estimation and testing in a model of environmental regulation with adverse selection. *Empirical Economics*, 30(1) :171–192, 2005.
- [13] Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics : theory and practice*. Princeton University Press, 2007.
- [14] Qi Li and Thanasis Stengos. Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. *International Economic Review*, pages 981–1000, 1994.
- [15] Oliver Linton. Second order approximation in the partially linear regression model. *Econometrica : Journal of the Econometric Society*, pages 1079–1112, 1995.
- [16] Elias Masry. Multivariate local polynomial regression for time series : uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6) :571–599, 1996.
- [17] EA Nadaraya. On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, 10(1) :186–190, 1965.
- [18] Whitney K Newey and T Chipty. *Efficient estimation of semiparametric models via moment restrictions*. Econometric Research Program, Princeton University, 1990.
- [19] Calyampudi Radhakrishna Rao, Calyampudi Radhakrishna Rao, Mathematischer Statistiker, Calyampudi Radhakrishna Rao, and Calyampudi Radhakrishna Rao. *Linear statistical inference and its applications*, volume 2. Wiley New York, 1973.
- [20] Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica : Journal of the Econometric Society*, pages 931–954, 1988.
- [21] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.
- [22] Paul Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society : Series B (Methodological)*, 50(3) :413–436, 1988.

- [23] Adonis Yatchew. An elementary estimator of the partial linear model. *Economics letters*, 57(2) :135–143, 1997.
- [24] Adonis Yatchew. *Semiparametric regression for the applied econometrician*. Cambridge University Press, 2003.

Rappels

A.1 Inégalités et Modes de convergence

A.1.1 Inégalités

Lemme 1

Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^q et définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, $(a_n)_n$ une suite de nombres positifs. Si $\sum_{n=1}^{\infty} \mathbb{P}(\|X_n\| > a_n)$ est fini, alors

$$X_n \leq a_n \text{ presque sûrement ou } X_n = O(a_n) \text{ presque sûrement} \quad (\text{A.1})$$

Preuve. Posons $A_n = \{\|X_n\| > a_n\}$. Alors $(A_n)_{n \geq 1}$ est une suite d'événements sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ telle que $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.

Etape 1 : Montrons que si $(u_n)_n$ est une suite de nombres réels telle que $\sum_{n=1}^{\infty} u_n < \infty$, alors $R_n = \sum_{k \geq n} u_k \xrightarrow{n \rightarrow \infty} 0$.

Posons $S_n = \sum_{k=1}^n u_k$; alors $\lim_{n \rightarrow \infty} S_n$ existe et est finie, donc la suite $(S_n)_n$ est une suite de

Cauchy. Donc pour tout $p, q \in \mathbb{N}$ tels que $N < p < q$, on a $|S_q - S_p| = \left| \sum_{k=p+1}^q u_k \right| < \varepsilon$.

En faisant tendre q vers ∞ , on déduit que $\forall p \geq N$, on a $\left| \sum_{k \geq p+1} u_k \right| < \varepsilon$.

On a donc montré que $\forall \varepsilon > 0$, $\exists n_0, n_0 = N + 1$, tel que $\forall n \geq N + 1$ on a $|R_n| < \varepsilon$. Ce qui montre que $R_n \xrightarrow{n \rightarrow \infty} 0$.

On déduit donc $\lim_{m \rightarrow \infty} \sum_{k \geq m} \mathbb{P}(A_n) = 0$.

Etape 2 : Montrons que $\mathbb{P}\left(\overline{\lim}_n A_n\right) = 0$ où $\overline{\lim}_n A_n = \limsup_n A_n = \bigcap_{n \geq 1} \left(\bigcup_{k \geq n} A_k \right)$.

On a pour tout $m \geq 1$,

$$\begin{aligned} 0 \leq \mathbb{P}\left(\overline{\lim}_n A_n\right) &\leq \mathbb{P}\left(\bigcup_{k \geq m} A_k\right) \\ &\leq \sum_{k \geq m} \mathbb{P}(A_k). \end{aligned} \quad (\text{A.2})$$

On déduit donc que $0 \leq \mathbb{P}\left(\overline{\lim}_n A_n\right) \leq \lim_{m \rightarrow \infty} \sum_{k \geq m} \mathbb{P}(A_k) = 0$.

Donc $\mathbb{P}\left(\overline{\lim}_n A_n\right) = 0$.

Etape 3 : Montrons que $\mathbb{P}\left(\liminf_n A_n\right) \leq \liminf_n \mathbb{P}(A_n) \leq \limsup_n \mathbb{P}(A_n) \leq \mathbb{P}\left(\overline{\lim}_n A_n\right)$

On rappelle d'abord la définition suivante :

$$\underline{\lim}_n A_n = \liminf_n A_n = \bigcup_{n \geq 1} \left(\bigcap_{k \geq n} A_k \right).$$

Comme la suite $\left(\bigcap_{k \geq n} A_k \right)_n$ est une suite croissante, par continuité croissante de la probabilité \mathbb{P} , on a

$$\mathbb{P}\left(\liminf_n A_n\right) = \mathbb{P}\left(\bigcup_{n \geq 1} \left(\bigcap_{k \geq n} A_k \right)\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k \geq n} A_k\right). \quad (\text{A.3})$$

Or

$$\mathbb{P}\left(\bigcap_{k \geq n} A_k\right) \leq \mathbb{P}(A_k) \quad \forall k \geq n,$$

en passant au inf sur le membre de droite, on déduit que

$$\mathbb{P}\left(\bigcap_{k \geq n} A_k\right) \leq \inf_{k \geq n} \mathbb{P}(A_k).$$

En passant à la limite sur cette dernière inégalité, on obtient $\lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k \geq n} A_k\right) \leq \lim_{n \rightarrow \infty} \inf_{k \geq n} \mathbb{P}(A_k)$.

