N° d'ordre :

Université de Saïda Dr- Moulay Tahar
Faculté des Sciences

Thèse

Présentée pour obtenir le diplôme de
Doctorat 3ème Cycle

Option :

Spécialité : Analyse stochastique et statistique des processus

Filière : Mathématiques

Par:

Souddi Youssouf

Thème

Sur la convergence faible des processus empiriques locaux

Thèse soutenue le 14 / 06 / 2022 devant le jury composé de :

Nom et prénom	Grade	Etablissement	Qualité.
KANDOUCI Abdeldjebbar	Prof.	Université Dr Moulay Tahar de Saïda.	Président.
MADANI Fethi	Prof.	Université Dr Moulay Tahar de Saïda.	Directeur de thèse.
BOUZEBDA Salim	Prof.	Alliance Sorbonne Univ., UTC, France.	Co-directeur de thèse.
RAHMANI Saädia	Prof.	Université Dr Moulay Tahar de Saïda.	Examinatrice.
BOUCHENTOUF A. Amina	Prof.	Université Djilali Liabess de Sidi Bel Abess.	Examinatrice.

Email:souddiyoucef@yahoo.fr

Acknowledgements

First of all, I thank **ALLAH**, for His helping and blessings to complete this thesis.

I want to express my great appreciation and deepest gratitude to my supervisors **Pr. Fethi Madani** and **Pr. Salim Bouzebda**, without whom this thesis would not have been emerged, I wold like to appraise their sound advices and carefuls guidance, their availability, their kindness and their constant support to carry out this work.
Again thank you.

I want to thank **Pr. KANDOUCI Abdeldjebbar** for chairing my committee members and for his help then, I would like to think my thesis committee **Pr. RAHMANI Saädia** , and **Pr. BOUCHENTOUF Angelika Amina** has agreed to participate in my jury and I thank them also for their encouragement, insightful comments, and meaningful questions.

I am very much thankful to all member of laboratory of stochastic models, statistics and applications, for their kindness and expertise, without forgetting my closest friends PhD students **Mr. Abdelmalik Keddi** and **Mr. Oussama Bouanani** and **Mr. Abdelhek Gendouzi**.

I am extremely grateful to my parents for their love, caring and sacrifices for my educate, their encouragement through my entire life.

Special thanks to my family and sister and don't forget her sons, my friends and everybody who was important to the successful realization of my thesis.

Dedication

I dedicate this work to
The Memory of Professor Mourid Tahar

List of Works

1. Publications:

   * "Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data" To appear in Int.J. Math.Stat.
     Salim Bouzebda, Fethi Madani and Youssouf Souddi.
   * "Some Characteristics of the Conditional Set-Indexed Empirical Process involving Functional Ergodic Data" Bull. Inst. Math. Acad. Sin. (N.S.), 16(4), 367–399.
     Youssouf Souddi, Fethi Madani and Salim Bouzebda.

2. Seminars and conferences presentations:

   * Paper presentation in the virtual national conference on new trends in applied mathematics (NCTAM 2021),title:"Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data" , India
   * Paper presentation in the virtual international conference of Mathematical Modelling in Biological science (October 26-28,2021)title:Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data, India.
   * Paper presentation in the virtual international conference of south Africa Mathematical science association (October22-24,2021)title:"Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data", South Africa
   * Presentation in the virtual international Student conference StudMath-IT (18-19 November 2021)title:"Conditional Empirical Process Involving Functional Data", Romaniei
   * Paper presentation in the virtual international conference of Fuzzy and computational Mathematics (October 28,2021)title:"Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process", India
   * Paper presentation in the international conference on Mathematics and Applications (ICMA 7-8 December 2021)title:"Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data", Blida, Algeria.
   * Poster presentation in the 1^{st} national conference on Applied science and Advanced Materials,(NCASAM December 20-22,2021), Higher Normal School of Technological Education,Title "Some Asymptotic Properties of the Conditional Set-Indexed Empirical Process Based on Dependent Functional Data", Skikda, Algeria.
   * Presentation in the national conference of Mathematics and applications (CNMA 2021) Center university of Mila,title: "About The Non Parametric Estimation of Conditional Empirical Process with Functional covarites" ,Algeria.

* Participation and Attendance in the Seminars organised by Laboratory of Stochastic Models, Statistics and Applications(LMSSA), Univerity Saida.

Summary

In this thesis, we focused the invariance principal for Nadaraya Watson conditional empirical process when the covarites are functional .

We propose set-indexed conditional empirical process where we establish the weak consistence and the asymptotic normality as well the density under some general conditions when the variables are stationary and strong mixing then we use our main results to the test of the conditional independence, than we extend our results to the ergodic data.

Résumé

Dans cette thèse, nous avons focalisé le principe d'invariance du processus empirique conditionnel lorsque les covariables sont fonctionnelles en introduisant l'estimateur du type Nadaraya Watson .

Nous avons proposé une processus empirique conditionnel indixé par une classe d'ensemble où nous établissons la consistance faible et la normalité asymptotique ainsi que la équicontinuité dans certaines conditions lorsque les variables sont stationnaires et fort mixtes puis nous utilisons nos principaux résultats pour tester l'indépendance conditionnel, et nous étendons nos résultats aux données ergodiques.

CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

1.1 Description and Contribution of Thesis

Nonparametric estimation constitutes a current research axis and important in statistics, this field of research is based on the study of nonparametric model for functional explanatory variable i.e. random variables with value in a space of infinite dimension, Consequently the literature including the study of functional data have grown considerably.

This thesis is to studyies the classical and important problem in nonparametric statistics for the theory of empirical process which is the invariance principal . In our work we are interested Nadaraya-Watson conditional empirical processes and we study some asymptotic properties of the constructed estimator when the covarites are functional.

As applications we investigate our results in this thesis by proposes a testing procedure for the justification of the theoretical results obtained. The test only needs a nonparametric estimator of the regression function depending on the explanatory variables which are significant under the null hypothesis, directly we use our empirical process as a models statistics for hypothesis testing.

The results presented in this thesis are far more general and they apply to a variety of interesting and novel situations in functional data analysis .

The thesis is organized as follows:

In the first chapter:
The next chapter is an introductory chapter, which presents a bibliographic study of problems related to statistical analysis of functional variables as well as non parametric estimation for conditional models whether in the finite dimension framework or infinite, we also give our model studied and a brief of the results.

In the second chapter:
We put some fundamental results of regression estimator in multivariate framework. We give the fundamental statistical properties of locally polynomial estimator of the regression function.
In the third chapter:
We put the definitions of basic elements of the thesis as the empirical process. We put the definitions of basic elements of the thesis as the functional variable, small ball probabilities, entropy ......etc.

In the fourth chapter:
We present our model Nadaraya Watson set-indexed conditional empirical process formed by strong mixing random variables and under some general conditions when the covariates are functional we establish the uniform weak consistence and the asymptotic normality as well density under some assumptions and richness of the index class and we apply our results for conditional independence test.
In the chapter five:
We extend the results of chapter 4 for the ergodic data.
In the last chapter:
we give the conclusion containing ours results with some comments and the prospects to focus on open future problems.

1.2 Bibliographic context

The theory of empirical processes plays a fundamental role in statistics with many applications can appear in too much theoretical and practical problems, one of the first applications of empirical process theory is to understand goodness of fit test statistics such as Kolmogorov-Smirnov statistic, Cramér-von Mises statistic used in the work of Darling (1957), in the areas of estimating some models statistics and to derive consistency and rate of convergence for them , a number of publications and books published during the past decades see for examples books Shorack and Wellner (1986), Pollard (1990), Van der Vaart and Wellner (1996), Dudley (1999), Van der Vaart and Wellner (2000), Van de Geer (2006), we mention some other applications such as empirical quantile refer Csörgo and Révész (1978),Deheuvels and Mason (1992) and to statistics of censored data one of the most popular statistics of censored is Kaplan-Meier empirical process,the M-estimators approach seeVan de Geer (2006), the Bootstrap methods Aenssler (1985), Radulović (1996), the Two-sample problem, copula processes Doukhan *et al.* (2009), Marcin (2017), U-statistic Dehling *et al.* (1987), for more applications of empirical process see the books mentioned previously.
In classical non parametric statistics the simplest sort of empirical process arises when trying to estimate a probability distribution, Glivenko (1933) and Cantelli (1933) showed the maximum difference between the empirical and true distribution functions converges

to zero when samples are independents and identically distributed, In literature a number of examples have been study asymptotic proprieties of empirical process considering the independents observations can be found in Gänssler and Stute (1979), pyke (1968), Révész (1976), we cite among many others Dudley (1978), Giné and Zinn (1984), Le Cam (1983), Pollard (1982), Bass and Pyke (1984), for conditional process Stute (1986a)proved almost sure and weak convergence results for kernel and nearest neighbor estimates of the conditional empirical function and Horváth and Yandell (1988) studied asymptotic of the kernel and the nearest neighbor type estimator of Nadaraya Watson empirical process, Interestingly, the statistical applications of empirical process was soon increase so empirical process also was soon extend to various types of mixing for example Yoshihara (1988) introduced a nearest-neighbour-type estimate of regression function and proved that the distribution of the estimate is asymptotically normal under some conditions when the sequence is ϕ-mixing, invariance principles studied by Doukhan (1995) considered β-framework we cite among Withers (1975) , Phillip (1984), Harel and Puri (1987), Massart (1988) however, In the last few years many authors interested to studies the asymptotic proprieties of empirical process considering multivariate conditional mixing framework Polonik and Yao (2002) studied invariance principal for set-indexed conditional empirical process where it was extended by Poryvaĭ (2005) for conditional empirical process indexed by functions but, situation where both X and Y are real or multivariate it has received significantly less attention currently.

Non-parametric estimation for statistics models based on kernel and their asymptotic properties according as well applications frequently used in the theory of empirical process, recall the density estimation has been the subject of a great deal of work, its field of application is very broad and covers various fields, such as the analysis of regression, chronological series and the theory of reliability, the most important nonparametric methods for density estimation are: kernel method introduced by Rosenblatt (1956)and Parzen (1962), the series method orthogonal studied among others by Schwartz (1967)and Watson (1964) and the method of histogram introduced by Graunt then developed by Scott and Tran (1994) and, Carbon*et al.* (1996).

Studying the links between two random variables is a very important question. aunt in statistics. Historically, this problem has been addressed for the first time in a geometric context by Galileo Galilei (1632), The main idea is to adjust a point cloud by a right to interpret the relationship between contaminated data.A mathematical formulation for this problem, known in the literature under the name of linear regression, was given by Legendre and Gauss independing, in (1805) and (1809) and is based on the principle of least squares. In statistics, this problem can be modeled as follows boasts: suppose we have two dependent random variables X and Y, the forecast of Y knowing X is done through X by an application r. In other words, we are looking for a function r such that r(X) is a good approximation of Y from a given criterion.In our non-parametric context, the first results were obtained by Tukey (1961) . While the kernel method estimate has been first used in 1964 separately by Nadaraya (1964)and Watson (1964). This estimation method has undergone continuous development. Indeed, Devroye (1978) established the almost sure uniform convergence of this estimator. The optimal convergence rate for nonparametric regression a was given by Stone (1982) , Collomb (1981),Collomb (1985) Collomb (1983) Collomb (1984) brings a decisive contribution to this model. These works are focused on using regression in forecasting time series. The first asymptotic results on the nonparametric estimate of the regression function on the α-mixing processes have been developed by Györfi *et al.* (2007). In this α-mixing framework, Vieu (1991) gave the asymptotically exact terms of the quadratic error of the estimator at core of the regression function. We refer to Bosq and Lecoutre (1987), Schimek (2000), Sarda and Vieu (2000) for a wide range of references.

The study of nonparametric models related to the conditional distribution has been widely considered in nonparametric statistics. Historically, the first results on these models were obtained by Roussas (1969) he treated the estimation of the conditional distribution function by the kernel method using Markov observations.He established the convergence in the probability of the constructed estimator.An alternative estimator for the same model was developed by Stone (1977)the later study the empirical estimator of conditional distribution function and applied the results obtained to the estimator of the conditional quantiles as the generalised inverse of the conditional distribution function Stute (1986b) added results on the almost complete convergence of the conditional distribution function of a vector random variables conditionally to vector explanatory variable added results on the almost complete convergence of the kernel estimator of the distribution function of a vector random variable conditionally to a vector explicative variable. The estimation of the conditional mode was treated for the first time by Collomb*et al* (1987). These authors showed the uniform convergence of the kernel estimator of this conditional model when the observations are ϕ mixing, Samanta (1989), studied the asymptotic normality of the kernel estimator of conditional quantiles when the observations are independent and identically distributed.The latter, in collaboration with Samanta (1990), obtained the same asymptotic property for a kernel estimator of the conditional mode by considering the case i.i.d. Roussas (1991) established the almost sure convergence of a kernel estimator.Of the conditional quantiles when the observations come from a Markov process. The contribution of Youndjé (1993) on the estimation of the conditional density is decisive. He addressed the question of the choice of the smoothing parameter by considering the two independent and dependent cases. Quintela and Vieu (1997) have treated the conditional mode as being the point that cancels the first order derivative of the conditional density and constructed an estimator for this model using the kernel estimator of the derivative of the conditional density.Ould-said (1997) studied the kernel estimator of the conditional mode from ergodic observations. We refer to Berlinet *et al.* (1998a),Louani and Ould-Said (1999) for the convergence in law of the kernel estimator of the conditional mode in the α-mixing case. The Berlinet *et al.* (1998b) gives a general theorem of the asymptotic normality of the conditional quantile estimators, independently of the correlation of the observations.Zhou and Liang (2000) used the L^1 approach to construct a conditional median estimator using α-mixing observations. They showed the asymptotic normality of this estimator. The convergence in Lp norm of the kernel estimator of the conditional density of a stationary Markov process was obtained by Laksaci and Yousfate (2002), Ioannides and Matzner-Lober (2002) constructed an estimator for the conditional mode, when, the observations are tainted by errors. In this article the authors focus on the almost sure convergence of the proposed estimator. While its asymptotic normality has been demonstrated by the same authors in Ioannides and Matzner-Lober (2004).

On the other hand, The modelization of functional variables that taking values in infinite dimensional spaces had received a lot of attention in the last few years, there are an increasing number of situation coming from different fields of applied sciences(environment, chemometrics, biometrics, medicine, econometrics,....) in which the collected data are curves,the study of statistical models adapted to such type of infinite dimensional data has been the subject of several works in the recent statistical literature good overviews about this literature can be found in Ramsay and Silverman (2005a), Bosq (2000), Ramsay and Silverman (2005b), Ferraty and Vieu (2006), Bosq and Blanke (2007), Shi and Choi (2011), Horváth and Kokoszka (2012), Zhang (2014), Bongiorno *et al.* (2014), Hsing and Eubank (2015) and Aneiros *et al.* (2017) and hundreds of papers and books have been published in this framework last decade.

The study of the problem of testing conditional independence has a long history. However, there are relatively few results on nonparametric tests when the vectors X, Y and Z are continuous. some examples of such tests can be found Su and White (2007), Su and White (2008) A primary role of hypothesis testing in empirical work is to justify model simplification. Whether one is testing a restriction implied by economic theory or an interesting behavioral property, the test asks whether imposing the restriction involves a significant departure from the data evidence,When the methods of analysis are widened to include nonparametric techniques, the need for model simplification is arguably even more important than with parametric modeling methods there is a large literature on testing independence with various method applied. for example authors have proposed tests based on a comparison between the empirical integrated regression and the estimated parametric integrated regression function under the specification in the null; see, for example, Brunk (1970), Hong-zhy and Bing (1991), Sue and Wei (1991) and Stute (1997). These tests are based on a marked empirical process and, in general, their null asymptotic distribution depends on certain features of the data generating process. The limiting distribution can be tabulated when the distribution of the regressors is known. Also Stute*et al.* (1998) and Koul and Stute (1999) suggest a transformation of the underlying empirical process, when the regression depends only on one variable, which is asymptotically distribution free under the null. Transformations when the regression model depends on more than one variable are still unexplored.
And for our needs a complete treatment of the necessary weak convergence theory, it is worthwhile to see the monograph of Van der Vaart and Wellner (1996) presents an

important collection of statistical tools for empirical process and also by Shorack and Wellner (1986).