On déduit donc de l'égalité (A.3) que $\mathbb{P}\left(\liminf_n A_n\right) \leq \liminf_n \mathbb{P}(A_n)$.

D'autre part, on a toujours

$$\inf_{k \geq n} \mathbb{P}(A_k) \leq \sup_{k \geq n} \mathbb{P}(A_k),$$

donc $\liminf_n \mathbb{P}(A_k) \leq \limsup_n \mathbb{P}(A_k)$.

A nouveau, on remarque que la suite $\left(\bigcup_{k \geq n} A_k \right)_n$ est une suite décroissante et en utilisant la

continuité décroissante de la probabilité \mathbb{P} , donc

$$\mathbb{P}\left(\limsup_n A_n\right) = \mathbb{P}\left(\bigcap_{n \geq 1} \left(\bigcup_{k \geq n} A_k\right)\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right). \quad (\text{A.4})$$

Or

$$\mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \geq \mathbb{P}(A_k) \quad \forall k \geq n,$$

donc en passant au sup on obtient $\mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \geq \sup_{k \geq n} \mathbb{P}(A_k)$.

D'où

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \geq \lim_{n \rightarrow \infty} \sup_{k \geq n} \mathbb{P}(A_k).$$

Et d'après l'égalité (A.4), on déduit que $\limsup_n \mathbb{P}(A_n) \leq \mathbb{P}\left(\overline{\lim}_n A_n\right)$.

Etape 4 : Déduction

On a $\mathbb{P}\left(\limsup_n A_n\right) = \mathbb{P}\left(\liminf_n A_n\right) = 0$, donc $\mathbb{P}(A_n) = 0$.

■

Lemme 2 (Inégalité de Markov)

Soient X une variable aléatoire à valeur dans \mathbb{R}^q de fonction de densité f et $\phi : \mathbb{R}^q \rightarrow \mathbb{R}$ une fonction telle que $\phi(x) \geq 0$ pour tout $x \in \mathbb{R}^q$. Soit $A \subset \mathbb{R}^q$ une partie non vide de \mathbb{R}^q et posons $\phi_A = \inf_{x \in A} \phi(x)$, alors

$$\phi_A \mathbb{P}(X \in A) \leq \mathbb{E}[\phi(X)]. \quad (\text{A.5})$$

Preuve. Si $\phi_A = 0$, l'inégalité reste vraie car la fonction ϕ est positive et donc $\mathbb{E}[\phi(X)] \geq 0$.

Dans le cas contraire, on peut écrire

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}^q} f(x)\phi(x)dx \geq \int_A f(x)\phi(x)dx \geq \phi_A \int_A f(x)dx = \phi_A \mathbb{P}(X \in A). \quad (\text{A.6})$$

■

Remarque A.1.1. Notons que si $\phi : \mathbb{R} \rightarrow \mathbb{R}$, $\phi(x) = |x|^k$ et $A = \{x \in \mathbb{R} \mid |x| \geq \alpha\}$ (avec $\alpha > 0$), alors on a $\phi_A = \alpha^k$, $\mathbb{P}(X \in A) = \mathbb{P}(|X| \geq \alpha)$ et $\mathbb{E}[\phi(X)] = \mathbb{E}[|X|^k]$. Donc l'égalité (A.6) devient

$$\mathbb{P}(|X| \geq \alpha) \leq \frac{\mathbb{E}[|X|^k]}{\alpha^k}, \quad (\text{A.7})$$

qui est sous la forme la plus connue et la plus utilisée.

Remarque A.1.2. Notons que si $\phi : \mathbb{R} \rightarrow \mathbb{R}$, $\phi(x) = \exp(ax)$ et $A = \{x \in \mathbb{R} / |x| \geq c\}$ (avec $a > 0$ et $c > 0$), alors on a $\phi_A = \exp(ac)$, $\mathbb{P}(X \in A) = \mathbb{P}(X \geq c)$ et $\mathbb{E}[\phi(X)] = \mathbb{E}[\exp(aX)]$. Donc l'égalité (A.6) devient donc

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}[\exp(aX)]}{\exp(ac)}. \quad (\text{A.8})$$

Lemme 3 (Inégalité de Chebychev)

Soit $g(\cdot)$ une fonction mesurable positive sur \mathbb{R} qui est croissante sur l'intervalle $]0, +\infty[$ et paire. Alors pour toute variable aléatoire X à valeurs dans \mathbb{R} et pour $\varepsilon > 0$, on a

$$\mathbb{P}[|X| > \varepsilon] \leq \frac{\mathbb{E}[g(X)]}{g(\varepsilon)}. \quad (\text{A.9})$$

Preuve. Puisque g est croissante sur $]0, +\infty[$, si $|X| > \varepsilon$, alors $g(|X|) > g(\varepsilon)$; et comme g est paire, on déduit que $g(X) > g(\varepsilon)$. Ainsi on peut écrire que

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X) dP_X \geq \int_{|X| > \varepsilon} g(X) dP_X \geq g(\varepsilon) \int_{|X| > \varepsilon} dP_X = g(\varepsilon) \mathbb{P}(|X| > \varepsilon).$$

■

Lemme 4 (Inégalité de Hölder)

Soient X_1 et X_2 deux variables aléatoires. Pour $p > 1$ et q son conjugué associé, ie, $\frac{1}{p} + \frac{1}{q} = 1$

$$\mathbb{E}[|X_1 X_2|] \leq \{\mathbb{E}[|X_1|^p]\}^{1/p} \{\mathbb{E}[|X_2|^q]\}^{1/q}. \quad (\text{A.10})$$

Preuve. On donne juste les grandes lignes de cette démonstration.

- Si $X_1 = 0$ p.s ou $X_2 = 0$ p.s, alors $X_1 X_2 = 0$ p.s et l'égalité est vraie.
- Supposons que $\{\mathbb{E}[|X_1|^p]\}^{1/p} > 0$ et $\{\mathbb{E}[|X_2|^q]\}^{1/q} > 0$.

Par la concavité de la fonction $x \mapsto \ln x$, on a

$$a \ln y + b \ln x \leq \ln(ay + bx),$$

donc

$$y^a x^b \leq ay + bx. \quad (\text{A.11})$$

On prend $a = \frac{1}{p}$, $b = \frac{1}{q}$, $y = \frac{|X_1|^p}{\mathbb{E}|X_1|^p}$ et $x = \frac{|X_2|^q}{\mathbb{E}|X_2|^q}$; et en appliquant l'inégalité (A.11), on a

$$\begin{aligned} y^{1/p} x^{1/q} &= \frac{|X_1|}{\{\mathbb{E}|X_1|^p\}^{1/p}} \frac{|X_2|}{\{\mathbb{E}|X_2|^q\}^{1/q}} \\ &\leq \frac{1}{p} \frac{|X_1|^p}{\mathbb{E}|X_1|^p} + \frac{1}{q} \frac{|X_2|^q}{\mathbb{E}|X_2|^q}. \end{aligned}$$

En passant à l'espérance, on obtient

$$\frac{\mathbb{E}\{|X_1 X_2|\}}{\{\mathbb{E}|X_1|^p\}^{1/p} \{\mathbb{E}|X_2|^q\}^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1.$$

Ce qu'il fallait démontrer.

■

Pour $p = q = 2$, on obtient l'**Inégalité de Cauchy-Schwartz**.

A.1.2 Modes de convergence

Soit $\{X_n\}_{n=0}^\infty$ une suite de variables aléatoires et soit X une variable aléatoire de même dimension que X_n (à valeur dans \mathbb{R}^q ou \mathbb{R}) définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 1 (Convergence en probabilité)

On dit que X_n converge en probabilité vers X et on note $X_n \xrightarrow{\mathbb{P}} X$ si

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}(\|X_n - X\| < \varepsilon) = 1. \quad (\text{A.12})$$

Définition 2

On dit que X_n converge en moyenne d'ordre r , où $r > 0$, et on note $X_n \xrightarrow{r^{\text{th}}} X$ si

$$\lim_{n \rightarrow \infty} \mathbb{E}\{\|X_n - X\|^r\} = 0. \quad (\text{A.13})$$

Remarque A.1.3. Si $r = 2$, on parle de convergence en moyenne quadratique et on note encore $X_n \xrightarrow{MSE} X$.

Lemme 5

La convergence en moyenne d'ordre r ($r > 0$) entraîne la convergence en probabilité ; ie,

$$\text{Si } X_n \xrightarrow{r^{\text{th}}} X \text{ alors } X_n \xrightarrow{\mathbb{P}} X. \quad (\text{A.14})$$

Preuve. En utilisant l'inégalité de Chebychev, on a

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \leq \frac{1}{\varepsilon^r} \mathbb{E}(\|X_n - X\|^r) \xrightarrow{n \rightarrow \infty} 0. \quad (\text{A.15})$$

■

Donc pour souvent montrer la convergence en probabilité, il sera plus pratique de montrer la convergence en moyenne quadratique.

Exemple A.1.1. Soit Y_n une suite de variables aléatoires i.i.d de loi $\mathcal{N}(0, 1)$. On pose $X_n = \frac{Y_n}{n}$; alors $X_n \xrightarrow{r^{\text{th}}} 0$. En effet

$$\mathbb{E}((X_n - 0)^2) = \frac{1}{n^2} \mathbb{E}(Y_n^2) = \frac{1}{n^2} \xrightarrow{n \rightarrow \infty} 0. \quad (\text{A.16})$$

Définition 3 (Convergence en distribution)

On dit que X_n converge en distribution vers X et on note $X_n \xrightarrow{d} X$ si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad (\text{A.17})$$

pour tout point x en lequel F est continue où F_n et F sont respectivement les fonctions de répartition de X_n et X .