1.3 Brief Presentation of Results

1.3.1 Notations:

We consider a sample of random elements $(X_1, Y_1), \ldots, (X_n, Y_n)$ copies of (X, Y) that takes its value in a space $\mathcal{E} \times \mathbb{R}^d$. The functional space $\mathcal{E}$ is equipped with a semi-metric $d_{\mathcal{E}}(\cdot, \cdot)$. The links between X and Y, estimating by functional operators associated to the conditional distribution of Y given X such as the regression operator, for some measurable set C in a class of sets $\mathscr{C}$,

$$\mathbb{G}(C \mid x) = \mathbb{E}\left(\mathbb{1}_{\{Y \in C\}} \mid X = x\right).$$

This regression relationship suggests to consider the following Nadaraya Watson-type conditional empirical distribution:

$$\mathbb{G}_n(C, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)},$$

where $K(\cdot)$ is a real-valued kernel function from $[0, \infty)$ into $[0, \infty)$ and h_n is a smoothing parameter C is a measurable set, and $x \in \mathcal{E}$.Concerning the semi-metric topology defined on $\mathcal{E}$ we will use the notation

$$B(x, t) = \{x_1 \in \mathcal{E} : d_{\mathcal{E}}(x_1, x) \leq t\},$$

for the ball in $\mathcal{E}$ with center x and radius t. We denote

$$\mathcal{F}_x(t) = \mathbb{P}(d_{\mathcal{E}}(x, X) \leq t) = \mathbb{P}(X \in B(x, t)),$$

which is the small ball probability function which we will note $\phi(h_n)$. We study the asymptotic behaviour of the conditional empirical process:

$$\widetilde{\nu}_n(C, x) = \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C, x) - \mathbb{E}(\mathbb{G}_n(C, x))\right), \text{ for } C \in \mathscr{C}.$$

1.3.2 Results:

Strong mixing data

Theorem 1 (Uniform Consistency). *For (X_t, Y_t) is geometrically strong mixing with $\beta > 2$. Let $\mathscr{C}$ be a class of measurable sets for which $\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)) < \infty$ for any $\varepsilon > 0$. Suppose* *further that* $\forall C \in \mathscr{C}$

$$|\mathbb{G}(C,y)f(y) - \mathbb{G}(C,x)f(x)| \longrightarrow 0, \quad as \;\; y \to x.$$

If $n\phi(h_n) \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$*, then*

$$\sup_{C\in\mathscr{C}} |\mathbb{G}_n(C,x) - \mathbb{E}\left(\mathbb{G}_n(C,x)\right)| \xrightarrow{\mathbb{P}} 0.$$

Theorem 2 (Asymptotic normality)**.** *Let **(H2)-(H5)(i)(ii)-(H6)-(H8)-(H9)(i)** hold and* (X_i, Y_j) *is geometrically strong mixing with* $\beta > 2$*, then* $n\phi(h_n) \to \infty$ *as* $n \to \infty$*. For* $m \geq 1$ *and* $C_1, \ldots, C_m \in \mathscr{C}$*,*

$$\{\widetilde{\nu}_n(C_i, x)_{i=1,\ldots,m}\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = \sigma_{ij}(x), i, j = 1, \ldots, m$ *and*

$$\sigma_{ij}(x) = \frac{C_2}{C_1^2 f_1(x)} \left(\mathbb{E}(\mathbb{1}_{\{Y\in C_i\cap C_j\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y\in C_i\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y\in C_j\}} \mid X = x)\right),$$

whenever $f_1(x) > 0$ *and*

$$C_1 = K(1/2) - \int_0^{1/2} K^{'}(s)\tau_0(s)ds, \;\; C_2 = K^2(1/2) - \int_0^{1/2} (K^2)^{'}(s)\tau_0(s)ds.$$

Theorem 3. *The process* (X_i, Y_i) *are exponentially strong mixing for each* $\sigma^2 > 0$*, let* $\mathscr{C}_\sigma \subset \mathscr{C}$ *be a class of measurable sets with*

$$\sup_{C\in\mathscr{C}_\sigma} \mathbb{G}(C,x) \leq \sigma^2 \leq 1,$$

and suppose that $\mathscr{C}$ *fulfils* (R_γ) *with* $\gamma \geq 0$*. Further, we assume that* $\phi(h_n) \to 0$ *and* $n\phi(h_n) \to +\infty$ *as* $n \to +\infty$*, such that*

$$n\phi(h_n) \leq \left(\Lambda_\gamma(\sigma^2, n)\right)^2,$$

and as $n \to +\infty$*, we have*

$$\frac{n\phi\left(\sigma^2 \log\left(\frac{1}{\sigma^2}\right)\right)^{1+\gamma}}{\log(n)} \to \infty.$$

Further we assume that $\sigma^2 \geq h^2$*. For* $\gamma > 0$ *and* $d = 1, 2$*, the later has to be replaced by* $\sigma^2 \geq \phi(h_n)\log\left(\dfrac{1}{\phi(h_n)}\right)$ *then for every* $\epsilon > 0$ *there exist a constant* $M > 0$ *such that*

$$\mathbb{P}\left(\sup_{C\in\mathscr{C}_\sigma} |\widetilde{\nu}_n(C \mid x)| \geq M\Lambda_\gamma(\sigma^2, n)\right) \leq \epsilon,$$

for all sufficiently large n*. Then the following function which provides the information on* *the asymptotic behaviour of the modulus of continuity*

$$\Lambda_\gamma(\sigma^2, n) = \begin{cases} \sqrt{\sigma^2 \log \frac{1}{\sigma^2}}, & \textit{if} \quad \gamma = 0; \\ \max\left((\sigma^2)^{(1-\gamma)/2}, n\phi(h_n)^{(3\gamma-1)/(2(3\gamma+1))}\right), & \textit{if} \quad \gamma > 0. \end{cases}$$

Theorem 4. *The process:*

$$\{\widetilde{\nu}_n(C \mid x) : C \in \mathscr{C}\},$$

converges in law to a Gaussian process $\{\widetilde{\nu}(C \mid x) : C \in \mathscr{C}\}$, *that admits a version with uniformly bounded and uniformly continuous paths with respect to* $\|\cdot\|_2$*−norm.*

1.3.3 Testing the independence:

We consider a sample of random elements $(X_1, Y_{1,1}, Y_{1,2}), \ldots, (X_n, Y_{n,1}, Y_{n,2})$ copies of (X, Y_1, Y_2) that takes its value in a space $\mathcal{E} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and define, for $(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\mathbb{G}_n(C_1 \times C_2, x) = \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_{i,1} \in C_1\}} \mathbb{1}_{\{Y_{i,2} \in C_2\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{1.1}$$

$$\mathbb{G}_{n,1}(C_1, x) = \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_{i,1} \in C_1\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{1.2}$$

$$\mathbb{G}_{n,2}(C_2, x) = \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_{i,2} \in C_2\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}. \tag{1.3}$$

We will investigate the following processes, for $(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\widehat{\nu}_n(C_1, C_2, x) = \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\mathbb{E}(\mathbb{G}_n(C_2, x))\right), \tag{1.4}$$

$$\breve{\nu}_n(C_1, C_2, x) = \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{G}_{n,1}(C_1, x)\mathbb{G}_{n,2}(C_2, x)\right). \tag{1.5}$$

Notice that we have

$$\begin{aligned} \breve{\nu}_n(C_1, C_2, x) \quad = \quad & \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\mathbb{E}(\mathbb{G}_n(C_2, x))\right) \\ & + \sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_2, x))\left(\mathbb{G}_n(C_1, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\right) \\ & - \sqrt{n\phi(h_n)}(\mathbb{G}_n(C_1, x))\left(\mathbb{G}_n(C_2, x) - \mathbb{E}(\mathbb{G}_n(C_2, x))\right). \end{aligned}$$

Hence we have

$$\begin{aligned}\breve{\nu}_n(C_1, C_2, x) &\overset{d}{=} \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\mathbb{E}(\mathbb{G}_n(C_2, x))\right) \\ &\quad + \sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_2, x))\left(\mathbb{G}_n(C_1, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\right) \\ &\quad - \sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_1, x))\left(\mathbb{G}_n(C_2, x) - \mathbb{E}(\mathbb{G}_n(C_2, x))\right) \\ &= \widehat{\nu}_n(C_1, C_2, x) + \mathbb{E}(\mathbb{G}_n(C_2, x))\widetilde{\nu}_n(C_1, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\widetilde{\nu}_n(C_2, x). \quad (1.6)\end{aligned}$$

One can show that, for $(A_1, B_1), (A_2, B_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\begin{aligned}&\operatorname{cov}(\widehat{\nu}_n(A_1, B_1, x), \widehat{\nu}_n(A_2, B_2, x)) \\ &= \frac{C_2}{C_1^2 f_1(x)}\left(\mathbb{E}(\mathbb{1}_{\{Y\in A_1\cap A_2\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y\in A_1\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y\in A_2\}} \mid X = x)\right) \\ &\quad \times \left(\mathbb{E}(\mathbb{1}_{\{Y\in B_1\cap B_2\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y\in B_1\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y\in B_2\}} \mid X = x)\right), \quad (1.7)\end{aligned}$$

whenever $f_1(x) > 0$. Let $\{\widehat{\nu}(C_1, C_2, x) : (C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2\}$ be a Gaussian process with covariance given in (4.13). Let us introduce the following limiting process, for $(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\breve{\nu}(C_1, C_2, x) = \widehat{\nu}(C_1, C_2, x) + \mathbb{G}(C_2, x)\widetilde{\nu}(C_1, x) - \mathbb{G}(C_1, x)\widetilde{\nu}(C_2, x).$$

We would test the following null hypothesis

$$\mathcal{H}_0 : Y_1 \text{ and } Y_2 \text{ are conditionally independent given } X = x.$$

Against the alternative

$$\mathcal{H}_1 : Y_1 \text{ and } Y_2 \text{ are conditionally dependent.}$$

Statistics of independence those can be used are

$$S_{1,n} = \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2} |\widehat{\nu}_n(C_1, C_2, x)|, \quad (1.8)$$

$$S_{2,n} = \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2} |\breve{\nu}_n(C_1, C_2, x)|. \quad (1.9)$$

A combination of Theorem 17 with continuous mapping theorem we obtain the following result.

Theorem 5. *We have under condition of Theorem 17, as $n \to \infty$,*

$$S_{1,n} \to \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2} |\widehat{\nu}(C_1, C_2, x)|, \quad (1.10)$$

$$S_{2,n} \to \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2} |\breve{\nu}(C_1, C_2, x)|. \quad (1.11)$$

Ergodic data

Theorem 6. *[Uniform Consistency] Suppose that the hypotheses **(H1)-(H7)** hold. Let* $\mathscr{C}$ *be a class of measurable sets for which*

$$\mathcal{N}\left(\varepsilon,\mathscr{C},\mathbb{G}\left(\cdot\mid x\right)\right)<\infty,$$

for any $\varepsilon>0$*. Suppose further that* $\forall C\in\mathscr{C}$

$$|\mathbb{G}(C,y)f(y)-\mathbb{G}(C,x)f(x)|\longrightarrow 0,\quad \textit{as}\ \ y\to x.$$

If $n\phi(h_n)\to\infty$ *and* $h_n\to 0$ *as* $n\to\infty$*, then*

$$\sup_{C\in\mathscr{C}}|\mathbb{G}_n(C,x)-\mathbb{E}\left(\mathbb{G}_n(C,x)\right)|\overset{\mathbb{P}}{\longrightarrow}0.$$

Theorem 7 (Asymptotic normality)**.** *Let **(H1)-(H7)** hold. Then as* $n\to\infty$*, for* $m\geq 1$ *and* $C_1,\ldots,C_m\in\mathscr{C}$*,*

$$\{\widetilde{\nu}_n(C_i,x)_{i=1,\ldots,m}\}\overset{\mathcal{D}}{\longrightarrow}\mathcal{N}(0,\Sigma),$$

where $\Sigma=\sigma_{ij}(x),i,j=1,\ldots,m$ *and*

$$\sigma_{ij}(x)=\frac{\mathfrak{C}_2}{\mathfrak{C}_1^2 f_1(x)}W_2(x),$$

whenever $f_1(x)>0$ *and*

$$\mathfrak{C}_1=k(1)-\int_0^1 K'(u)\tau_0(u)d(u),\ \ \mathfrak{C}_2=K^2(1)-\int_0^1(K^2)'(u)\tau_0(u)du.$$

Theorem 8. *Suppose that **(H1)-(H7)** hold. For each* $\sigma^2>0$*, let* $\mathscr{C}_\sigma\subset\mathscr{C}$ *be a class of measurable sets with*

$$\sup_{C\in\mathscr{C}_\sigma}\mathbb{G}(C,x)\leq\sigma^2\leq 1,$$

and suppose that $\mathscr{C}$ *fulfils* (5.3) *with* $\gamma\geq 0$*. Further, we assume that* $\phi(h_n)\to 0$ *and* $n\phi(h_n)\to+\infty$ *as* $n\to+\infty$*, such that*

$$n\phi(h_n)\leq\Big(\Lambda_\gamma(\sigma^2,n)\Big)^2,$$

and as $n\to+\infty$*, we have*

$$\frac{n\phi\left(\sigma^2\log\left(\frac{1}{\sigma^2}\right)\right)^{1+\gamma}}{\log(n)}\to\infty.$$

Further we assume that $\sigma^2\geq h^2$*. For* $\gamma>0$ *and* $d=1,2$*, the later has to be replaced by* $\sigma^2\geq\phi(h_n)\log\left(\frac{1}{\phi(h_n)}\right)$*, then under conditions of Theorem 20 we have the process:*

$$\{\widetilde{\nu}_n(C,x):C\in\mathscr{C}\},$$

converges in law to a Gaussian process $\{\widetilde{\nu}(C,x) : C \in \mathscr{C}\}$ *, that admits a version with uniformly bounded and uniformly continuous paths with respect to* $\|\cdot\|_2-$*norm with covariance* $\sigma_{ij}(x)$ *given in Theorem 20.*

The hypotheses and the proofs and details of the conditions imposed of the results will be given in chapter 4 and 5

CHAPTER 2

ON THE NONPARAMETRIC ESTIMATION OF REGRESSION FUNCTION

2.1 Introduction

Estimation theory is one of the most basic branches of statistics. This theory is usually divided into two main components, namely, parametric estimation and non-parametric estimation. The problem of nonparametric estimation consists, in most of the cases, in estimating, from observations, an unknown function, element of a certain functional class. Remember that a non-parametric procedure is defined regardless of the distribution or law of the sample of observations. More specifically, we talk about estimation method non-parametric when this is not reduced to the estimation of a finite number of real parameters associated with the law of the sample. Central problems in statistics is that of the estimation of functional characteristics associated with the law of observations, for example, the density function or the regression function (in a multivariate model). One of the most frequently encountered models in parametric or non-parametric statistics is the regression model, a description of which is given below. We have a sample, composed of n independent pairs of random variables $(X_1, Y_1), ..., (X_n, Y_n)$ copier of (X, Y)a generic element of this sample in the non-parametric regression model, we typically assume the existence of a function $r(.)$ which expresses the mean value of the response variable Y as a function of the input variable X:

$$Y_i = r(X_i) + \epsilon_i \quad for \quad 1 \leq i \leq n \quad and \quad \epsilon_i \stackrel{d}{=} \epsilon \sim \mathcal{N}(\mu, \sigma^2). \tag{2.1}$$

The error made is, in the classic case, modeled by a Gaussian random variable, which will generally be chosen independent of the observations$\{X_i : 1 \leq i \leq n\}$ and of mean μ zero. This last hypothesis considerably simplifies the calculations and the expression of asymptotic properties linked to the estimation of the regression function, under such a simplified model, will not be considered in our work. We consider the more delicate problem posed by the estimation of the regression function, without assumption particular on the law of the couple (X, Y) other than that of the existence of r(·) (supposed sufficiently regular), and superior moments of suitable order of X and Y There are two main cases for the model (2.1) depending on the probabilistic nature of the data $\{(X_i, Y_i) : 1 \leq i \leq n\}$. The first case is the simplest, and is called a device experimental fixed effects (or "fixed design") It corresponds to the situation where the $X_i = x_i$ are fixed (i.e. constants p.s., or, equivalently, deterministic or degenerate).

Example 1. *The regular experimental system we assume $X_i = x_i = i/n$ and $r(.)$ a function of $[0, 1]$ in $\mathbb{R}$ such as*

$$Y_i = r(i/n) + \epsilon_i, \quad for \quad 1 \leq i \leq n.$$

The second case, called an experimental device with random effects (or "random design") denotes the model where the data $\{X_i : 1 \leq i \leq n\}$ are strictly random (or non degenerate). We will essentially study this latter model, which is clearly more general. Note also that only models with independent observations will be analyzed, the study of the dependency case adding only technical difficulties.

We will now present the regression function more explicitly, in the framework of the univariate random model. Let (X, Y) be a couple of real random variables admitting an attached density on $\mathbb{R}^2$ denoted $f_{X,Y}$and a marginal density f_x. the variable Y is assumed to be integrable i.e. $\mathbb{E}(|Y|) < \infty$ We can then properly define the regression function or conditional expectation of Y knowing $X = x$, by

$$r(x) = \mathbb{E}(Y \mid X = x) = \frac{\displaystyle\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{\displaystyle\int_{\mathbb{R}} f_{X,Y}(x, y) dy} = \frac{m(x)}{f_X(x)} \tag{2.2}$$

when the density $f_X(x)$ is different than zero the problem of estimating r(·) is of the non-parametric type, i.e. the regression function belongs to a nonparametric (infinidimensional) set. For example, we can assume that r(·) belongs to the function class $\mathcal{F}$ consisting of continuous functions on [0,1] (cf. example 1 above), when the density support is the interval [0,1]. For the study of the minimax properties of the regression function estimators, the nonparametric classes of functions encountered are of the Hölder, Sobolev or Besov type The regression function r(x) defined above in(2.2) realizes (for all x fixed) the best approximation of Y knowing $X = x$, in the least squares sense, assuming Y of an integrable square. In this first chapter, we will discuss some methods of constructing estimators of regression by the kernel method. Then, we will focus our work on the statistical properties of estimators (convergence, speed of convergence) as well as their optimality The estimators we consider belong to the large class of estimators linear (i.e. linear as a function of observations Y_i):

Definition 2.1.1. *the estimator $\hat{r}_n(x)$ of $r_n(x)$ is said to be a linear estimator of the nonparametric regression if*

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

where the weight function $W_{ni}(.)$ does not depend on the observations Y_i. The class of linear estimators groups the majority of the regression estimators, i.e. estimators by spline functions, by projection or orthogonal series, by wavelets, and by the kernel method. In the next section, we will present the famous kernel estimator of regression introduced by Nadaraya (1964) and Watson (1964) and some of its essential properties. We will then be interested in the asymptotic optimality of this estimator, then in the locally polynomial estimation of the regression, which is one of the most effective approaches today. For a bibliographic review of older work on non-parametric regression, we cite the articles by Collomb (1981) and Stone (1977).