Définition 4 (Convergence presque-sûre)

On dit que X_n converge presque-sûrement ($p.s$) vers X et on note $X_n \xrightarrow{p.s} X$ ou encore $X_n \rightarrow X$ $p.s$ si

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1. \quad (\text{A.18})$$

Définition 5 (Convergence faible)

Une suite de variables aléatoires $Z_n(\cdot)$ converge faiblement vers $Z(\cdot)$ si

$$\mathbb{E}(g(Z_n)) \longrightarrow \mathbb{E}(g(Z)) \quad (\text{A.19})$$

pour toute fonction continue bornée $g(\cdot)$.

Théorème 1 (Théorème de convergence dominée)

Soit $(g_n)_{n \geq 0}$ une suite de fonctions mesurables et intégrables définies sur \mathcal{S} . On suppose qu'il existe une fonction m définie sur \mathcal{S} telle que $|g_n(x)| \leq m(x)$ p.s pour tout $x \in \mathcal{S}$ et que $\int_{\mathcal{S}} m(x) < \infty$. Si $\lim_{n \rightarrow +\infty} g_n(x) = g(x)$ pour tout $x \in \mathcal{S}$, alors

$$\lim_{n \rightarrow +\infty} \int_{\mathcal{S}} g_n(x) dx = \int_{\mathcal{S}} \left[\lim_{n \rightarrow +\infty} g_n(x) \right] dx = \int_{\mathcal{S}} g(x) dx. \quad (\text{A.20})$$

Preuve. Voir [Rao \(19\)](#) ■

Théorème 2 (Théorème de convergence dominée version stochastique)

Soit $\{X_n\}_{n=0}^{\infty}$ une suite de variables aléatoires. On suppose que $X_n \xrightarrow{\mathbb{P}} X$ et qu'il existe une variable aléatoire Y telle que $|X_n| \leq Y$ p.s et $\mathbb{E}(Y) < \infty$, alors

$$\mathbb{E}\{|X_n - x|^r\} \longrightarrow 0 \quad \text{et} \quad \mathbb{E}(X_n^r) \longrightarrow \mathbb{E}(X^r). \quad (\text{A.21})$$

Preuve. Voir [Rao \(19\)](#) ■

Définition 6

Soient $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites de nombres réels, avec $b_n > 0$ pour tout $n \in \mathbb{N}$.

- i) On écrit $a_n = \mathcal{O}(1)$ si $(a_n)_{n \in \mathbb{N}}$ reste bornée lorsque $n \rightarrow \infty$, i.e, $|a_n| \leq C \forall n \in \mathbb{N}$ avec C étant une constante.

De façon similaire, on écrit $a_n = \mathcal{O}(b_n)$ si $\frac{a_n}{b_n} = \mathcal{O}(1)$ ou encore $a_n \leq C b_n \forall n \in \mathbb{N}$ et avec C constant.

- ii) On écrit $a_n = o(1)$ si la suite $(a_n)_{n \in \mathbb{N}}$ converge vers 0, ie, $a_n \xrightarrow{n \rightarrow \infty} 0$.

On écrit $a_n = o(b_n)$ si $\frac{a_n}{b_n} \xrightarrow{n \rightarrow \infty} 0$.

Exemple A.1.2. 1. Si $a_n = \frac{n}{n+1}$, puisque $a_n \leq 1 \forall n$, alors $a_n = \mathcal{O}(1)$.

2. Si $a_n = \frac{10}{n+1}$, alors $a_n = o(1)$.

3. Si $b_n = n + 5$ et $a_n = n$ alors $a_n = \mathcal{O}(b_n)$ car pour tout $n \in \mathbb{N}$, $a_n \leq b_n$.

4. Si $a_n = \frac{1}{n^2}$ et $b_n = \frac{1}{n}$, alors $a_n = o(b_n)$ car $\frac{a_n}{b_n} = \frac{1}{n} \rightarrow 0$.

Exemple A.1.3. Grâce aux inégalités $\sqrt{a_n + b_n} \leq \sqrt{a_n} + \sqrt{b_n} \leq 2\sqrt{a_n + b_n}$, on a

$$\mathcal{O}(\sqrt{a_n + b_n}) = \mathcal{O}(\sqrt{a_n}) + \mathcal{O}(\sqrt{b_n}).$$

Remarque A.1.4. Les notions de o et \mathcal{O} sont aussi valables pour des fonctions au voisinage d'un point, en particulier si f et g sont deux fonctions réelles définies sur \mathbb{R} et $x_0 \in \mathbb{R}$,

$$f = o(g) \iff \exists h \mid f = gh \text{ avec } \lim_{x \rightarrow x_0} h(x) = 0,$$

$$f = \mathcal{O}(g) \iff \exists h \mid f = gh \text{ avec } h \text{ bornée au voisinage de } x_0.$$

Et on a les propriétés suivantes :

$$o(f + g) = o(f) + o(g)$$

$$o(\lambda f) = o(f)$$

$$o(\mathcal{O}(f)) = o(f)$$

$$\mathcal{O}(o(f)) = o(f).$$

Définition 7

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^q ou \mathbb{R} .

- i) On dit que la suite $(X_n)_{n \geq 0}$ est bornée en probabilité si pour tout $\varepsilon > 0$, il existe un réel $M_\varepsilon \equiv M$ et un entier $N_\varepsilon \equiv N$ tels que :

$$\mathbb{P}(\|X_n\| > M) \leq \varepsilon \quad \forall n \geq N. \quad (\text{A.22})$$

L'on écrit $X_n = \mathcal{O}_p(1)$ pour indiquer que la suite $(X_n)_{n \geq 0}$ est bornée en probabilité.

- ii) On écrit $X_n = o_p(1)$ si $X_n \xrightarrow{\mathbb{P}} 0$, i.e, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|X_n - 0\| < \varepsilon) = 1. \quad (\text{A.23})$$

Remarque A.1.5. Si $X_n = \mathcal{O}(1)$, alors $X_n = \mathcal{O}_p(1)$.

Mais la réciproque est fausse. En effet, en considérant la suite $\{X_n\}_{n=1}^\infty$ de variables aléatoires i.i.d de loi $\mathcal{N}(0, 1)$, on a $X_n = \mathcal{O}_p(1)$ mais $X_n \neq \mathcal{O}(1)$.

Pour souvent montrer qu'une suite de variables aléatoires est bornée en probabilité, on pourra se servir du théorème suivant :

Théorème 3

Soit $\{X_n\}_{n=1}^\infty$ une suite de variables aléatoires, et soient a_n et b_n deux suites de nombres positifs non aléatoires. Alors :

- (i) Si $\mathbb{E}(\|X_n\|) = \mathcal{O}(a_n)$, alors $X_n = \mathcal{O}_p(a_n)$.
(ii) Si $\mathbb{E}(\|X_n\|^2) = \mathcal{O}(b_n)$, alors $X_n = \mathcal{O}_p(b_n^{1/2})$.

Preuve. (i) Si $\mathbb{E}(\|X_n\|) = \mathcal{O}(a_n)$, alors il existe $M_0 > 0$ tel que $0 \leq \mathbb{E}(\|X_n\|) \leq M_0 a_n$.

Donc $a_n \geq 0$ et on peut écrire que $\mathbb{E}\left(\left\|\frac{X_n}{a_n}\right\|\right) \leq M_0$.

Soit $\varepsilon > 0$. Pour $M = \frac{M_0}{\varepsilon}$ et par l'inégalité de Markov on a :

$$\mathbb{P}\left(\left\|\frac{X_n}{a_n}\right\| > M\right) \leq \frac{\mathbb{E}\left(\left\|\frac{X_n}{a_n}\right\|\right)}{M} \leq \varepsilon.$$

D'où le résultat.

(ii) De la même façon, puisque $\mathbb{E}(\|X_n\|^2) = \mathcal{O}(b_n)$, alors il existe $M_1 > 0$ tel que $\mathbb{E}\left(\left\|\frac{X_n}{b_n^{1/2}}\right\|^2\right) \leq M_1$.

Soit $\varepsilon > 0$. Pour $M = \sqrt{\frac{M_1}{\varepsilon}}$ et par l'inégalité de Chebychev, on a :

$$\mathbb{P}\left(\left\|\frac{X_n}{b_n^{1/2}}\right\| > M\right) \leq \frac{\mathbb{E}\left(\left\|\frac{X_n}{b_n^{1/2}}\right\|^2\right)}{M^2} \leq \varepsilon.$$

■

Remarque A.1.6. Si $X_n = o_p(1)$ alors, $X_n = \mathcal{O}_p(1)$.

Mais la réciproque est fausse.

Exemple A.1.4. (i) Soit $\{X_n\}_{n=1}^\infty$ une suite de variables aléatoires telle que $\mathbb{E}|X_n| \leq C < \infty$ ou $\mathbb{E}(\|X_n\|^2) \leq C < \infty$ pour tout $n \in \mathbb{N}$, alors $X_n = \mathcal{O}_p(1)$.

(ii) Soit $\{X_n\}_{n=1}^\infty$ telle que $\mathbb{E}(X_n) = o(1)$ et $\text{Var}(X_n) = o(1)$, alors $X_n = o_p(1)$. En effet,

$$\mathbb{E}(\|X_n\|^2) = \text{tr}\{\mathbb{E}(X_n X_n')\} = \mathbb{E}(X_n)\mathbb{E}(X_n') + \text{tr}(\text{Var}(X_n)) = o(1).$$

Donc $X_n \xrightarrow{MSE} 0$, et comme la convergence en moyenne quadratique implique la convergence en probabilité, on déduit que $X_n \xrightarrow{\mathbb{P}} 0$; ie, $X_n = o_p(1)$.

A.2 Loi faible des grands nombres et Théorème central limite

Lemme 6 (Loi des grands nombres de Khinchin)

Si X_1, \dots, X_n est une suite i.i.d d'observations de moyenne $\mu < \infty$, alors

$$\bar{X}_n \equiv \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \mathbb{E}(X_1) = \mu \quad (\text{A.24})$$

Preuve. Si $\text{Var}(X_i) = \sigma < \infty$, nous allons montrer la convergence en moyenne quadratique

de \bar{X}_n vers μ ($\bar{X}_n \xrightarrow{MSE} \mu$). On a

$$\begin{aligned}\mathbb{E}[(\bar{X}_n - \mu)^2] &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[(X_i - \mu)(X_j - \mu)] \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^n \mathbb{E}[(X_i - \mu)^2] + \sum_{i=1}^n \sum_{j=1}^n \underbrace{\text{Cov}(X_i, X_j)}_{=0} \right\} \\ &= \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.\end{aligned}$$

Donc $\bar{X}_n \xrightarrow{MSE} \mu$ et on déduit que $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$.