Definition 2.1.2. *A kernel function $K(u) : \mathbb{R} \to \mathbb{R}$ is any function which satisfies:*

$$\int_{-\infty}^{+\infty} k(u)du = 1$$

Some examples

(a) Box kernel:$k(u) = \frac{1}{2}\mathbb{1}_{[-1,+1]}(u)$;

(b) Triangle kernel :$k(u) = (u+1)\mathbb{1}_{[-1,0]}(u) + (1-u)\mathbb{1}_{[0,+1]}(u)$;

(c) Quadratic kernel:$k(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{[-1,+1]}(u)$;

(d) Gaussian kernel:$k(u)\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-u^2}{2}\right)$;

2.1.1 Dependency measure

Mixture

The notion of strong mixing has been introduced by Rosenblatt (1956) as a dependency structure and was used in that paper in the proof of a central limit theorem. This mixing condition has two advantages. First, it is the least restrictive among the various mixing conditions existing in classical literature (see Doukhan (1994),ch11). Then the likely knowledge about this type of dependance is sufficient pushes (see Rio (2000)) to allow us to carry out the study of non parametric forecast which is our main concern.

Definition 1. *For the sequence $\{(X_t, Y_t)\}$ we define the α mixing coefficient by*

$$\alpha(j) = \sup \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{j}^{-\infty}} \{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|\},$$

where $\mathcal{F}_s^t$ denotes σ-algebra generated by $\{(X_t, Y_t), s \le i \le t\}$.

The sequence is called strong mixing if the coefficient α verify $\lim_{n\to\infty} \alpha(n) = 0$

Definition 2.1.3. *The sequence $\{(X_t, Y_t)\}$ is called geometrically strong mixing if: $\alpha(j) \leq aj^{-\beta}$ for some $a > 0$ and $\beta > 1$, and exponentially strong mixing if: $\alpha(k) \leq b\gamma^k$ for some $b > 0$ and $0 < \gamma < 1$.*

Comments on limit theory under α-mixing: Under α-mixing and other similar conditions (including ones reviewed below), there has been a vast development of limit theory — for example, CLTs, weak invariance principles, laws of the iterated logarithm, almost sure invariance principles, and rates of convergence in the strong law of large numbers. For example, the CLT in Rosenblatt (1956) evolved through subsequent refinements by several researchers into the following "canonical" form. (For its history and a generously detailed presentation of its proof, see e.g. [Bradley (2007), v1, Theorems 1.19 and 10.2].
Several other classic strong mixing conditions:
As indicated above, the terms "α-mixing" and "strong mixing condition" (singular) both refer to the condition $\alpha(n) \to 0$. (A little caution is in order; in ergodic theory, the term "strong mixing" is often used to refer to the condition of "mixing in the ergodic-theoretic sense", which is weaker than α-mixing as noted earlier.) The term "strong mixing conditions" (plural) can reasonably be thought of as referring to all conditions that are at least as strong as (i.e. that imply) α-mixing. In the classical theory, five strong mixing conditions (again, plural) have emerged as the most prominent ones: α-mixing itself and four others that will be defined here. Recall our probability space $(\Omega, \mathcal{F}, \mathcal{P})$. For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, define the following four "measures of dependence":

$$\phi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{A}, \mathbb{P}(A) > 0} |\mathbb{P}(B \mid A) - \mathbb{P}(B)|; \tag{2.3}$$

$$\psi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{A}, \mathbb{P}(A) > 0, \mathbb{P}(B) > 0} |\mathbb{P}(B \cap A)/(\mathbb{A}\mathbb{A}) - 1|; \tag{2.4}$$

$$\rho(\mathcal{A}, \mathcal{B}) = \sup_{f \in \mathcal{L}^2(\mathbb{A}), g \in \mathcal{L}^2(\mathbb{B})} |Corr(f, g)|; \tag{2.5}$$

$$\beta(\mathcal{A}, \mathcal{B}) = \sup(1/2) \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \tag{2.6}$$

where the latter supremum is taken over all pairs of finite partitions $(A_1, A_2, \ldots, A_I)$ and $(B_1, B_2, \ldots, B_J)$ of Ω such that $A_i \in \mathcal{A}$ for each i and $B_j \in \mathcal{B}$ for each j. In (2.5), for a given σ-field $\mathcal{D} \subset \mathcal{F}$, the notation $\mathcal{L}^2(\mathcal{D})$ refers to the space of (equivalence classes of) square-integrable,$\mathcal{D}$-measurable random variables.
Now suppose $X := (X_k, k \in Z)$ is a strictly stationary sequence of random variables on $(\Omega, \mathcal{F}, \mathcal{P})$. For any positive integer n, define the dependence coefficient

$$\phi(n) = \phi(X, n) := \phi(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{0}^{\infty}) \tag{2.7}$$

And define analogously the dependence coefficients $\psi(n)$, $\rho(n)$, and $\beta(n)$. Each of these four sequences of dependence coefficients is trivially nonincreasing. The (strictly stationary) sequence X is said to be:

"ϕ-mixing" if $\phi(n) \to 0$ as $n \to \infty$;

"ψ-mixing" if $\psi(n) \to 0$ as $n \to \infty$;

"ρ-mixing" if $\rho(n) \to 0$ as $n \to \infty$; and

"absolutely regular", or "β-mixing", if $\beta(n) \to 0$ as $n \to \infty$. one has the following well known inequalities:

$$2\alpha(\mathcal{A}, \mathcal{B}) \leq \phi(\mathcal{A}, \mathcal{B}) \leq (1/2)\psi(\mathcal{A}, \mathcal{B});$$

$$4\alpha(\mathcal{A}, \mathcal{B}) \leq \rho(\alpha(\mathcal{A}, \mathcal{B})) \leq \psi(\mathcal{A}, \mathcal{B})$$

and

$$\rho(\alpha(\mathcal{A}, \mathcal{B})) \leq 2\left(\phi(\mathcal{A}, \mathcal{B})\right)^{1/2}\left(2\phi(\mathcal{B}, \mathcal{A})\right)^{1/2} \leq 2\left(\phi(\mathcal{A}, \mathcal{B})\right)^{1/2}$$

For a history and proof of these inequalities, see e.g. [Bradley (2007), v1, Theorem 3.11]. As a consequence of these inequalities and some well known examples, one has the following "hierarchy" of the five strong mixing conditions here:

(i) ψ-mixing implies ϕ-mixing.

(ii) ϕ-mixing implies both ρ-mixing and β-mixing (absolute regularity).

(iii) ρ-mixing and β-mixing each imply α-mixing(strong mixing)

(iv) Aside from “transitivity”, there are in general no other implications between these five mixing conditions. In particular, neither of the conditions ρ-mixing and β-mixing implies the other.

For all of these mixing conditions, the "mixing rates" can be essentially arbitrary, and in particular, arbitrarily slow.

Remark 1. *the α mixing coefficient such that $0 \leq \alpha \leq \frac{1}{4}$ this coefficient is notably weaker than other noted mixing coefficient β, ϕ, ρ and ψ (see Doukhan (1994) and Rio (2000)). the results obtained in the case α mixing will therefore concern a wider class of processes.*

2.2 Nadaraya-Watson estimator

Suppose we have an n-sample $(X_1, Y_1), ..., (X_n, Y_n)$ random variables with real values, same law as the couple (X,Y). We propose to build an estimator $\hat{r}_n(x)$ of the regression function from pairs of observations $\{(X_1, Y_1), ..., (X_n, Y_n)\}$ The first estimator encountered in the literature is the Nadaraya-Watson kernel estimator (Nadaraya (1964) and Watson (1964)),noted **[NW]** estimator it is built from a kernel function K(.) and a bandwidth h, analogously to the kernel estimator of the density function $f_X(.)$ introduced by Parzen (1962) and Rosenblatt (1956) noted [PR] We recall the definition of the estimator [PR],

$$\hat{f}_{X,n}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), x \in \mathbb{R} \tag{2.8}$$

First, we designate a window or bandwidth $h_n : n \geq 1$ (possibly strictly positive numbers verifying

$$h_n \longrightarrow 0, \quad \text{when} \quad n \longrightarrow \infty$$

the bandwidth $h = h_n$ denotes a sequence indexed by n=1,2,....but dependence on n does will not always be specified in order to lighten the ratings. The kernel function K:$\mathbb{R} \to \mathbb{R}$ will be assumed to be measurable and satisfying certain basic hypotheses among those set out below:

(K.1) K is bounded ie : $\sup_{u\in\mathbb{R}} |K(u)| < \infty$;

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

(K.3) $k(.) \in L_1(\mathbb{R})$ i.e. $\int |K(u)|du < \infty$;

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

The estimator [NW] is presented as a weighted local average of the values Y_i and is defined by

$$\hat{r}_n^{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} \times \mathbb{1}\left\{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \neq 0\right\}, \tag{2.9}$$

where $\mathbb{1}\{\} = \mathbb{1}_{\{\}}$ designates the indicator function. Remember that for any event A Borel-measurable,

$$\mathbb{1}(A) = \begin{cases} 1, & \text{if A is checked ;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.10}$$

Similarly, we can define the estimator [NW] by,

$$\hat{r}_n^{NW}(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}, & \text{when } \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \neq 0 \\ \frac{1}{n}\sum_{i=1}^{n} Y_i, & \text{otherwise.} \end{cases} \tag{2.11}$$

The kernel K determines the shape of the neighborhood around the point x and the window h controls the size of this neighborhood, i.e. the number of observations taken to local average . Intuitively, it is natural that the window h is preponderant for the consistency of the estimator [NW]. This observation will be confirmed in the next section and in the following paragraph .

Posing

$$\hat{m}^n(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right) \quad \textit{Kernel estimator of } m(x) \tag{2.12}$$

we notice that the estimator [NW] can be written $\hat{r}_n^{NW}(x) = \hat{m}^n(x)/\hat{f}_{X,n}(x)$. This last formulation is common in the literature and consists of a good first approach to the estimator [NW]. De facto, we will treat the random numerator and denominator separately in order to obtain the usual asymptotic properties of the estimator [NW], because it is difficult to work directly with a random quotient. The method consists then to linearize the deviation $\hat{r}_n^{NW}(x) - r(x)$ in terms of $\hat{f}_{X,n}(x) - f_X(x)$ and $\hat{m}^n(x) - m(x)$. This technique is central (even systematic) in nonparametric regression,

First comments on the estimator [NW]

The estimator [NW] (2.9) is linear in the sense of the definition 2.1.1 with as function weight $W_{ni}^{NW}(.)$ defined by

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

Remark 2. *For a more general discussion of the weight function in the non-parametric regression framework and an exposure of certain conditions necessary for its consistency, we will cite the pioneering article of Stone (1977)Also note that, by restricting our study to positive kernels (i.e., such as $K \geq 0$), the indicator function presented in (2.9) disappears.*

Among the two parameters K (functional) and h (numeric) to be selected, the window h determines the degree of smoothing of the estimator [NW].Suppose the estimator is only evaluated at observation points $\{X_i : 1 \leq i \leq n\}$ then, when K is at compact support we get

$$\lim_{h \to 0} \hat{r}_n^{NW}(X_i) = K(0)Y_i/K(0) = Y_i$$

Specifically, we have

$$\lim_{h \to 0} \hat{r}_n^{NW}(X_i) = \begin{cases} Y_i, & \text{when } x = X_i, \quad \forall 1 \leq i \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

When h tends to zero, the estimator [NW] therefore tends to reproduce the data, the curve obtained is close to an interpolation of the points$\{(X_i, Y_i) : 1 \leq i \leq n\}$ It's a phenomenon of under-smoothing, the variance of the estimator is too large. On the other side

$$\lim_{h \to 0} \hat{r}_n^{NW}(X_i) = \frac{\sum_{i=1}^{n} K(0)Y_i}{\sum_{i=1}^{n} K(0)} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

When h tends to infinity, we have a phenomenon of highlighting, the estimator $\hat{r}_n^{NW}(x)$ tends to $n^{-1}\sum_{i=1}^{n} Y_i$ which is a function independent of x. The deterministic error or bias is too big. This observation indicates that the statistical properties of the estimator [NW] depend on the window or smoothing parameter h, which we will have to choose in order to balance the bias and the variance.

Now, we will discuss one of the many ways to build the estimator of the regression function introduced by Nadaraya and Watson. For an intuitive justification of the estimator [NW], let's recall the definition of the bivariate density kernel estimator, natural extension of (2.8),

$$\hat{f}_{X,Y,n}(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) K\left(\frac{y-Y_i}{h}\right), \tag{2.13}$$

By replacing in (2.2) the joint density $f_{X,Y}$ and the marginal density $f_X$by their respective kernel estimators [PR], we find the estimator [NW] defined in (2.9) or (2.11). The following proposition follows

Proposition 2.2.1. *If the kernel K is symmetrical (or of order 1), we obtain the following equalities*

$$\hat{r}_n^{NW}(x) = \frac{\int_{\mathbb{R}} y\hat{f}_{X,Y,n}(x,y)dy}{\int_{\mathbb{R}} \hat{f}_{X,Y,n}(x,y)dy} = \int_{\mathbb{R}} y\hat{f}_{X,Y,n}(x,y)dy/\hat{f}_{X,n}(x) \tag{2.14}$$

from (2.13) We have

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

similar

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

which demonstrates (2.14).
The definition (2.13) leads us to introduce the estimator [NW] in the multivariate framework When the explanatory or predictive variable X has values $\mathbb{R}^P$, for a certain $p \in \mathbb{N}$ fixed, The estimators [PR] and [NW] are defined by:

$$\hat{f}_{X,n}(x) = \frac{1}{nh^p}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), x \in \mathbb{R}^p \tag{2.15}$$

and

$$\hat{r}_n^{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} \times \mathbb{1}\left\{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \neq 0\right\}, \tag{2.16}$$

Above $K : \mathbb{R}^p \to \mathbb{R}$ denotes a multivariate function defined as the product of univariate kernels K_j (possibly identical for 1,...,j,) such as

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

Remark 3.

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

a subset of $\mathbb{R}$ corresponding to the space of all possible windows. The definition (2.15) of the density estimator [PR] is a special case of the estimator following:

$$\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i),$$

with

$$K_h(X) = \prod_{j=1}^{p} K_j(\frac{x_j}{h_j})$$

It is possible to present the multivariate [PR] estimator in an even more general context . Let H be a non-singular $p \times p$ matrix (i.e. not admitting an eigenvalue null and therefore invertible) belonging to the space of square matrices $\mathcal{M}_p(\mathbb{R})$ we use the multivariate kernel $K : \mathbb{R}^p \to \mathbb{R}$ which satisfies the following conditions:

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

(K.2) $\int_{\mathbb{R}} uK(u)du = 0$ property of symmetry

So the kernel density estimator is defined, in its most general form, by

$$\hat{f}_{X,n}(x) = \frac{1}{n|\mathbf{H}|} \sum_{i=1}^{n} K(\mathbf{H}^{-1}(x - X_i)), \quad x \in \mathbb{R}^p \tag{2.17}$$

where $|\mathbf{H}|$ denotes the determinant of the matrix $\mathbf{H}$ By taking up the above notations in the definitions (2.15) and (2.16) the window matrix is the form $\mathbf{H} = h\mathbf{I}_p$ where $\mathbf{I}_p$ designates the matrix $p \times p$ identity. In other words, we have chosen in each direction the same window $h = h_i, i = 1, ..., p$ The kernel $\mathbf{K}$ can also be spherical, that is to say as

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

where W denote a univariate kernel with compact support and$\|.\|_p$ is the Euclidean norm on $\mathbb{R}^p$ However, when we base ourselves on the definition (2.15)the kernel support is rather rectangular, reference is made to Scott (1992) p. 152-155, for more details on density and regression estimation in the multivariate framework

Alternative estimators

The random denominator in (2.9) is a major drawback, especially for the study derivatives of the estimator [NW]. Within the framework of the experimental device where the variables

X_i are ordered, Gasser and Müller (1979) proposed the following estimator:

$$\hat{r}_n^{GM}(x) = \sum_{i=1}^{n} \left\{ \int_{s_i-1}^{s_i} K\left(\frac{x-t}{h}\right) dt \times Y_i \right\}, \tag{2.18}$$

with $s_i = (X_i + X_{i+1})/2, X_0 = -\infty$ and $X_{n+1} = +\infty$ This estimator is linear at meaning of definition 2.1.1, with a weight function without denominator and summable from (2.18), the weight function is defined by

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

Gasser and Müller's [GM] estimator is a modification of an earlier version developed by Priestley and Chao (1972) . For a complete study of the estimator [GM], we cite the work of Müller (1988).
When the marginal density function f_X is known, there is a slightly different from the estimator [NW], proposed by (cf.Johnston (1979) and Johnston (1982)),

$$\hat{r}_n^J(x) = \frac{1}{nh} \sum_{i=1}^{n} f_X(X_i)^{-1} Y_i K\left(\frac{x-X_i}{h}\right) / f_X(x) \tag{2.19}$$

The estimator $\hat{r}_n^J(.)$also refers to the experimental device with fixed effects because the density function f_X is known. The bias of the estimator is close to the estimator [NW] . Following Wand and Jones (1995) , p. 152, we present the estimator

$$\hat{r}_n^*(x) = \frac{1}{nh} \sum_{i=1}^{n} f_X(X_i)^{-1} Y_i K\left(\frac{x-X_i}{h}\right), \tag{2.20}$$

which has a better bias than the estimator [NW] or the estimator $\hat{r}_n^J(x)$.The bias of the estimator defined in (2.20) is equivalent to that of the locally linear estimator the restriction of our presentation of regression estimators to the kernel method may be excused by the following remark: two other important classes of estimators, the splines and the nearest neighbors correspond to estimators with kernel constructed with particular windows, of the form $f_X^{-\alpha}, 0 \leq \alpha \leq 1$(cf.Jennen-Steinmetz and Gasser (1988), for appropriate references).