Si $\mathbb{V}ar(X_i)$ n'est pas finie, voir [Rao](#)[(19), P.112]

■

Lemme 7 (Théorème central limite de Lindeberg-Levy)

Si X_1, \dots, X_n est une suite i.i.d d'observations de moyenne $\mu < \infty$, et de variance σ^2 alors

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n X_i - \mu \right\} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad (\text{A.25})$$

Preuve. voir [Rao](#)[(19), P.127] ■

Lemme 8 (Théorème central limite de Lindeberg-Feller)

Soit X_1, \dots, X_n une suite d'observations indépendantes avec $\mathbb{E}(x_i) = \mu_i$, $\mathbb{V}ar(X_i) = \sigma_i^2$ et de fonction de répartition $F_i(\cdot)$. On définit $\bar{\sigma}_n = \left[\frac{1}{n} \sum_{i=1}^n \sigma_i^2 \right]^{1/2}$. Si

$$\lim_{n \rightarrow \infty} \left\{ \max_{1 \leq i \leq n} \frac{\sigma_i}{\sqrt{n} \bar{\sigma}_n} \right\} = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{1}{n \bar{\sigma}_n^2} \sum_{i=1}^n \int_{|x - \mu_i| > \varepsilon \bar{\sigma}_n \sqrt{n}} (x - \mu_i)^2 dF_i(x) = 0 \quad (\text{A.26})$$

alors

$$\frac{1}{\sqrt{n} \bar{\sigma}_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} \mathcal{N}(0, 1). \quad (\text{A.27})$$

Preuve. Voir [Rao](#) [(19), P.127] ■

Lemme 9 (Théorème central limite de Lyapunov)

Soit $\{Z_{n,i}\}$ une suite de variables aléatoires indépendantes telle que $\mathbb{E}(Z_{n,i}) = \mu_{n,i}$ et $\text{Var}(Z_{n,i}) = \sigma_{n,i}^2$ avec $\mathbb{E}|Z_{n,i}|^{2+\delta} < \infty$ où $\delta > 0$. Soit $S_n = \sum_{i=1}^n Z_{n,i}$, et soit $\sigma_n^2 = \sum_{i=1}^n \sigma_{n,i}^2$. Si $\sigma_n^2 = \sigma^2 + o(1)$ (σ^2 étant une constante), et

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}|Z_{n,i} - \mu_{n,i}|^{2+\delta} = 0 \text{ pour un certain } \delta > 0, \quad (\text{A.28})$$

alors

$$\sigma_n^{-1} (S_n - \mathbb{E}(S_n)) = \sigma_n^{-1} \sum_{i=1}^n [Z_{n,i} - \mathbb{E}(Z_{n,i})] \xrightarrow{d} \mathcal{N}(0, 1). \quad (\text{A.29})$$

Preuve. Voir Gnedenko 1962 (8) ■

A.3 Estimation dans les modèles de régression linéaire : Estimateur des moindres carrés ordinaires (MCO) et Estimateur des moindres carrés généralisés (MCG)

Les modèles de régression modélisent le comportement espéré que peut avoir une variable Y (variable expliquée) étant donné un vecteur d'observations X (variables explicatives et régresseurs). Le plus connu des modèles de régression est le modèle de régression linéaire donné par :

$$Y_i = \beta_0 + X_i^T \beta + u_i \quad i = 1, \dots, n,$$

où $X_i \in \mathbb{R}^q$ et β est un vecteur ligne de taille $q \times 1$ (vecteur des paramètres inconnus). Ici on suppose que la fonction de lien est une fonction linéaire de la forme $g(x) = x^T \beta$ ($x, \beta \in \mathbb{R}^q$), et donc une estimation de g se ramène à une estimation paramétrique du paramètre $\beta = (\beta_1, \dots, \beta_q)$.

Nous supposons que les données collectées suivent le modèle suivant :

$$y_i = \beta_1 x_{i1} + \dots + \beta_q x_{iq} + u_i \quad i = 1, \dots, n \quad (\text{A.30})$$

où

- les x_{ij} sont des nombres connus
- le vecteur des paramètres à estimer est $\beta = (\beta_1, \dots, \beta_q)^T$

— les u_i sont des variables aléatoires inconnues, erreurs.

Le modèle de régression linéaire est l'écriture matricielle de (A.30) donnée par

$$Y = X\beta + u \quad (\text{A.31})$$

où

— $Y = (y_1, \dots, y_n)^T$

— X est une matrice de taille $n \times q$ connue, appelée matrice d'expérience

— $u = (u_1, \dots, u_n)^T$ est le vecteur centré des erreurs.

Définition 8

On appelle estimateur des moindres carrés $\hat{\beta}$ de β la valeur suivante

$$\hat{\beta} = \underset{\beta_1, \dots, \beta_q}{\operatorname{argmin}} \sum_{i=1}^n \left(y_i - \sum_{j=1}^q \beta_j x_{ij} \right)^2 = \underset{\beta_1, \dots, \beta_q}{\operatorname{argmin}} (Y - X\beta)^T (Y - X\beta). \quad (\text{A.32})$$

Suivant la nature et les hypothèses faites sur le vecteur erreur u , on parlera d'estimateur des moindres carrés ordinaires $\hat{\beta}_O$ ou d'estimateur des moindres carrés généralisés $\hat{\beta}_G$. Les propriétés asymptotiques de ces estimateurs sont très connus, donc nous ne nous y attardons pas dessus et le Théorème suivant les résume :

Théorème 4**i) Estimateur des moindres carrés ordinaires**

On suppose que :

H1) La matrice X est de plein rang

H2) $\mathbb{E}(u) = 0$; $\mathbb{V}ar(u) = \Sigma_u = \sigma^2 I_n$, avec I_n la matrice identité d'ordre n

L'estimateur des moindres ordinaires est

$$\hat{\beta}_O = (X^T X)^{-1} X^T Y. \quad (\text{A.33})$$

Il est sans biais et de variance minimale parmi les estimateurs sans biais de β : on dit encore qu'il est optimal.

ii) Théorème d'Aïtken

On suppose que :

H1) La matrice X est de plein rang ie $\text{rang}(X) = q$

H2) $\mathbb{E}(u) = 0$; $\mathbb{V}ar(u) = \Sigma_u = \sigma^2 \Omega$, avec Ω une matrice symétrique définie positive telle que $\text{rang}(\Omega) = n$

L'estimateur des moindres carrés généralisés ou estimateur d'Aitken est

$$\hat{\beta}_G = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y. \quad (\text{A.34})$$

Il est optimal parmi les estimateurs sans biais de β .

Preuve. Voir [Rao\[\(19\), P.223\]](#) ■

A.4 Rappel d'algèbre

A.4.1 Projection

Définition 9

Soient E un espace vectoriel, et F, G deux sous espaces vectoriels de E tels que $E = F \oplus G$, ie, pour tout $x \in E$, il existe un unique couple $(x_F, x_G) \in F \times G$ tel que $x = x_F + x_G$. On appelle **projection de E sur F parallèlement à G** l'application $p : E \rightarrow E$ telle que, pour tout $x \in E$, $p(x) = x_F$.

Remarque A.4.1. L'application p ainsi définie est une application linéaire d'image $\text{Im}(p) = F$ et de noyau $\text{Ker}(p) = G$. De plus, $p \circ p = p$: on dit que p est idempotente.

Définition 10

Un endomorphisme $p \in \mathcal{L}(E)$ est appelé **projecteur** si $p \circ p = p$.

Définition 11

Soient E un espace euclidien, et F un sous espace vectoriel de E ; alors $E = F \oplus F^\perp$. On appelle **projection orthogonale de E sur F** et on note p_F , la projection de E sur F parallèlement à F .

Remarque A.4.2. Toute matrice carrée idempotente et symétrique est une matrice de projection orthogonale sur un sous espace de \mathbb{R}^n .

Exemple A.4.1. Si E est un espace euclidien de dimension finie, et X_1, \dots, X_p p vecteurs de E . On note par $F = \langle X_1, \dots, X_p \rangle$ le sous espace vectoriel engendré par ces p vecteurs et $X = [X_1, \dots, X_p]$ la matrice de taille $n \times p$ dont les colonnes sont formées de ces vecteurs et dont le rang est p . Alors la matrice de la projection orthogonale sur le sous espace vectoriel F est $P_F = X(X^T X)^{-1} X^T$ et celle sur l'orthogonale F^\perp de F est $P_{F^\perp} = I_n - P_F$. On vérifie aisément que ces matrices sont idempotentes et symétriques : donc elles sont des matrices de projections orthogonales.

A.4.2 Produit de Kronecker des matrices

Définition 12

Soient $A = (a_{ij})$ une matrice de taille $m \times n$ et $B = (b_{ij})$ une matrice de taille $p \times q$. Le produit de Kronecker de ces deux matrices est noté $A \otimes B$ et est défini par

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \quad (\text{A.35})$$

On écrit encore de façon compact $A \otimes B = [a_{ij}B]$ et c'est une matrice de taille $(mp) \times (nq)$.

Quelques propriétés du produit de Kronecker

Soient A_1, A_2 deux matrices de taille $m \times n$ et B_1, B_2 deux matrices de taille $p \times q$. On pose pour $i = 1, 2$: $D_i = A_i \otimes B_1$, $E_i = A_1 \otimes B_i$ et $C_i = A_i \otimes B_i$. Alors

$$D_1 + D_2 = (A_1 + A_2) \otimes B_1$$

$$E_1 + E_2 = A_1 \otimes (B_1 + B_2)$$

$$\alpha C_i = (\alpha A_i \otimes B_i) = (A_i \otimes \alpha B_i)$$

$$C'_i = A'_i \otimes B'_i$$

Si A et B sont deux matrices carrées de $m \times m$ et $n \times n$ respectivement et inversibles, alors $A \otimes B$ est inversible et