2.3 Consistency of The Nadaraya Watson Estimator

The kernel regression estimator is therefore dependent on the choice of two parameters, the window h and kernel K. We will see in the following sections that the crucial parameter is the window to get good asymptotic properties. However, the kernel should not to be neglected, it reduces the bias of our estimator by relying on the regularity properties of the regression curve. In this section we will determine the conditions on the window and the kernel necessary for the consistency of the estimator [NW]. We obtain the consistency of the estimators of the type [NW], via the following biasvariance decomposition:

$$\mathbb{E}\left[\left(\hat{r}_n^{NW}(x) - r(x)\right)^2\right] = Var\left[\hat{r}_n^{NW}(x)\right] + \left(\mathbb{E}\left(\hat{r}_n^{NW}(x) - r(x)\right)\right)^2 \tag{2.21}$$

We denote by $\overset{L_2}{\to}$ respectively $\overset{\mathbb{P}}{\to}$ norm convergence L_2(respectively in probability) when (2.21) tends to zero, it follows

$$\hat{r}_n^{NW}(x)\overset{L_2}{\to}r(x) \quad \textit{which} \quad \textit{implies} \quad \hat{r}_n^{NW}(x)\overset{\mathbb{P}}{\to}r(x) \tag{2.22}$$

In view of (2.22), a simple study of the criteria of convergence towards zero of the bias and the variance above will specify the conditions necessary for the consistency of the estimator [NW]. We also note that the loss characterized above is a very measure practice of the performance of our estimator, it will be used to determine the asymptotically optimal parameters.

2.3.1 Variance calculation

We begin the study of the estimator [NW] by calculating its variance and its expression asymptotic see Blondin (2004). The kernel K is supposed to verify the hypotheses (K.1–4). We note that (K.1) and (K.3) imply that K(·) is twice integrable . We ask, for convenience

$$\sigma^2(x) = Var(Y \mid X = x) = \frac{1}{f_X(x)} \int y^2 f_{X,Y}(x, y)dy - (r(x))^2$$

when this expression is well defined.

Proposition 2.3.1. *we suppose $\mathbb{E}(Y^2) < \infty$ At each point of continuity of functions r(x),$f_X(x)$ and $\sigma^2(x)$ such as $f_X(x) > 0$,*

$$Var\left(\hat{r}_n^{NW}(x)\right) = \frac{1}{nh} \times \left(\frac{\sigma^2(x)}{f_X(x)} \int_{\mathbb{R}} K^2(u)du\right)(1 + o(1)) \tag{2.23}$$

where the term $o(1)$ tend to 0 when $h \to 0$

Using Bochner's lemma , we obtain easily

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

when $h \to 0$ let the function $s(x) = \displaystyle\int y^2 f_{X,Y}(x, y)dy$ we have

$$\begin{aligned} Var\left(\hat{m}_n(x)\right) &= \frac{1}{nh^2}\left(\mathbb{E}\left(Y^2K^2\left(\frac{x - X_i}{h}\right)\right) - \mathbb{E}\left(YK\left(\frac{x - X_i}{h}\right)\right)^2\right) \\ &= \frac{1}{nh}\left(\int_{\mathbb{R}} K^2(u)s(x - hu) - h\left(\int_{\mathbb{R}} K(u)r(x - hu)\right)^2\right) \\ &= \frac{1}{nh}s(x)\int_{\mathbb{R}} K^2(u)du(1 + o(1)), \end{aligned}$$

Similarly

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

Let be the vector

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

and $\sum(A_n(x))$ its covariance variance matrix. It follows

$$\sum(A_n(x))=\frac{1}{nh}\begin{pmatrix}f_X(x) & r(x)\\ m(x) & s(x)\end{pmatrix}\int_{\mathbb{R}}K^2(u)du(1+o(1))$$

Noting that

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

We get

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

Remark 4. *In the asymptotic expression of the terms of variance of kernels estimators , we invariably find the quantity:*

$$\int_{\mathbb{R}}K^2(u)du=\|K\|_2^2. \tag{2.24}$$

To ensure the finiteness of this integral, we can choose the kernel function $K(\cdot)$ with boundary variation on $\mathbb{R}$ and compact support, noting that the latter assumptions clearly imply (K.1–3)For asymptotic optimality, the variance minimal will be obtained by minimizing (2.24) along K in a certain class of kernels fixed In conclusion, if the window h_n satisfies the conditions $h_n \longrightarrow 0$ and $nh_n \longrightarrow \infty$ when $n \to \infty$ the variance of the estimator [NW] tends to zero.

Multidimensional extension

Let x and u be vectors of $\mathbb{R}_p$ The asymptotic variance has an expression similar to univariate case. We recall that

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

Where $K:\mathbb{R}^p\to\mathbb{R}$ the kernel function product of univariate kernel verifying (K.1–4).

Proposition 2.3.2. *Blondin (2004) We suppose $\mathbb{E}(Y^2) < \infty$.so at each point of continuity of functions m(x),$f_X(x)$ and $\sigma^2(x)$ such as $f_X(x) > 0$, we have*

$$Var\left(\hat{r}_n(x)\right) = \frac{1}{nh^p} \times \left(\frac{\sigma^2(x)}{f_X(x)} \int_{\mathbb{R}} K^2(u)du\right)(1 + o(1)) \tag{2.25}$$

Where the term $o(1)$ tend to 0 when $h \to 0$ We obtain,

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

$$Var\left(\hat{m}_n(x)\right) = \frac{1}{nh}\left[s(x)\int_{\mathbb{R}} K^2(u)du\right](1 + o(1))$$

The rest of the demonstration is similar to the univariate framework and will not be presented by conciseness if the window $h_n$satisfies the conditions $h_n \longrightarrow 0$ and $nh_n^p \longrightarrow \infty$ when $n \to \infty$ the variance of the multivariate estimator [NW] tends to zero.

2.3.2 Calculation of bias

Blondin (2004)

The treatment of bias is purely analytical and is essentially based on Taylor's development we must assume certain regularity conditions on the functions r(x) and $f_X(.)$ which will determine the order of asymptotic bias as a function of the parameter smoothing h. The estimator [NW] is in the form of a random quotient, it is why we generally use the following approximation as a centering term

$$\tilde{\mathbb{E}}(\hat{r}_n^{NW}(x)) = \frac{\mathbb{E}\left(\hat{m}_n(x)\right)}{\mathbb{E}\left(\hat{f}_{X,n}(x)\right)} \tag{2.26}$$

The formula (2.26) is easier to handle and allows in particular the linearization of the deviation

$$\tilde{d_n}(x) = \hat{r}_n^{NW}(x) - \tilde{\mathbb{E}}(\hat{r}_n^{NW}(x))$$

we obtain

$$\tilde{d_n}(x) = \left[\hat{m}_n(x) - \mathbb{E}(\hat{m}_n(x))\right] \times \frac{1}{\mathbb{E}\left(\hat{f}_{X,n}(x)\right)} - \left(\hat{f}_{X,n}(x) - \mathbb{E}(\hat{f}_{X,n}(x))\right) \times \frac{\hat{m}_n(x)\hat{f}_{X,n}(x)}{\mathbb{E}\left(\hat{f}_{X,n}(x)\right)}$$

The proposition below demonstrated by Nadaraya (1989) (cf. p. 116-117,) justifies the choice of the centering term (2.26)

Proposition 2.3.3. *When Y is bounded and $nh_n \longrightarrow \infty$,*

$$\mathbb{E}(\hat{r}_n^{NW}(x)) = \tilde{\mathbb{E}}(\hat{r}_n^{NW}(x)) + \mathcal{O}((nh)^{-1}) \tag{2.27}$$

when $\mathbb{E}(Y^2) < \infty$ and $nh_n^2 \longrightarrow \infty$

$$\mathbb{E}(\hat{r}_n^{NW}(x)) = \tilde{\mathbb{E}}(\hat{r}_n^{NW}(x)) + \mathcal{O}((n^{1/2}h)^{-1}) \tag{2.28}$$

We use the following identity

$$\frac{1}{\hat{f}_{X,n}(x)} = \frac{1}{\mathbb{E}\left(\hat{f}_{X,n}(x)\right)} - \frac{\hat{f}_{X,n}(x) - \mathbb{E}\left(\hat{f}_{X,n}(x)\right)}{(\hat{f}_{X,n}(x))^2} + \frac{\left(\hat{f}_{X,n}(x) - \mathbb{E}\left(\hat{f}_{X,n}(x)\right)\right)^2}{\hat{f}_{X,n}(x)\left(\mathbb{E}\left(\hat{f}_{X,n}(x)\right)\right)^2}$$

We multiply by $m_n(x)$ on both sides, then we pass to the expectation

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

let $s(x) = \int_{\mathbb{R}} y^2 f_{X,Y}(x,y)dy$ We calculate the asymptotic variance of $\hat{m}_n(x)$ then $\hat{f}_{X,n}(x)$via Bochn's lemma

$$\begin{aligned}
Var(\hat{m}_n(x)) &= \frac{1}{nh}\int_{\mathbb{R}} K^2(u)s(x-uh)du - \frac{1}{n}\left(\int_{\mathbb{R}} K(u)r(x-uh)\right)^2 \\
&\approx \frac{1}{nh}s(x)\int_{\mathbb{R}} K^2(u)du.
\end{aligned}$$

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

Using the Cauchy-Schwarz inequality combined with the above formulas, we obtain

$$a_n(x) = \mathcal{O}\left(\frac{1}{nh}\right) \tag{2.29}$$

When the variable Y is bounded i.e $|Y| \leq M$ for a certain constant M fixed, we note that the estimator [NW] is also naturally bounded,

$$\frac{\hat{m}_n(x)}{\hat{f}_{X,n}(x)} = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} \leq \frac{\sum_{i=1}^{n} M \times K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} = M \tag{2.30}$$

This last inequality (2.30) makes it possible to bounded $b_n(x)$,

$$\begin{aligned} b_n(x) &\leq M \times \mathbb{E}\left(\left[\hat{f}_{X,n}(x) - \mathbb{E}(\hat{f}_{X,n}(x))\right]^2\right) \\ &\approx \frac{M}{nh} f_X(x) \int_{\mathbb{R}} K^2(u)du = \mathcal{O}\left(\frac{1}{nh}\right) \end{aligned} \tag{2.31}$$

The relations (2.29) and (2.31) entail (2.27). when $\mathbb{E}(Y^2) < \infty$ we have

$$\begin{aligned} b_n(x) &\leq \mathbb{E}\left(\max_{1\leq i\leq n} |Y_i| \left[\hat{f}_{X,n}(x) - \mathbb{E}(\hat{f}_{X,n}(x))\right]^2\right) \\ &\leq \left(\sum_{i=1}^{n} Y_i^2\right)^{1/2} \times \left(\mathbb{E}\left[\left(\hat{f}_{X,n}(x) - \mathbb{E}(\hat{f}_{X,n}(x))\right)^4\right]\right)^{1/2} \\ &= \sqrt{n}\left(\mathbb{E}(Y^2)\right)^{1/2} \times \mathcal{O}\left(\frac{1}{nh}\right) = \mathcal{O}\left(\frac{1}{n^{1/2}h}\right) \end{aligned} \tag{2.32}$$

The relations (2.29) and (2.32) imply (2.28), the demonstration is completed
We are now ready to state the asymptotic bias of the estimator [NW]. We will assume the bounded variable Y, so that (2.27) is satisfied. We will see than the estimator bias [NW], according to the regularity properties of the curve regression, is a functional of the derivatives of regression.

Proposition 2.3.4. *suppose that $r(.)$ and $f_X(.)$ are $C^2(\mathbb{R})$ class and that the kernel K is order 2 i.e. such that*

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

we have,when $h \longrightarrow 0$ and $nh \longrightarrow \infty$,

$$\mathbb{E}(\hat{r}_n^{NW}(x)) - r(x) = \frac{h^2}{2} \times \left[\left(m''(x) + 2m'(x)\frac{f'_X(x)}{f_X(x)} \int_{\mathbb{R}} u^2K(u)du\right)\right](1 + o(1)) \tag{2.33}$$

Remark 5. *we note that the term $o(1)$ in (2.33) above decompose as follows $(\mathcal{O}(h) + \mathcal{O}((nh)^{-1}))$ from (2.27)*

$$\begin{aligned} \mathbb{E}(\hat{r}_n^{NW}(x)) - r(x) &= (\mathbb{E}\left[K((x-X)/h)\right])^{-1} \\ &\times \left(\int \frac{1}{h}K\left(\frac{x-t}{h}\right)m(t)dt - m(x) + m(x) - r(x)\int \frac{1}{h}K\left(\frac{x-t}{h}\right)f_X(t)dt\right) \\ &\approx \frac{h^2}{2} \times (f_X(x))^{-1} \times (m''(x) - r(x)f''_X(x)) \times \int_{\mathbb{R}} u^2K(u)du \\ &= \frac{h^2}{2}\left(r''(x) + 2r''(x)\frac{f'_X(x)}{f_X(x)}\right) \times \int_{\mathbb{R}} u^2K(u)du \end{aligned} \tag{2.34}$$

The sign $\approx$ above denotes an error of the order $\mathcal{O}(h)$ or $o(1)$ from to the lemma of Bochner. Proposition 2.3.3 and (2.34) imply (2.33). The term asymptotic bias reveals the derivative of the functions m(x) and $f_X(.)$. This is due to the fact that the estimator [NW] performs a least squares approximation locally constant values Y_i . The estimator [NW] therefore suffers from a high bias in the region where the derivative of the true regression function is large. The bias can also be large when $f'_X/f_X(x)$ is large. In comparison, under assumptions similar to those of Proposition 2.3.4, the estimator [GM] has a better bias:

$$\mathbb{E}(\hat{r}_n^{NW}(x)) - r(x) = \frac{h^2}{2} \times \left[\left(m''(x) \times \int_{\mathbb{R}} u^2 K(u)du\right)\right](1 + o(1)) \tag{2.35}$$

The form of the above asymptotic bias is preferable from a statistical point of view, because it does not depend on the density f_X and its derivative. For example, if the curve of regression is a straight line, the main bias term disappears whatever the form of the marginal density f_X. When the regression function admits additional regularity conditions, it is possible to reduce the asymptotic bias of the estimator [NW] using a **kernel superior order.** Let q be a fixed natural integer

Definition 2.3.5. *The kernel K is called kernel of order q if it satisfies the following conditions*

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

To illustrate the usefulness of higher order kernels, we consider the simple example density estimation. The density kernel bias is written

$$\mathbb{E}\left(\hat{f}_{X,n}(x)\right) - f_X(x) = \int (f_X(x - hu) - f_X(x))\, K(u)du$$

Now suppose that the density $f_X(x)$ admits bounded derivatives up to the order q in the vicinity of point x. So we get, via Taylor development,

$$\mathbb{E}\left(\hat{f}_{X,n}(x)\right) - f_X(x) = \sum_{k=1}^{q-1}\left(h^k \frac{(-1)^k}{k!} f_X^{(k)}(x) \int u^k K(u)du\right) + \mathcal{O}(h^q) \tag{2.36}$$

The formula (2.36) above clearly shows the importance of the nuclei of which the first moments are zero: a kernel of order q reduces the bias to order $\mathcal{O}(h^q)$ modulo some regularity assumptions. In the multivariate framework, we have the following orthogonality conditions,

$$\int_{\mathbb{R}^p}\left(\prod_{i=1}^{n}\right) \times K(u_1, ..., u_p)du_1....du_p = 0 \quad \textit{when} \quad \sum_{i=1}^{n} s_i = 1, 2, ..., q-1 \tag{2.37}$$

If (2.37) is verified and

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

The multivariate kernel K(·) is called the multivariate kernel of order q, that is, all of its moments up to the order q-1 are zero.
for convenience, we denote by $[\mu_j(K)]$ the moment of order k associated with the kernel function K (·), when $j \in \mathbb{N}$.

Proposition 2.3.6. *[Blondin (2004)] suppose that r(.) and* $f_X(.)$ *are* $C^q(\mathbb{R})$ *class and that the kernel K is order 2 i.e. such that* $[\mu_0(K)] = 1$ *,*$[\mu_j(K)] = 0, \quad 1 \le j \le q-1, \quad$ *and*

$[\mu_j(K)] < \infty$ *,when* $h \longrightarrow 0$ *and* $nh \longrightarrow \infty$*, we have*

$$\mathbb{E}(\hat{r}_n^{NW}(x)) - r(x) = \frac{h^q}{q!} \times \left[\left(r^{(q)}(x) + q \times r^{q-1}(x) \frac{f'_X(x)}{f_X(x)} \right) [\mu_j(K)] \right] (1 + o(1)) \tag{2.38}$$

First, we consider the expectation of $\hat{m}_n$:

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

then

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

The rest of the demonstration is similar to the demonstration of the proposition 2.3.4

Multidimensional extension : $X \in \mathbb{R}$

We specify some notations necessary for the presentation of asymptotic bias in the multivariate frame. let $f : \mathbb{R}^p \to \mathbb{R}$any multivariate function. We designate by Q the operator on f defined by,

$$\mathrm{Q}\left[f\right](x) = \int_{\mathbb{R}^p} \left[u^T (\nabla^2 f(x)) u \right] K(u) du$$

where $\nabla^2 f(x)$ denotes the Hessian matrix of partial derivatives of order 2 of the function f(·) at point x.

Proposition 2.3.7. *when* Y *is boundary and* $nh^p \longrightarrow \infty$*,*

$$\mathbb{E}(\hat{r}_n^{NW}(x)) = \tilde{\mathbb{E}}(\hat{r}_n^{NW}(x)) + \mathcal{O}((nh^p)^{-1}) \tag{2.39}$$

suppose that $r(.)$ *and* $f_X(.)$ *are* $C^2(\mathbb{R}^p)$ *class and that the kernel* K *is order 2 we have* $h \to 0$ *and* $nh^p \longrightarrow \infty$*,*

$$\mathbb{E}(\hat{r}_n^{NW}(x)) - r(x) = \frac{h^2}{2} \left(\frac{\mathrm{Q}\left[m\right](x) - r(x)\mathrm{Q}\left[f_X\right](x)}{f_X(x)} \right) (1 + o(1)) \tag{2.40}$$

We can also formulate the asymptotic bias (2.40) more explicitly but less compact

$$\frac{h^2}{2} \left(\sum_{j=1}^{p} \left(\frac{\partial^2}{\partial x_j} m(x) + 2 \left(\frac{\partial}{\partial x_j} m(x) \right) \left(\frac{\partial}{\partial x_j} f_X(x) \right) \frac{1}{f_X(x)} \right) \int_{R^p} u_j^2 K(u) du \right) \tag{2.41}$$

2.4 Asymptotic optimality and choice of parameters

In last section, we have established the necessary and sufficient conditions on the window h_n to get the consistency of the estimator [NW] $h_n \longrightarrow 0$ and $nh_n \longrightarrow \infty$ when $n \to \infty$ We now propose to determine the optimal window, within the meaning of a certain criterion asymptotic efficacy. We will look for the window that minimizes the L_2 loss associated with the estimator [NW] by fixing the kernel K in a certain class. Then, we will be interested in the optimality of the kernel. We denote by $\mathcal{K}[q]$ the class of kernels of order q with compact and bounded support. We suppose, throughout this section, that the kernel $K \in \mathcal{K}[q]$. The bounded K hypothesis and with compact support is very classic in non-parametric regression, it implies in particular the integrability of the various moments of the kernel function K(·). Under the hypotheses of proposition 2.3.6, we have,

$$\begin{aligned}\mathbb{E}(\hat{r}_n^{NW}(x)) - r(x) &= \frac{h^q}{q!} \times \left[\left(r^{(q)}(x) + q \times r^{q-1}(x)\frac{f'_X(x)}{f_X(x)}\right)[\mu_j(K)]\right](1 + o(1)) \\ &= \frac{h^q}{q!} \times [b(x;q)]\,(1 + o(1)) \qquad (2.42)\end{aligned}$$

Under the hypotheses of proposition 2.3.1, via (2.23), it follows

$$\begin{aligned}Var(\hat{r}_n^{NW}(x)) &= \frac{1}{nh} \times \left(\frac{\sigma^2(x)}{f_X(x)}[\mu_0(K^2)]\right)(1 + o(1)) \\ &= \frac{1}{nh} \times [v^2(x)](1 + o(1)). \qquad (2.43)\end{aligned}$$

These asymptotic developments are recurrent in asymptotic optimization, because the optimal window equilibrate bias and variance. There are basically two types of procedures for the selection of the smoothing parameter: the local approach and the overall. In view of punctual or uniform results, we will choose the appropriate procedure, that is to say the local approach for results of the point convergence type and the global approach for uniform convergence type results.

Local selection criteria: AMSE

We consider as efficiency criterion the famous mean square **error or MSE ("mean squared error")**. From formulas (2.42) and (2.43), we can present the theorem specifying the exact asymptotic behavior of the quadratic risk of the estimator [NW] $\hat{r}_n^{NW}(x)$ at point x.

Theorem 9. *Blondin (2004) Under the hypotheses of propositions 2.3.6 and 2.3.1, we obtain,*

$$\begin{aligned}[MSE]\left(\hat{r}_n^{NW}(x)\right) &= \mathbb{E}\left[\left((\hat{r}_n^{NW}(x)) - r(x)\right)^2\right] \\ &= \left(\frac{h^{2q}}{(q!)^2} \times [b(x;q)]^2 + \frac{1}{nh} \times [v^2(x)]\right)(1 + o(1)) \qquad (2.44)\end{aligned}$$

from (2.42) and (2.43)

$$\begin{aligned}[MSE]\left(\hat{r}_n^{NW}(x)\right) &= \left(\mathbb{E}\left[\left((\hat{r}_n^{NW}(x)) - r(x)\right)^2\right]\right)^2 + Var\left((\hat{r}_n^{NW}(x))\right) \\ &= \frac{h^{2q}}{(q!)^2} \times [b(x;q)]^2\,(1 + o(1)) + \frac{1}{nh} \times [v^2(x)](1 + o(1)). \qquad (2.45)\end{aligned}$$

From theorem 9 and formula (2.44), we get the expression of the error asymptotic quadratic mean or AMSE ("asymptotic mean squared error"):

$$[MSE]\left(\hat{r}_n^{NW}(x)\right) = \frac{h^{2q}}{(q!)^2} \times [b(x;q)]^2 + \frac{1}{nh} \times [v^2(x)] = [AMSE](h, K) \qquad (2.46)$$

Note that the asymptotic quadratic risk (2.46) depends on the kernel K and the window h associated with the estimator [NW]. We first assume the kernel K fixed.The optimal window, within the meaning of the local criterion for minimizing the AEMS at point x, is then obtained by minimizing From h the quantity (2.46), that is to say

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

the bandwidth $h^{MSE}(K)$ is solution of the following equation:

$$\frac{h^{2q}}{(q!)^2} h^{2q-1} \times [b(x;q)]^2 + \frac{1}{nh^2} \times [\upsilon^2(x)] = 0$$

when $[b(x;q)] \neq 0$ we obtain

$$\begin{aligned} h^{MSE}(K) &= n^{-1/(2q+1)} \left(\frac{(q!)^2 [\upsilon^2(x)]}{2q\left([b(x;q)]\right)^2} \right)^{1/(2q+1)} \\ &= n^{-1/(2q+1)} \left(\frac{q!(q-1)!\left(\frac{\sigma^2(x)}{f_X(x)}[\mu_0(K^2)]\right)}{2\left(r^{(q)}(x) + q \times m^{(q-1)}(x)\frac{f_X'(x)}{f_X(x)}\right)^2 [\mu_q(K)]^2} \right)^{1/(2q+1)} \end{aligned} \tag{2.47}$$

the bandwidth $h^{MSE}(K)$ therefore asymptotically minimizes the MSE of the estimator [NW] to point x (local criterion). After calculations, it follows

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

To simplify our writing, we can consider the special case q=2, which corresponds to the study frame where the nucleus is positive or of order 2 from (2.47), when q=2

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

We get, as a result:

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

For convenience, we introduce the notations:

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

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

Corollary 1. *We assume the hypotheses of Theorem 9 verified. We have, if $\hat{r}_n^{NW}(x)$ is built with the window $h = h^{MSE}(K)$ (oracle estimator)*

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

Remark 6. *the random function $\hat{r}_n^{NW}(x)$ define in (2.11) with the window $h^{MSE}(K)$is not an estimator, stricto sensu, because it depends on the regression function at to estimate. This type of function is called pseudo-estimator or oracle estimator in the literature. The corollary above therefore has no interest in practice because it does not allow to build an estimator. It is however possible to replace the unknown quantities by consistent preliminary estimators.*

The optimal bandwidth $h^{MSE}(K)$ allows to determine the optimal speed of convergence of the quadratic risk (close to $1/n$) when the kernel is fixed in the function class $\mathcal{K}[q]$ We are now interested in the optimality of the kernel on $\mathcal{K}[q]$ It should be noted that the choice of the kernel only has an impact on the limit constant, via C[K,q]. The problem of the optimal choice of the K core is summarized as follows:

$$K_{opt}^{MSE} = \text{argmin}_{K\in\mathcal{K}[q]}\left\{[\mu_0(K^2)]^{2q/(2q+1)}[\mu_q(K)]^{2/(2q+1)}\right\} \tag{2.48}$$

Note that the Epanechnikov kernel is a solution of the problem (2.48) when q=2 and the support of the kernel $[-1,1]$. We recall the definition of the Epanechnikov kernel

$$K^E(u) = \frac{3}{4}\left(1-u^2\right)\mathbb{1}\{|u|\le 1\}$$

which provides the minimum value $C[K^E,2] = 3^{4/5}5^{-6/5}$ We can then give the expression of the corresponding optimal window:

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

Global selection criterion: AMISE

Now, we are interested in the estimation of the regression function on an interval $I \subseteq \mathbb{R}$ and the overall risk of the estimator [NW] over this interval. We introduce for this the error mean integrated squared error (MISE),

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

from Tonelli-Fubini theorem

Theorem 10. *Blondin (2004) Suppose the hypotheses of propositions 2.3.6 and 2.3.1*

$$[MISE]\left(\hat{r}_n^{NW}(x)\right) = n^{-1/(2q+1)}\left(\frac{h^{2q}}{(q!)^2}\int_I ([b(x,q)])^2\,dx + \frac{1}{nh}[\upsilon^2(x)]dx\right)(1+o(1)) \quad (2.49)$$

The optimal window, within the meaning of the overall criteria for minimizing AMISE ("asymptotic mean integrated squared error ") over the interval I, is given by

$$h_{n,opt}^{MISE}(x) = n^{-1/(2q+1)}\left(\frac{q!(q-1)!\int_I[\upsilon^2(x)]dx}{2\int_I([b(x,q)])^2\,dx}\right)^{1/(2q+1)} \quad (2.50)$$

Again, the optimal window depends on unknown parameters and therefore cannot be used in practice. It is proposed to remedy this obstacle via a reference method, the cross validation, presented in the next section.

2.5 The cross validation

In this section, we assume the kernel K fixed, and we are only interested in the choice of window h. We observed in the previous paragraphs that the efficiency of the estimator [N W] is linked to the smoothing parameter, the window h. Choose the window in order to equilibrate a stochastic term (the variance) and a deterministic term (the bias), if possible independently of the regularity properties of the regression curve. In the previous section, the optimal window which minimizes the integrated quadratic risk (MISE) is obtained under specific regularity assumptions and then depends on quantities unknown, functional of the distribution of the couple (X, Y). In order to build a non-oracle estimator that minimizes the quadratic error, other methods must be used the most common of which is called the cross-validation procedure. The main idea of cross validation consists in minimizing, with respect to h, the estimate of a measurement of the MISE. The window h is then not be deterministic, it depends on the observations, like plug-in methods which we will talk about in the next paragraph.

Let $(X,Y),(X_1,Y_1),(X_2,Y_2)$,. . ., random variables i.i.d. with values $\mathbb{R}_p\times\mathbb{R}$. We consider kernel estimators, with random window (or "data-driven bandwidth ") of the form,

$$\hat{h} = \hat{h}_n = h_n\left\{(X_1,Y_1),...,(X_n,Y_n),x\right\} \in H_n, \qquad x \in \mathbb{R}_p$$

When H_n denotes a subset of $\mathbb{R}_+^n$ (i.e., the area of variation $\hat{h}$), let d(.,.) a certain distance, used to define the risk, which will be used to measure the effectiveness of a some m estimator of the regression function. In order to simplify the exposure of the cross validation procedure, we will work with the regression estimator [NW], which will be noted

$\hat{r}_h$ to emphasize its dependence on h,

$$\hat{r}_h(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)} \quad when \quad \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \neq 0,$$

The window selection method is said to be asymptotically optimal by relative to the distance d when we have

$$\lim_{n\to\infty}\left[\frac{d(\hat{r}_k, r)}{\inf\limits_{h\in H_n} d(\hat{r}_k, r)}\right] \overset{a.s}{=} 1 \tag{2.51}$$

Where the notation $\overset{a.s}{=} 1$ indicates almost sure equality. Thereafter, we designate by $\omega(\mathring{u})$ a positive and arbitrary weight function. The different distances considered in this section are:

Quadratic Mean Error:

$$d_M = (\hat{r}, r) = \frac{1}{n}\sum_{i=1}^{n} \left(\hat{r}(X_i) - r(X_i)\right)^2 \omega(X_i);$$

Integrated Quadratic Error

$$d_I = (\hat{r}, r) = \int_{\mathbb{R}^p} \left(\hat{r}(X_i) - r(X_i)\right)^2 \omega(X_i) f_X(x) dx;$$

Conditional Average Integrated Quadratic Error:

$$d_C = (\hat{r}, r) = \mathbb{E}\left(d_I(\hat{r}, r) \mid X_1, ..., X_n\right);$$

Remark 7. *Each of these d_M, d_I or d_C error measures is decompose into one bias squared term and variance term.*

Now, we will present the procedure for selecting the random window $\hat{h}$ for the distance d_I. We can decompose $d_I = (\hat{r}, r)$ as follows

$$\begin{aligned} d_I = (\hat{r}, r) &= \int_{\mathbb{R}^p} \Big(\hat{r}(X_) - r(X_i)\Big)^2 \omega(X_i) f_X(x) dx \\ &= \int_{\mathbb{R}^p} \hat{r}_h^2(x) f_X(x) d(x) - 2\int_{\mathbb{R}^p} \hat{r}_h(x) r(x) \omega(X) f_X(x) d(x) \\ &+ \int_{\mathbb{R}^p} r^2(x) \omega(X) f_X(x) d(x) \end{aligned}$$

As the last integral is independent of h, to minimize the loss associated with the distance d_I as a function of h, it suffices to minimize

$$\int_{\mathbb{R}^p} \hat{r}_h^2(x) f_X(x) d(x) - 2\int_{\mathbb{R}^p} \hat{r}_h(x) r(x) \omega(X) f_X(x) d(x) \tag{2.52}$$

However, this is not practical in practice because the latter quantity depends on unknown functions $r(.)$ and $f_X(.)$. The classic method to get around this difficulty consists in replacing these terms by their empirical versions. We notice that the second term of the integral

$$\int_{\mathbb{R}^p} \hat{r}_h(x) r(x) \omega(X) f_X(x) d(x) = \mathbb{E}\left(\hat{r}_h(x) Y \omega(X)\right)$$

It follows as a natural estimator

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{r}_i(x) Y_i \omega(X_i)\right)$$

where $\hat{r}_i$ Is the estimator called "leave-one-out", defined by

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

The leave-one-out estimator is simply the [N-W] estimator built with the (n-1) random couples $(X_1, Y_1), ..., (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), ..., (X_n, Y_n)$ Likewise, it is possible to approximate the first integral term of (2.52) by

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{r}_i^2(X_i)\omega(X_i)\right)$$

In short, it seems reasonable to choose the window h which minimizes the empirical version of (2.52), ie h which minimizes:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{r}_i^2(X_i)\omega(X_i)\right) - \frac{2}{n}\sum_{i=1}^{n}\left(\hat{r}_i(X_i) Y_i \omega(X_i)\right)$$

This last quantity is equal to

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{r}_i^2(X_i) - Y_i\right)^2 \omega(X_i) - \frac{1}{n}\sum_{i=1}^{n}(Y_i)^2 \omega(X_i)$$

where the second term does not depend on h and therefore does not intervene in minimization. The window selection criterion is reduced to: choose $\hat{h}$ which minimizes

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{r}_i(X_i)\right)^2 \omega(X_i). \tag{2.53}$$

This method is well known in the statistical literature and is called a **cross-validation procedure**. concerning the estimate non-parametric regression. The cross-validation procedure can be interpreted as the best choice of h which makes $\hat{r}_i(X_i)$ an efficient estimator of Y_i at meaning of (2.53). Under the assumptions (A.1–6), p. 1467-1468,Härdle (1985), we have the theorem following

Theorem 11. *The cross-validation procedure, choosing $\hat{h}$ which minimizes CV(h), is asymptotically optimal, in the sense of (2.51),with respect to the distances d_M, d_I and d_C.*

2.6 Asymptotic normality

The first demonstration of the asymptotic normality of the estimator [NW] is due to Schuster (1972) . We also refer to Theorems 1.3 and 1.4 p. 117-120 of Nadaraya (1989) and to Theorem 4.2.1 p. 99 de Härdle (1990) , who propose other demonstration methods. The kernel K is supposed to be bounded, compact and orderly 2. The window h_n is chosen

Theorem 12. *Suppose Y bounded or admitting a moment of order $l > 2$. The functions $f_X(.)$ and $r(x)$ are assumed to be twice continuously differentiable over $\mathbb{R}$ At each point of continuity of $\sigma^2(x)$ such as $f_X(x) > 0$,*

$$(nh)^{1/2}\left(\hat{r}_n^{NW}(x) - r(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(B(x), \upsilon^2(x)\right) \tag{2.54}$$

with

$$\upsilon^2(x) = \frac{\sigma^2(x)}{f_X(x)} \int_{\mathbb{R}} K^2(u)du \qquad (asymptotic\ \ variance),$$

and with

$$B(x) = \left(r''(x) + 2r'(x)\frac{f'_X(x)}{f_X(x)}\right) \int_{\mathbb{R}} u^2 K(u)du \qquad (asymptotic\ \ bias)$$

For a number d of points $x_1, ..., x_d$ of continuity, we have,

$$\left\{(nh)^{1/2}\left(\frac{\hat{r}_n^{NW}(x_i) - r(x_i)}{\upsilon(x_i)}\right)\right\}_{i=1}^{d} \xrightarrow{\mathcal{L}} \mathcal{N}_d\left(B(x_i)_{i=1}^d, \mathbb{I}_d\right)$$

Where $\mathbb{I}_d$ denotes the d-dimensional identity matrix.