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

A.5 Codes R

```
#tracé des densité à noyau
plot(density(0,bw=0.4, kernel="gaussian"), xlim=c(-1.0,1.0),
     ylim=c(0.0,1.0), lwd=2, col="green", xlab="")
lines(density(0,bw=0.4, kernel="epanechnikov"), xlim=c(-1.0,1.0),
```

```

ylim=c(0.0,1.0), lwd=2)
lines(density(0,bw=0.4, kernel="triangular"), xlim=c(-1.0,1.0),
ylim=c(0.0,1.0), lwd=2, col="purple")
lines(density(0,bw=0.4, kernel="rectangular"), xlim=c(-1.0,1.0),
ylim=c(0.0,1.0), lwd=2, col="red")
lines(density(0,bw=0.4, kernel="cosine"), xlim=c(-1.0,1.0),
ylim=c(0.0,1.0), lwd=2, col="pink")
lines(density(0,bw=0.4, kernel="optcosine"), xlim=c(-1.0,1.0),
ylim=c(0.0,1.0), lwd=2, col="blue")
lines(density(0,bw=0.4, kernel="biweight"), xlim=c(-1.0,1.0),
ylim=c(0.0,1.0), lwd=2, col="darkgreen")
legend(0.7,1, legend = c("Gaussien", "Epanechnikov", "Triangular",
"Rectangular", "Cosine", "Triweight", "Biweight"),

col = c("green", "black", "purple", "red", "pink", "blue", "darkgreen"),
lty=1, cex = 0.5)

#test sur le jeu de données
data <- c(-0.57,0.25,-0.08,1.4,-1.05,-1,0.37,-1.15,0.73,1.59)
m_data <- mean(data)
sd_data <- sd(data)
x <- seq(-3,3,0.1)
plot(x, dnorm(x, m_data, sd_data), xlab = "", ylab = "", type = "l",
col="red")
lines(density(data, bw=1.5, kernel = "epanechnikov"), xlim=c(-3,3))
lines(density(data, bw=1, kernel = "epanechnikov"), xlim=c(-3,3),
col="green")
lines(density(data, bw=0.5, kernel = "epanechnikov"), xlim=c(-3,3),
col="blue")
legend(1.5,0.4, legend = c("Courbe f1", "kernel bw=1.5", "kernel bw=1",
"kernel bw=0.5"), col = c("red", "black", "green", "blue"), lty = 1,
cex = 0.6)
hist(data, xlab = "") #histogramme

```

Code R pour le tracé de la densité estimée des données Y

```
xx=seq(0,15, by=0.1)
plot(xx,dexp(xx,0.5),lwd=4,col="red",ylim=c(0,0.25),xlim=c(2,15),
type="l")
par(new=T)
Y=rexp(1000,0.5)
plot(density(Y,bw=0.5),xlim=c(2,15),ylim=c(0,0.25),lwd=2)
```

Code R pour la simulation de la courbe de $\hat{g}(z)$

```
n=600
g=function(x){
  set.seed(52)
  Z_it=runif(n,-1,1)
  sd(Z_it)
  h=sd(Z_it)*(n)^(-1/5)
  Nu_it=rnorm(n,0,1)
  U_it=runif(n,-1,1)
  Y_it=sin(2*Z_it)+U_it+Nu_it
  val=c()
  i=1
  while (i <= length(x)) {
    val[[i]]=sum(Y_it*dnorm((Z_it-x[i])/h))/sum(dnorm((Z_it-x[i])/h))
    i=i+1}
  return (val) }
x=seq(-1,1,1/149)
plot(x,sin(2*x),main="",type="l",col='green',xlab="",ylab="",lwd=2)
par(new = T)
plot(x,g(x),main="",type="l",col="blue",xlab="",ylab="", ylim=c(-1,1),
lwd=2)
legend(-0.6, 0.5, legend=c("sin2x", "g^(x)"), col=c("green", "blue"),
lty=1:2, cex=0.8)
```

Notations et Symboles

Les notations suivantes sont utilisées dans notre mémoire :

$(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé

\mathbb{E} : Espérance mathématique

Var : Variance mathématique

x^T : la transposée de $x \in \mathbb{R}^q$ avec $q \in \mathbb{N}^*$

$\|x\| = \sqrt{x_1 + \dots + x_q}$: est la norme euclidienne du vecteur $x = (x_1, \dots, x_q)^T \in \mathbb{R}^q$

Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction différentiable sur \mathbb{R} , $f^{(k)}$ désigne la dérivée k -ième de f

$f : \mathbb{R}^q \rightarrow \mathbb{R}$ est une fonction différentiable sur \mathbb{R}^q , $\partial_s f(x)$ désigne la dérivée partielle d'ordre 1 par rapport à x_s et $\partial_{ss'} f(x)$ désigne la dérivée partielle d'ordre 2 par rapport à x_s et $x_{s'}$ au point $x = (x_1, \dots, x_s, \dots, x_{s'}, \dots, x_q)' \in \mathbb{R}^q$

tr désigne l'opérateur trace

Si $r : \mathbb{R}^q \rightarrow \mathbb{R}$ est une fonction sur \mathbb{R}^q , nous utiliserons la notation $r(x + hv) \equiv r(x_1 + h_1 v_1, \dots, x_q + h_q v_q)$ pour x , h et v dans \mathbb{R}^q

\otimes : désigne le produit de Kronecker entre deux matrices

i.i.d : Indépendants et identiquement distribués

MSE : Mean squared error, erreur quadratique moyen