Multidimensional extension: $X \in \mathbb{R}^p$

In order to properly state the theorem concerning asymptotic normality, we recapitulate some essential hypotheses, related to the control of bias and variance in the multivariate framework.

Let V_x be a neighborhood of point x. We assume the following conditions on the distribution of the couple (X,Y):

(H1) All partial derivatives of order 2 of r(x) exist on V_x;

(H2) All partial derivatives of order 2 of $f_X(.)$ exist and are continuous on V_x furthermore $f_X(u) > 0$ for all $u \in V_x$;

(H3) the joint density $f_{X,Y}(u, y)$ is continuous on $V_x \times \mathbb{R}$, and all partial order derivatives 2 with respect to the components of the vector u exist and are continuous on $V_x \times \mathbb{R}$

(H4) In the multivariate framework, the kernel function K:$\mathbb{R}^p \to \mathbb{R}$ satisfied
K has a compact support such that $\int_{\mathbb{R}^p} K^2(u)du < \infty$
K is order 2

the window $h = h_n$ verify $h \to 0$ and $nh^p \to \infty$ More precisely, for a bias-variance equilibrate , we choose h of the order $n^{-1/(4+p)}$ We recall the expression of variance and asymptotic bias of the estimator $\hat{r}_n^{NW}(x)$ via (2.25)

$$\frac{1}{nh^p} \times \left(\frac{Var(Y \mid X = x)}{f_X(x)} \int_{\mathbb{R}^p} K^2(u)du \right) = \frac{1}{nh^p} \upsilon^2(x)$$

from(2.40),

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

Assuming the above hypotheses verified, it follows asymptotic normality in the multivariate framework.

Theorem 13. *Müller and Song (1993)*

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

2.7 Estimation of regression by local polynomial method

The estimation of the regression function by the local polynomial method is based on a simple generalization of the estimator [NW]. The main thrust of the approach locally polynomial is to consider the problem of regression from the angle of least squares. Intuitively, this approach is full of common sense, denoting that the regression function m(.) is itself a solution to a least squares problem. Through convenience, we recall the definition of the estimator [NW]: when $K \geq 0$

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

We have, when $K \geq 0$,

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

The regression estimator $\hat{m}_N^{NW}(x)$ can therefore be regarded as the solution of the following weighted least squares problem:

$$\arg\min_{\theta \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - \theta)^2 K\left(\frac{x - X_i}{h}\right) \tag{2.55}$$

In other words, the estimator $\hat{m}_N^{NW}(x)$ is obtained by a least approximation locally constant squares. The principle of locally polynomial estimation consists by the local fit of a polynomial of degree p to the data $\{(X_i, Y_i) : 1 \leq i \leq n\}$. The aim of this section is to present the locally polynomial estimators as well as their fundamental statistical properties.

2.7.1 Construction and definition of the local polynomial estimator

Let p be a fixed natural number. We are looking to fit the polynomial

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

to the data (X_i, Y_i), via the weighted least squares method. First, we assume the existence of the (p + 1)-th derivative of the regression function m(.) at point x. This assumption, although difficult to verify in practice, is essential to theoretically validate the construction of the locally polynomial estimator. We can approximate the regression function m(x) by the locally a polynomial order p. It follows, via the Taylor expansion around point x,

$$\begin{aligned} m(z) &\approx m(x) + m'(x)(z-x) + \frac{m''(x)}{2}(z-x)^2 + ... + \frac{m^{(p)(x)}}{p!}(z-x)^p \\ &\approx \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(z-x)^j = \sum_{j=0}^{p} \beta_j (z-x)^j, \end{aligned} \quad (2.56)$$

when z is located in a neighborhood of point x. Now we locally fit the polynomial (2.56) to the data $\{(X_i, Y_i) : 1 \leq i \leq n\}$by the weighted least squares method with the weight function $K\left(\frac{.-x}{h_n}\right)$ we must minimize with respect to the vector$\beta = (\beta_0, ..., \beta_p)^T \in \mathbb{R}^{p+1}$the following quantity

$$\sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x) \right)^2 K\left(\frac{X_i - x}{h_n}\right) \quad (2.57)$$

As with the estimator [NW], the parameters K and h_n determine the shape and size of the neighborhood around point x. be $\hat{\beta} = (\hat{\beta}_0, ..., \hat{\beta}_p)^T \in \mathbb{R}^{p+1}$the vector which minimizes the expression (2.57)From the equality in (2.56)the k-th derivative $m^{(k)}(x)$ can therefore be estimate by $\hat{\beta}_k \times k!$; for $k = 0, 1, ..., p$ The following definition follows:

Definition 2. *The statistics*

$$\hat{m}_n^{(k)}(x; p) = \hat{\beta}_k \times k!, \qquad 0 \leq k \leq p \quad (2.58)$$

*is **the locally polynomial estimator** of order p of the k-th derivative of the regression $m_{(k)}(x)$,and noted estimator [LP](p) of $m_{(k)}(x)$.*

When k=p=0, we find the estimator [NW], $\hat{m}_n(x; 0) = m^{(k)}(x)$ A particularly interesting example is the case p=1 and k=0. The estimator $\hat{m}_n(x; 1)$ of the regression function is called **the locally linear estimator** and noted $\hat{m}_n^{LL}(x)$. From to (2.57)and (2.58)it is equal to $\hat{\beta}_0$ when $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ designates the solution vector of the following least squares equation:

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

More explicitly, the estimator [LL] is defined by:

$$\hat{m}_n^{LL}(x) = \frac{\hat{r}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{r}_{n,1}(x)\hat{f}_{n,1}(x)}{\hat{f}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{f}_{n,1}(x)\hat{f}_{n,1}(x)} \quad (2.59)$$

where

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

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

We will see, subsequently, that the estimators [LP] are greater than the estimators with [NW] (2.9) and [GM] (2.18) nuclei in the framework of the random experimental set-up. From to (Fan (1992)) [45], the estimator [LL] or [LP](1) has a better bias than the estimator [NW] and better variance than the estimator [GM]. In addition, the estimator [LL] has good minimax properties, it is the best estimator on the class of functions of bounded second derivative regression we refer to the works of (Fan and Gijbels (1996)) for a complete exposition of the properties of estimators [LP] with many statistical applications.

2.7.2 Bias and variance Calculation

Locally polynomial estimators arise from a least squares problem see Blondin (2004) It is preferable to adopt a matrix notation in this context. Let X_x be the matrix associated with our experimental device:

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

We put

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

denotes the transposition, for a vector or a matrix. We now assume the invertibility of the square matrix, we denote by W_x the diagonal matrix $n \times n$ of weight:

$$W_x = W = diag\left(K\left(\frac{X_i - x}{h_n}\right)\right)$$

The least squares problem (2.57) can be summarized as follows:

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

where the sign^T denotes the transposition, for a vector or a matrix we now assume the invertibility of the matrix $X^T W X \in \mathcal{M}_{p+1}(\mathbb{R})$.

Remark 8. *More generally, if the matrix $X^T W X \in \mathcal{M}_{p+1}(\mathbb{R})$ is definite positive the estimator [LP](p) belongs to the class of linear estimators (LL) below*

According to least squares theory, the solution vector is given by

$$\hat{\beta} = (X^T W X)^{-1} X^T W y \tag{2.60}$$

This last equality (2.60) allows you to easily formulate the conditional bias and variance of the estimator $\hat{\beta}$
We recall the definition of the vector β

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

according to (2.56) we have $\mathbb{X}$ the set of variables $X_i, 1 \leq i \leq n$.we define $m = (m(X_1, ..., m(X_n))^T$ and $r = m - X\beta$ the residual vector. It follows, according to (2.60)

$$\begin{aligned} \mathbb{E}\left(\hat{\beta} \mid \mathbb{X}\right) &= (X^T W X)^{-1} X^T W m \\ &= \beta + (X^T W X)^{-1} X^T W r \end{aligned} \tag{2.61}$$

we have

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

where $\sigma^2(X_i) = Var\left(\hat{\beta} \mid \mathbb{X}\right)$ The conditional variance-covariance matrix is

$$Var\left(\hat{\beta} \mid \mathbb{X}\right) = (X^T W X)^{-1}(X^T \Sigma X)(X^T W X)^{-1} \tag{2.62}$$

the expressions (2.61) and (2.62) are not directly usable, because they depend on unknown quantities: the vector of residuals r and the matrix Σ Ruppert and Wand (1994) obtained asymptotic expansions for the bias and variance of the locally polynomial estimator $\hat{m}_n^{(k)}(x;p)$ defined in (2.58).
Before stating the theorem, we recall some useful notations. The moments of K and K^2 are denoted by $[\mu_j(K)] = \int_{\mathbb{R}} u^j K(u)du$ and $[\mu_j(K^2)] = \int_{\mathbb{R}} u^j K^2(u)du$ with $j \in \mathbb{N}$

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

We designate by $e_{k+1} = (0, ..., 0, 1, 0, ..., 0)^T$ the (k+1)-th vector unit in $\mathbb{R}^{p+1}$.

Theorem 2.7.1 (Ruppert and Wand (1994)). *the reader can see also Blondin (2004) P-30 We suppose $f_X(x) > 0$ and the functions $f_X(.)$, $m^{p+1}(.)$ and $\sigma^2(.)$ continue in a* *neighborhood of point x. Window h satisfies $h \to 0$ and $nh \to \infty$. So we get*

$$Var\left(\hat{m}_n^{(k)}(x;p) \mid \mathbb{X}\right) = (k!)^2 \times e_{k+1}^T S^{-1}\bar{S}S^{-1}\frac{\sigma^2(x)}{nh^{1+2k}f_X(x)} + o_{\mathbb{P}}\left(\frac{1}{nh^{1+2k}}\right) \quad (2.63)$$

When p-k is odd,

$$Biais\left(\hat{m}_n^{(k)}(x;p) \mid \mathbb{X}\right) = (k!)^2 \times e_{k+1}^T S^{-1}\frac{c_p}{(p+1)!}m^{(p+1)}(x)h^{p+1-k} + o_{\mathbb{P}}\left(h^{p+1-k}\right) \quad (2.64)$$

When p-k is even, assuming $f'_X(x) > 0$ and $m^{(p+2)}$ continues in a neighborhood of point x as well as $nh^3 \to \infty$. the asymptotic conditional bias is given by,

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

From the above theorem, it is clear that there is a difference between the case $p-k$ even and the odd $p-k$ case. When $p-k$ even, the principal bias term in $\mathcal{O}(h^{p+1})$ vanishes via the kernel symmetry K. On the other hand, when p-k odd, the asymptotic bias term has a simple expression where there are no derivative terms such as $f'_X(x)$. We notice that when $p = k = 0$, we do indeed find the asymptotic bias of the estimator [NW]. Of a From a practical and theoretical point of view, we will favor the odd p-k case (cf. the section 3.3 of Fan and Gijbels (1996)), where the form of the bias is more appreciable from a theoretical point of view. The best representation of the estimators [LP] is obtained by the method of "kernels equivalent ", that is to say by rewriting asymptotically the estimators [LP] under a more classical form close to the estimator [NW]. We introduce the following notation

$$S_{n,j} = \sum_{i=1}^{n}(X_i - x)^j K\left(\frac{X_i - x}{h_n}\right) \quad (2.65)$$

we have $S_n = X^T W X$ the square matrix of dimension p+1 also defined by,

$$S_n = \{S_{n,j+1}\}_{0\leq j,l\leq p}$$

From (2.60)

$$\begin{aligned}\hat{\beta}_k = e_{k+1}^T\hat{\beta} &= e_{k+1}^T S_n^{-1}X^T W y \\ &= \sum_{i=1}^{n} W_k^n\left(\frac{X_i - x}{h_n}\right)Y_i \end{aligned} \quad (2.66)$$

We notice that

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

It follows

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

We finally get,

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

where

$$W_k^n(t) = e_{k+1}^T S_n^{-1}\{1, th, ..., (th)^p\}^T K(t) \tag{2.67}$$

The estimator $\hat{\beta}_k$ therefore has a conventional form, except that the kernel W_k^n depends on the X_i points and their location. This intuitively explains why the locally polynomial estimate adapts to different experimental devices as well as to the across the density support. We now state a fundamental property estimators [LP](p).

Lemma 2.7.2. *The weight function $W_k^n(.)$satisfies the following condition:*

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

Above $\delta_{k,q}$ denotes the Kronecker symbol.

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

As a consequence of Lemma (2.7.2), the finite distance bias of the estimator $\hat{\beta}_k$ is zero when the function $m^{(k)}(.)$ to be estimated is a polynomial of degree less than or equal to p. This property highlights one of the practical advantages of estimation by the local polynomials for bias reduction, compared to the use of kernels high orders. Indeed, the bias is zero at fixed n and not asymptotically. In others terms, the estimator [LP](p) has the property of reproducing polynomials of degree $q \leq p$ (cf. proposition 1.12, p.32, (Tsybakov (2003))).

We continue the investigation of the properties of the weight function W_k^n We note that, when $h \to 0$ and$nh \to \infty$

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

via an application of Bochner's lemma and the law of large numbers. It follows

$$S_n = nf_X(x)HSH\left(1 + o_{\mathbb{P}}(1)\right) \tag{2.68}$$

where $H = diag1, h, ..., h^p$ By substituting the formula (2.68)in the definition (2.67) of $W_k^n(.)$ we get

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

It follows,

$$\hat{\beta}_k = \frac{1}{nh^{k+1}f_X(x)} \sum_{i=1}^{n} Y_i K_k^* \left(\frac{X_i - x}{h_n} \right) \left(1 + o_{\mathbb{P}}(1)\right) \tag{2.69}$$

with

$$K_k^*(t) = e_{k+1}^T S^{-1} 1, h, ..., h^p K(t) \tag{2.70}$$

The kernel in (2.70) is called the "equivalent kernel" and is very useful for express the asymptotic properties of the estimator [LP](p). The kernel (2.70) checks for following moment conditions

$$\int_{\mathbb{R}} u^q K_k^*(u) = \delta_{k,q} \qquad 0 \leq k, q \leq p \tag{2.71}$$

The kernel equivalent $K_k^*(u)$ is therefore simply a kernel of order (k, p + 1) We denote it accordingly $K_{k,p}^*(u)$ in order to underline the dependence in p. The conditional variance and bias of the estimator $\hat{m}_n^{(k)}(x;p)$ specified in (2.63) and (2.64) respectively, can be expressed as a function of the equivalent kernel $K_{k,p}^*(.)$,leading us to the following asymptotic expressions:

$$Var\left(\hat{m}_n^{(k)}(x;p) \mid \mathbb{X}\right) = \left(\frac{1}{nh^{1+2k}} \times \frac{\sigma^2(x)}{f_X(x)} \left((k!)^2 \int_{\mathbb{R}} K_{k,p}^*(u)du \right) \right) \left(1 + o_{\mathbb{P}}(1)\right) \tag{2.72}$$

and

$$Biais\left(\hat{m}_n^{(k)}(x;p) \mid \mathbb{X}\right) = \left(h^{p+1-k} \times \frac{m^{(p+1)}(x)}{(p+1)!} \left(k! \int_{\mathbb{R}} u^{p+1} K_{k,p}^*(u)du \right) \right) \left(1 + o_{\mathbb{P}}(1)\right) \tag{2.73}$$

These asymptotic expansions are easily obtained by relying on the formulas (2.69) and (2.71). The optimal window within the meaning of the local AMSE minimization criterion is obtained at from (2.73) and (2.72)

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

where

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

From these formulas, there are different procedures to choose the optimal window to from the data are based on the minimization of the conditional MISE, defined by,

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

The MISE-optimal window therefore has an asymptotic expression, from (2.72) and (2.73),

$$h^{MISE}(x) = C_{0,p}(K)\left(\frac{\int_{\mathbb{R}} \sigma^2(x)dx}{\int_{\mathbb{R}} f_X(x)(m^{(p+1)}(x))^2}\right)^{1/(2p+3)} n^{-1/(2p+3)} \tag{2.74}$$

Remember that plug-in type strategies are based on the replacement in (2.74) integrals unknown by consistent estimators.

Finally, for a comparative state of the art of the various techniques for estimating the regression function by the kernel method, we cite(Chu and Marron (1991)) and (Hastie and Loader (1993)).

CHAPTER 3

BASIC TOOLS OF EMPIRICAL PROCESS AND FUNCTIONAL VARIABLES

3.1 Basic Tools of Empirical Process

The theory of empirical processes is very useful because many statistics can be expressed as functionals of the empirical distribution function noted F_n.
Let$\{X_i : i \geq 1\}$ a sequence of random variables i.i.d with values on $\mathcal{X} = \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$More precisely, we can see the variable X_i as an application such as $X_i : \Omega \to \mathcal{X}$ for each $i \geq 1$. The function of empirical distribution based on $X_1, ..., X_n$ is defined by

$$F_n(t) = \frac{1}{n}\sharp\{X_i \leq t : 1 \leq i \leq n\} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{X_i \leq t\right\}, \quad t \in \mathbb{R}^p$$

To insist on the fact that the function F_n is random, i.e. dependent on $\omega \in \Omega$, we can use the following script:

$$F_n(t,\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{X_i(\omega) \leq t\right\}, \quad t \in \mathbb{R}^p$$

The theory on the empirical distribution function was primarily developed for p=1, i.e. for real random variables. Reference is made to the article by Gänssler and Stute (1979) and to the book by Shorack and Wellner (1986) for a complete exposition of the properties of F_n in the univariate framework. We notice, first of all, that F_n is the distribution function associated with the empirical measurement of the n-sample $X_i : 1 \leq i \leq n$, defined by

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where δ_{X_i} denote the Dirac measure at point $x \in \mathbb{R}$. When we look at F_n as a discrete random measure, it follows, for a given score function φ,

$$\int_{\mathbb{R}^p} \varphi(t) F_n(dt) = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i)$$

Thus, for φ integrable, we clearly obtain,

$$\int_{\mathbb{R}^p} \varphi(t) F_n(dt) \to \int_{\mathbb{R}^p} \varphi(t) F(dt) = \mathbb{E}\left(\varphi(Y)\right) \quad \textit{almost surely}$$

Example 2. *For the particular choice of*$\varphi(.) = h^{-p} K(\frac{X-.}{h})$ *we find the multivariate kernel estimator [PR] of the density*

$$\int_{\mathbb{R}^p} \varphi(t) F_n(dt) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

The empirical process is defined by:

$$\nu_n(t) = \sqrt{n}\left(F_n(t) - F(t)\right) \tag{3.1}$$

3.2 Empirical process indexed by sets

Let $U_1, U_2, ...$a sequence of independent and uniformly distributed random vectors in $[0,1]^d$. Let $\mathscr{C}$ the class Borelian on $[0,1]^d$. and let $\mathcal{C}$ any subclass of $\mathscr{C}$ We then introduce the uniform empirical process indexed by the set $\mathcal{C}$

$$\nu_n(C) = \sqrt{n}\left(\lambda_n(C) - \lambda(C)\right) \quad , C \in \mathcal{C}$$

where λ denotes the Lebesgue measure on $\mathbb{R}^d$ and λ_n is the uniform empirical measure indexed by $\mathscr{C}$ such as

$$\lambda_n(C) = n^{-1} \sum_{i=1}^{n} \mathbb{1}\{U_i \in C\} \quad C \in \mathscr{C},$$

3.3 weak convergence

Definition 3.3.1. *Let* $(\mathbb{D}, d)$ *be a metric space, and let* $\mathbb{P}_n$ *and* $\mathbb{P}$ *be Borel probability measures on* $(\mathbb{D}, \mathscr{D})$*, where* $\mathscr{D}$ *is the Borel* σ*-field on* $\mathscr{D}$ *, the smallest* σ*-field containing all the open sets. Then the sequence* $\mathbb{P}_n$ *converges weakly to* P*, if and only if:*

$$\int_{\mathbb{D}} f d\mathbb{P}_n \longrightarrow \int_{\mathbb{D}} f d\mathbb{P}$$

for all $f \in C_b(\mathbb{D})$ *Here* $C_b(\mathbb{D})$ *denotes the set of all bounded, continuous, real functions on* $\mathbb{D}$ *.*

Theorem 3.3.2. *Glivenko (1933) and Cantelli (1933)*
If $X_1, X_2, ...$ *are i.i.d random variables with distribution* F *then:* $\|F_n - F\|_\infty \to 0$ *almost surly.*

Theorem 3.3.3. *Donsker (1952)*
If $X_1, X_2, ...$ are i.i.d random variables with distribution F then: $\sqrt{n}\,(F_n - F)$ converge in distribution to standard Brownian bridge $\mathbb{G}$.[1]

3.4 Glivenko-Cantelli and Donsker classes

It is necessary to impose conditions on our functions classes if we wish to obtain uniform convergence results. Here we being interested in at find the necessary conditions why a functions class is either Glivenko-Cantelli or Donsker
For any probability measure ν and for $p > 0$, $f \in \mathscr{F}$ we note $\|f\|_{p,\nu} = \int |f(\omega)|^p d\nu(\omega)$ the norm $L^P(\nu)$ and we recall also the function Φ is the envelope for the functions class if $|f(\omega)| \leq \Phi(\omega)$ almost surly for all element $f \in \mathscr{F}$.
We use the condition in the brackets the class function will be Glivenko-Cantelli if

$$N_{[\ ]}(\varepsilon, \mathscr{F}, L_1(\nu)) < \infty.$$

With the covering numbers 3.5.6 the condition sufficient so that $\mathscr{F}$ Glivenko-Cantelli is

$$\sup_{\nu:\|\Phi\|_{\nu,1}<\infty} N(\varepsilon\|\Phi\|_{\nu,1} < \infty, \mathscr{F}, L_1(\nu)) < \infty.$$

In the same way the condition sufficient so that $\mathscr{F}$ Donsker is

$$\int_0^{\infty} \sqrt{\log N_{[\ ]}(\varepsilon, \mathscr{F}, L_1(\nu))}d\varepsilon < \infty.$$

By the way so that $\mathscr{F}$ Donsker is

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

Vapnik Červonenkis classes

Definition 3.4.1. *Let $\mathcal{C}$ be a subset of $\mathcal{X}$ and $\{x_1, ..., x_n\}$ be n points in the same space. We say $C \in \mathcal{C}$* picks out *$Y \subseteq \{x_1, ..., x_n\}$ if $C \cap \{x_1, ..., x_n\} = Y$ We say that $\mathcal{C}$* shatters *points $\{x_1, ..., x_n\}$ if every subset of $\{x_1, ..., x_n\}$ is picked out by some set $C \in \mathcal{C}$:*

$$\{\{x_1, ..., x_n\} \cap C : C \in \mathcal{C}\} = 2^{\{x_1,...,x_n\}}.$$

*If exists some fnite n such that C shatters no set of size n, then we say $\mathcal{C}$ is a VC-class of sets.The smallest such n is called the VC-*index *of $\mathcal{C}$, denoted by V(C).*

The interest of these classes in the following result it gives us an exponential boundary of the covering number see(Van der Vaart and Wellner (1996)) for the VC-classes of functions

[1] A standard Brownian bridge is a continuous stochastic process $(\mathbb{G}(t) : 0 \leq t \leq 1)$ such that $\mathbb{G}(t) = B(t) - tB(1)$ where B is a standard Brownian motion.

$\mathscr{F}$ an envelope Φ we have for all probability measures ν any $p \geq 0$

$$\sup_{\nu:\|\Phi\|_{\nu,p}<\infty} N(\varepsilon\|\Phi\|_{\nu,1} < \infty, \mathscr{F}, L_p(\nu)) < KV(\mathscr{F})(16e)^{V(\mathscr{F})}\varepsilon^{-p(V(\mathscr{F}))-1},$$

where K is constant universal and $0 < \varepsilon < 1$.
So if Φ integrable the VC-class will be also Glivenko-Cantelli class while if Φ integrable square VC-class will be Donsker moreover the VC function classes has many very useful stability properties, we return the reader to Van der Vaart and Wellner (1996) for more details on VC-class and entropy conditions.

3.5 Basic Tools of functional variables

In this section we give some principal definitions and tools that we used to solve our problems related by functional variables.

3.5.1 Introduction

There is a close connection to traditional statistics where usually the properties of a random variable or a multivariate random variable is studied. During the last three decades Functional data analysis (FDA) became a popular tool to handle data entities that are random functions. Usually, discrete and noisy versions of them are observed. Oftentimes, these entities are highdimensional spatial objects,It should also be mentioned that the progress made in the functional data processes have made it possible to offer the possibility to statisticians to have more and more often observations of so called functional variables, i.e. curves These data are modeled as being realizations of a random variable taking its values abstract space or possibly infinite dimension.

3.5.2 Functional variables

Nonparametric statistics have been developed intensively where the variable takes values in an infinite dimensional space which is called functional variable has received increasing interests in recent literature this type of variable is found in different fields, such meteorology, quantitative chemistry, biometrics, econometrics, nonparametric statistics in the functional context took a lot of attention currently in fact hundreds of Books and papers have been published in this framework last decade.

Definition 3.5.1. *A random variable X is called functional variable (f.v.) if it takes values in an infinite dimensional space (or functional space) An observation x of X is called functional data.*

3.5.3 Nonparametric Statistics for Functional Data

Traditional statistical methods fail as soon as we deal with functional data. Indeed, if for instance we consider a sample of finely discretized curves, two crucial statistical problems appear. The first comes from the ratio between the size of the sample and the number of variables (each real variable corresponding to one discretized point). The second, is due to the existence of strong correlations between the variables and becomes an ill-conditioned problem in the context of multivariate linear model. So, there is a real necessity to develop statistical methods/models in order to take into account the functional structure of this kind of data. Most of existing statistical methods dealing with functional data use linear modelling for the object to be estimated. Key references on methodological aspects are those by Ramsay et Silverman (1997) and Ramsay and Silverman (2005a) , while applied issues are discussed by Ramsay et Silverman (2002) and implementations are provided by Clarkson*et al* (2005). Note also that, for some more specific problem, some theoretical support can be found in Bosq (2000). On the other hand, nonparametric statistics have been developped intensively. Indeed, since the beginning of the sixties, a lot of attention has been paid to free-modelling (both in a free-distribution and in a free-parameter meaning) statistical models and/or methods. The functional feature of these methods comes from the nature of the object to be estimated (such as for instance a density function, a regression function, ...) which is not assumed to be parametrizable by a finite number of real quantities. In this setting, one is usually speaking of **Nonparametric Statistics** for which there is an abundant literature. For instance, the reader will find in Härdle (1990) a previous monograph for applied nonparametric regression, while Schimek (2000) and Akritas and Politis (2003) present the state of the art in these fields. It appears clearly that these techniques concern only classical framework, namely real or multidimensional data. However, recent advances are mixing nonparametric free-modelling ideas with functional data throughout a double infinite dimensional framework see Ferraty and Vieu (2003), Demongeot *et al.* (2017),Naceri (2016),Lian (2011), Yang*et al.* (2014) among other.

Definition 3.5.2. *Let $\mathcal{Z}$ a random variable valued in some infinite dimensional space F and let ϕ be a mapping defined on F and depending on the distribution of $\mathcal{Z}$. A model for the estimation of ϕ consist in introducing some constraint of the from*

$$\phi \in \mathcal{C}$$

the model is called a functional parametric model for estimation of ϕ if $\mathcal{C}$ is indexed by finite number of elements of F.Otherwise, the model is called a functional nonparametric model.

3.5.4 Functional data

One never observes an integral function over its entire trajectory. This would require a measuring instrument with an in- ternal recording speed. Even the fastest quotes on the fully computerized financial markets are spanned by a few milliseconds. When the functional data arrive they are for these reasons always in vector form. Thus we shall not observe, for example $X(t)$ ∀ t but we shall have $[X(t_1), X(t_2), ..., X(t_p)]$ where the t_j constitute a discretization grid. From the phenomenon studied p can vary between several units and several million. This type of data is not new and has been studied for a long

time using multivariate techniques (seeing X as a random vector in $\mathbb{R}^p$ to continue the previous example). But there are two problems.
-If the frequency of discretization of the curves is high (i.e. if p is large) we can find ourselves in situations where the size of X is of the order or even greater than the size of the sample itself. This situation can pose prohibitive problems both from the theoretical point of view and from the numerical aspects. This problem is common with that of many problems of statistics in large dimensions.
-By treating X as a vector, we completely lose its true nature, that of process in continuous time or more generally of function. The derivation operation, for example, does not make sense in this context. It is logical then to ask the question of alternative methods in which, by failing to grasp X (t) $\forall$ t, one could be satisfied with an approximation $\tilde{X}$ which would be a real function.

Some examples of functional data

The statistic for functional data or functional data analysis studies observations which are not real or vector variables but random curves.
Examples:

- The temperature curve recorded at a given point on the globe is a completely random continuous process. If the temperature is observed during N days it may be interesting to cut out the starting curve on N curves which plot the temperature for each of the observation days. Each of these daily curves can then be seen as an element of a sample of size N constituted of functional data

- Currently, experiments are carried out on the INRA campus to study the growth of maize plants from different varieties and subjected to explicit conditions, Different errors. For each maize plant the measuring instruments collect a function which is indeed random (it depends on the varietal of maize, experimental conditions and other Fluctuations ...)

In the two preceding examples the random curves depend on time but the situation may be different. The spectrometric analysis of the materials (which aims to deduce physicochemical properties by examining a light spectrum from the material).Also produces random curves indexed by a wavelength (and more by time).

Functional Datasets

Since the middle of the nineties, the increasing number of situations when functional variables can be observed has motivated different statistical developments, that we could quickly name as *Statistics for Functional Variables (or Data)*. We are determinedly part of this statistical area since we will propose several methods involving statistical functional sample $X_1, ..., X_n$ Let us start with a precise definition of a functional dataset.

Definition 3.5.3. *A functional dataset* $\chi_1, ..., \chi_n$ *is the observation of n functional variables* $\mathcal{X}_1, ..., \mathcal{X}_n$ *identically distributed as* $\mathcal{X}$.

This definition covers many situations, the most popular being curves datasets. We will not investigate the question of how these functional data have been collected, which is linked with the discretization problems. According to the kind of the data, a preliminary stage consists in presenting them in a way which is well adapted to functional processing. If the grid of the measurements is fine enough, this first important stage involves usual numerical approximation techniques (see for instance the case of spectrometric data presented in Ferraty and Vieu (2006) . In other standard cases, classical smoothing methods can be invoked (see for instance the phonemes data and the electrical consumption curves discussed inFerraty and Vieu (2006). There exist some other situations which need more sophisticated smoothing techniques, for instance when the repeated measures per subjects are very few (sparse data) and/or with irregular grid. This is obviously a parallel and complementary field of research but this is far from our main purpose which is nonparametric statistical treatments of functional data. From now on, we will assume that we have at hand a sample of functional data.

Some Fields for application of functional data:

In terms of applications (in imagery, agro-industry, geology, econometrics, ...), We present below some concrete examples:

Study of the El N*iño* phenomenon

It is a data set from the study of a Quite important climatological phenomenon. This phenomenon is commonly called El N*iño* It is a great sea current which occurs in an exceptional way.(on average one Twice a decade) along the Peruvian coast at the end of winter. This current causes global climatic disruption. The data set is consisting of monthly ocean surface temperature readings 1950 in an area in the north of Peru (co-ordinates $0 - 10°$ South, $80 - 90°$ West) in which the El N*iño* sea current can appear. These data and their are available on the web site of the American climate forecast center: http: //www.cpc.ncep.noaa.gov/data/indices/. It should be noted that the evolution of temperatures over time is actually a continuous phenomenon. The number of measurements allows to take into consideration the functional nature of the data (cf. Figure 3.1). From these data, we can be interested in the prediction of the evolution of the phenomenon from data collected in previous years.

In animal Biology :

studies on the laying of Mediterranean flies have been carried out and summarized by curves giving, for each fly, the quantity of eggs laid as a function of time (cf. Figure 3.2)

In Biology:

for the study of variations in growth curves (cf. Rao, 1958 and Figure 3.3), and more recently, for the study of variations in the angle of the knee during walking (cf.Ramsay et Silverman (2002)). Note that a huge amount of functional data is product and asks only to have the adequate methodology for its treatment, in particular mass spectrometric data (cf. for cancer Figure 3.4).

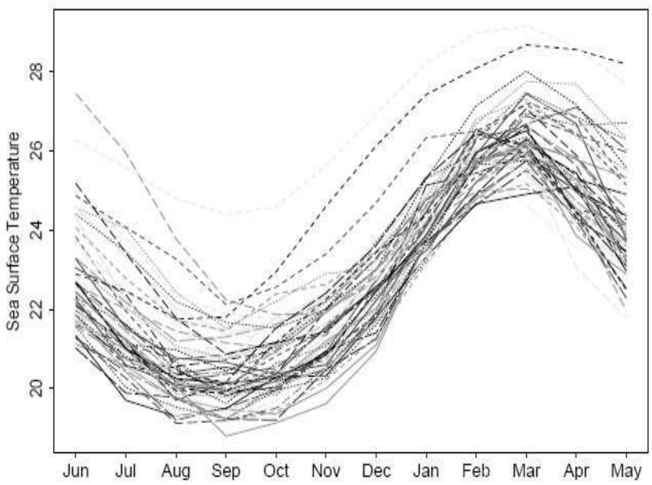

Figure 3.1: The curves corresponding to the current El N*iñ*o

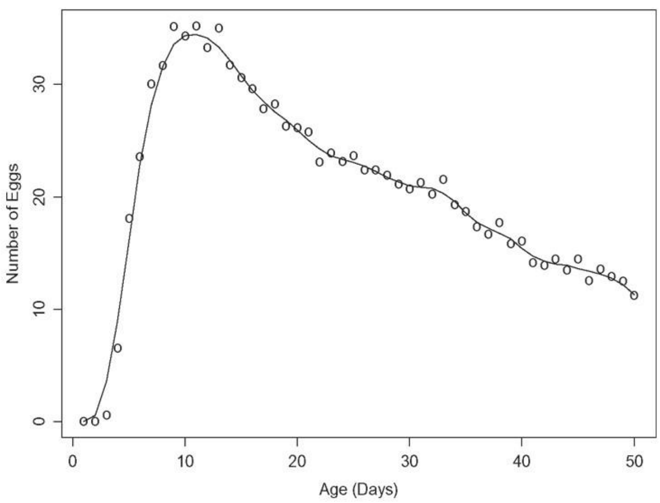

Figure 3.2: A curve of the number of daily eggs laid by flies

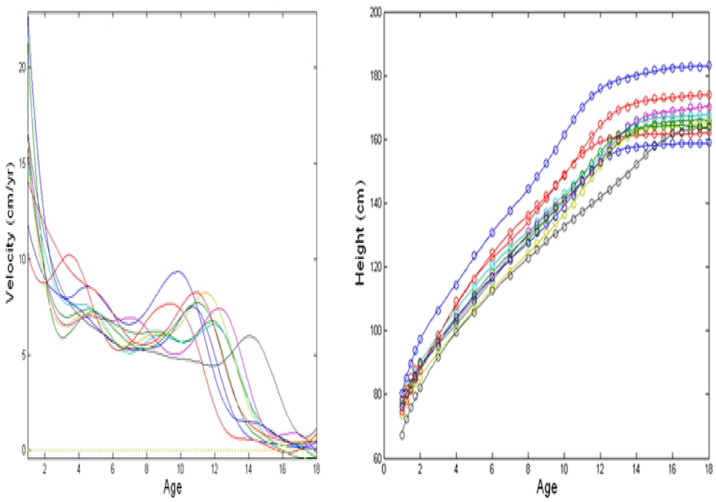

Figure 3.3: Curves growth

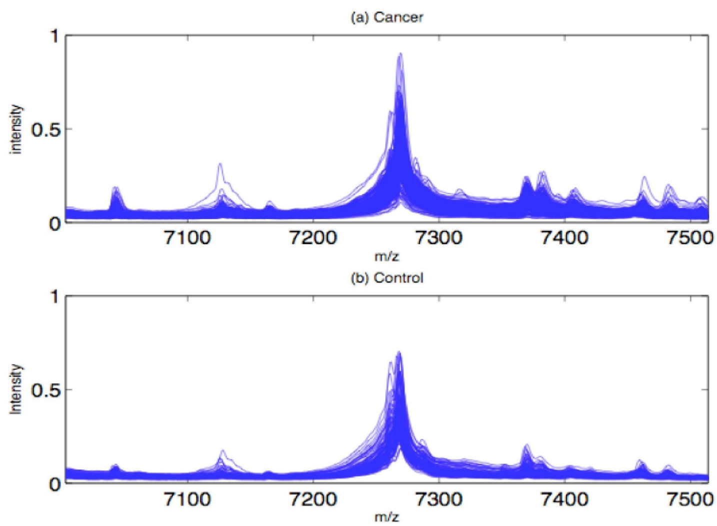

Figure 3.4: Mass spectrometric curves on cancer cells

3.5.5 Semi-metric space and Small Ball Probabilities

Semi-metric space

To study data it is often to have a notion of distance between them. Its well known that in finite dimension all metrics are equivalent this in no longer the case in infinite dimension, which is why the chose of the metric (and therefore of associated topology) is even more crucial for the study of functional random variables than it is in multivariate statistics.In addition to available metrics, it is often interesting to consider semi-metrics

Definition 3.5.4. *d is a semi-metric on some space F as soon as:*

1. $\forall x \in F, d(x, x) = 0.$
2. $\forall (x, y, z) \in F \times F \times F, d(x, y) \leq d(x, z) + d(z, y).$

Allowing a wider range of possible topologies that can be chosen depending on the nature of the data and the problem under consideration. this is why we have chosen in thesis to consider and study functional variable defined as random variables with values in semi-metric space $(\mathcal{E}, d)$ of infinite dimension. A part from allowing the modeling of more general phenomena, another interest of using a semi-metric rather than a metric is that it can constitute an alternative to the problems posed by large data dimension.Indeed, We can take a semi-metric defined from a projection of our functional data in space of smaller size than by performing in functional principal component analysis of our data (Besse and Ramsay (1986),Yao and Lee (2006)) or by projecting them on a finite basis (wavelets,spline ....)this reduce the size of the data and increase the speed of convergence of the methods used while preserving the functional nature of the data .We can choose the basis on which we project based on the knowledge we have of nature of functional variable. For example wa can choose the fourier basis if we assume that the functional variable observed is periodic we can refer to Ramsay et Silverman (1997), Ramsay and Silverman (2005a) for more complete discussion of the value of using different approximation methods by projection of functional data. Further discussion of the value of using different types of semi-metric is made in Ferraty and Vieu (2006) especially in Section (3.4).It can be remember that the choice of the semi-metric makes it possible both to take account of more varied situation and to be able to circumvent the scourge of the dimension.This choice,however;should not be made lightly but taking into account the nature of the data the and problem under study .

Small Ball Probabilities

The curse of dimensionality is well known phenomenon in nonparametric regression on multivariate variable see Stone (1982), in multivariate nonparametric regression,convergence rates (for the dispersion part) are expressed in terms of h_n^d.In the functional case we adopt more general concentration notations called small probabilities and express our asymptotic results in function of these quantities, small ball probabilities are defined by:

$$\phi(h_n) = \mathbb{P}(X \in B(x, h)) \tag{3.2}$$

the way they decrease to zero have a great influence on the convergence rate of the kernel estimator. In the case of finite dimension spaces, that is $\mathcal{E} = \mathbb{R}^d$ it can be seen that $\phi(h_n) = C(d)h^d l(x) + o(h^d)$,where $C(d)$is the volume of the unit ball in $\mathbb{R}^d$.Furthermore, in infinite dimensions there exist many examples fulfilling the decomposition above. we quote the following (which can be found in Ferraty *et al.* (2007)):

1. $\phi(h_n) = l(x)h^\upsilon$ for some $\upsilon > 0$ with $\tau_x(s) = s^\upsilon$;
2. $\phi(h_n) = l(x)h^\upsilon \exp(-Ch^{-p})$ for some $\upsilon > 0$ and $p > 0$ with $\tau_x(s)$ is the Dirac's function;
3. $\phi(h_n) = l(x) \left|\ln(h)\right|^{-1}$ with $\tau_x(s) = \mathbb{1}_{]0,1]}(s)$ the indicator function in $]0, 1]$.

This notion of small ball probabilities will play a major role both from theoretical and practical points of view. Because the notion of ball is strongly linked with the semi-metric d, the choice of this semi-metric will become an important stage. Note that the rates of convergence of our nonparametric functional estimates will be systematically linked with d through the behaviour, around 0, of the small ball probability function $\phi(h_n)$ .It exists in the literature a fairly large number of probabilistic results that study the way where these probabilities of small balls tend to 0 when d is a norm (see for example, Li and Shao (2001), Lifshits *et al.* (2006) and Gao et Li (2007)). We can also refer to the work of Dereich (2003) (Chapter 7) which is devoted to the behavior of probabilities of small balls whose centers are random. Through this work we can see, for example, that in the case of non-smooth processes such as motion Brownian or the Ornstein-Uhlenbeck process, these probabilities of small balls are exponential form (with respect to hn)and that consequently the speed of convergence of our estimators is in power of ln (n) (Ferraty *et al.* (2006), paragraph 5 and Ferraty and Vieu (2006a), paragraph 13.3.2, for further discussion on this topic) Moreover the choice of the semi-metric is expected to be a crucial point as long as we focus on the applied aspects.

3.5.6 Kernel functions

We will consider only two kinds of kernels for functional variables

Definition 3.5.5.

i) A function K from $\mathbb{R}$ into $\mathbb{R}^+$ such that $\int K = 1$ is called a kernel of type I if there exist two real constants $0 < \mathrm{C}_1 < \mathrm{C}_2 < \infty$ such that:

$$\mathrm{C}_1 \mathbb{1}_{[0,1]} \leq K \leq \mathrm{C}_2 \mathbb{1}_{[0,1]}.$$

ii)A function K from $\mathbb{R}$ into $\mathbb{R}^+$ such that $\int K = 1$ is called a kernel of type II if its support is $[0, 1]$ and if its derivative K'exists on $[0, 1]$ and satisfies for two real constants $-\infty < \mathrm{C}_2 < \mathrm{C}_1 < 0$:

$$C_2 \leq K' \leq C_1.$$

3.5.7 Topological considerations

Entropy

Definition 3.5.6. *Let $(\mathcal{H}, d)$ be a semimetric space then the ε-covering number $N(\varepsilon, \mathcal{H}, d)$is defined as the smallest number of balls of radius ε required to cover $\mathcal{H}$. The ε-entropy number is $H(\varepsilon, \mathcal{H}, d) = \log N(\varepsilon, \mathcal{H}, d)$.*

Definition 3.5.7. *Let $(\mathcal{H}, d)$ be a semimetric space.If l,$u \in \mathcal{H}$,then the bracket $[l, u] = \{f \in \mathcal{H} : l(x) \leq f(x) \leq u(x) \forall x\}$defines a subset of functions squeezed pointwise between l and u. We call $d(l, u)$ the size of the bracket.The ε-bracketing number $N_B(\varepsilon, \mathcal{H}, d)$ is defined as the smallest number of brackets of size at most ε required to cover $\mathcal{H}$ i.e. $N_B(\varepsilon, \mathcal{H}, d) = \inf \left\{ n : \exists l_1, u_1, ..., l_n, u_n \quad s.t. \quad \bigcup_{i=1}^{n} [l_i, u_i] = \mathcal{H} \quad and \quad d(l_n, u_n) \leq \varepsilon \right\}$ The ε-bracketing entropy number is $H_B(\varepsilon, \mathcal{H}, d) = \log N_B(\varepsilon, \mathcal{H}, d)$.*

The purpose of this section is to emphasize the topological components of our study. Indeed,as indicated in Ferraty and Vieu (2006), all the asymptotic results in nonparametric statistics for functional variables are closely related to the concentration properties of the probability measure of the functional variable X. Here,we have moreover to take into account the uniformity aspect. To this end,let $\mathcal{S}_{\mathcal{F}}$ be a fixed subset of $\mathcal{H}$; we consider the following assumption:

$$\forall x \in \mathcal{S}_{\mathcal{F}}, 0 < C\mathcal{F}_x(h) \leq \mathbb{P}(X \in B(x, h)) \leq C^{'}\mathcal{F}_x(h) < \infty$$

We can say that the first contribution of the topological structure of the functional space an beviewed through the function $\mathcal{F}_x$ controlling the concentration of the measure of probability of the functional variable on a small ball. Moreover, for the uniform consistency, where the main tool is to cover a subset $\mathcal{S}_{\mathcal{F}}$ with a finite number of balls, one introduces an other topological concept defined as follows:

Some examples

Kolmogorov's entropy

Definition 3.5.8. *Let $\mathcal{S}_{\mathcal{F}}$ be a subset of a semi-metric space $\mathcal{H}$, and let $\varepsilon > 0$ be given. A finite set of points $x_1, x_2, ..., x_N$ in $\mathcal{F}$ is called an ε-net for $\mathcal{S}_{\mathcal{F}}$ if $\mathcal{S}_{\mathcal{F}} \subset \bigcup_{k=1}^{N} B(x_k, \varepsilon)$*
The quantity $\psi_{\mathcal{S}_{\mathcal{F}}} = log(N_\varepsilon(\mathcal{S}_{\mathcal{F}}))$, where $N_\varepsilon(\mathcal{S}_{\mathcal{F}})$ is the minimal number of open balls in $\mathcal{F}$ of radius ε wich is necessary to cover $\mathcal{S}_{\mathcal{F}}$,is called the Kolmogorov's ε-entropy of the set $\mathcal{S}_{\mathcal{F}}$

This concept was introduced by Kolmogorov in the mid-1950's see (Kolmogorov and Tikhomirov (1959)) and it represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy ε. Therefore, the choice of the topological structure (with other words, the choice of the semi-metric) will play a crucial role when one is looking at uniform (over some subset $\mathcal{S}_{\mathcal{F}}$) of $\mathcal{F}$ asymptotic results. More precisely, we will see thereafter that a good semi-metric can increase the concentration of the probability measure of the functional variable X as well as minimize the ε-entropy of the subset $\mathcal{S}_{\mathcal{F}}$. In an earlier contribution (see, Ferraty *et al.* (2006)) we highlighted the phenomenon of concentration of the probability measure of the functional variable by computing the small ball probabilities in various standard situations. We will devote Section 3.5.7 to discuss the behaviour of the Kolmogorov's ε-entropy in these standard situations. Finally, we invite the readers interested in these two concepts (entropy and small ball probabilities) or/and the use of the Kolmogorov's ε-entropy in dimensionality reduction problems to refer to respectively, Kuelbs and Li (1993) or/and Theodoros and Yannis (1997). We will start Example 3 by recalling how this notion behaves in unfunctional case (that is when $\mathcal{F} = \mathbb{R}^P$). More interestingly (from statistical point of view) is Example 4 since it allows to construct, in any case, a semi-metric with reasonably "small" entropy.

Example 3. *(Compact subset in finite dimensional space) :*
A standard theorem of topology guaranties that for each compact subset $\mathcal{S}_{\mathcal{F}}$ of $\mathbb{R}^P$ and for each $\varepsilon > 0$ there is a finite ε-net and we have for any $\varepsilon > 0$,

$$\psi_{\mathcal{S}_{\mathcal{F}}}(\varepsilon) \leq \mathrm{C}_p \, log(1/\varepsilon).$$

More precisely, Chate and Courbage (1997) have shown that, for any $\varepsilon > 0$ the regular polyhedron in $\mathbb{R}^P$ with length r can be covered by $([2r\sqrt{p}/\varepsilon] + 1)$ balls, where [m] is the largest integer which is less than or equal to m. Thus, the Kolmogorov's ε-entropy of a polyhedron P_r in $\mathbb{R}^P$ with length r is

$$\forall \varepsilon > 0, \psi_{P_r}(\varepsilon) \sim p \, log([2r\sqrt{p}/\varepsilon] + 1).$$

Example 4. *(Compact subset in a Hilbert space with a projection semimetric):*
The projection-based semi-metrics are constructed in the following way. Assume that H is a separable Hilbert space, with inner product $< .,. >$ and with orthonormal basis $\{e_1, ..., e_j, ...\}$, and let k be a fixed integer, $k > 0$. As shown in Lemma (13.6) of Ferraty and Vieu (2006), a semi-metric d_k on $\mathcal{H}$ can be defined as follows

$$d_k(x, x') = \sqrt{\sum_{i=1}^{k} < x - x', e_j >^2}. \tag{3.3}$$

Let χ be the operator defined from $\mathcal{H}$ into $\mathbb{R}^k$ by

$$\chi(x) = (< x, e_1 >, ..., < x, e_k >),$$

and let d_{eucl} be the euclidian distance on $\mathbb{R}^k$, and let us denote by $\mathbf{B}_{eucl}(.,\, .)$ an open ball of $\mathbb{R}^k$ for the associated topology. Similarly, let us note by $\mathbf{B}_k(.,.)$ an open ball of $\mathcal{H}$ for *the semi-metric d_k. Because is a continuous map from $(\mathcal{H}, d_k)$ into $(\mathbb{R}^k, d_{eucl})$, we have that for any compact subset $\mathcal{S}$ of $(\mathcal{H}, d_k)$, $\chi(\mathcal{S})$ is a compact subset of $\mathbb{R}^k$. Therefore, for each $\varepsilon > 0$ we can cover $\chi(\mathcal{S})$ with balls of centers $z_i \in \mathbb{R}^k$:*

$$\chi(\mathcal{S}) \subset \bigcup_{i=1}^{d} \mathbf{B}_{eucl}(z_i, r), \quad with \quad dr^k = \mathbf{C} \quad for\,some \quad \mathbf{C} > 0 \tag{3.4}$$

For $i = 1, ..., d$, let x_i be an element of $\mathcal{H}$ such that $\chi(x_i) = z_i$. The solution of the equation $\chi(x) = z_i$ is not unique in general, but just take x_i to be one of these solutions. Because of (3.3), we have that

$$M\chi^{-1}(\mathbf{B}_{eucl}(z_i, r)) = \mathbf{B}_k(x_i, r) \tag{3.5}$$

Finally, (3.4) and (3.5) are enough to show that the Kolmogorov's ε-entropy of $\mathcal{S}$ is

$$\psi\mathcal{S}(\varepsilon) \approx \mathbf{C}k\,log\left(\frac{1}{\varepsilon}\right).$$

3.5.8 Some Models Conditional In Non-parametric Statistics Functional

Let (X, Y) a copies $(X_i, Y_i)_{i=1,...,n}$ couples with the same law of random variables with values in $\mathcal{E} \times \mathbb{R}$, where $\mathcal{E}$ the functional space equipped by semi-metric $d_{\mathcal{E}}(.,.)$, defines the regression function of $\psi(Y)$ given $X = x$ by:

$$\mathbb{G}(x) = \mathbb{E}(\psi(Y) \mid X = x) \tag{3.6}$$

whereψ a known measurable function (see Ferraty and Vieu (2006)),
The Kernel estimator of the regression function $\mathbb{G}$ proposed by Nadaraya (1964) and Watson (1964) is :

$$\mathbb{G}_n(\psi(Y), x) = \frac{\sum_{i=1}^{n} \psi(Y) K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)},$$

Here K(.) is the kernel function values $\mathbb{R}$ and h_n is the smoothing parameter.
The first results in nonparametric functional statistics were developed by Ferraty and Vieu (2000)concern the estimation of the function regression with explanatory variable of fractal dimension. They established the almost complete convergence of a kernel estimator of this model non parametric in the i.i.d. case Drawing inspiration from recent developments of the theory of probability of small balls,Ferraty and Vieu (2004) have generalized these last results to the α-mixing case and they exploited the importance of non-parametric modeling of functional data in applying their study to curve discrimination and forecasting. In the framework of α-mixing functional observations, Masry (2005) showed the asymptotic normality of the estimator of Ferraty and Vieu (2004) for the regression function. Note that the function(3.6) can group several non-parametric models.

By choosing $\psi(Y) = \mathbb{1}_{\{Y\in C\}}$

$$\mathbb{G}(x) = \mathbb{E}\left(\mathbb{1}_{\{Y\in C\}} \mid X = x\right) = \mathbb{P}\left(\mathbb{1}_{\{Y\in C\}} \mid X = x\right). \tag{3.7}$$

The Nadaraya-Watson Estimator of The Conditional Distribution Function $\mathbb{G}(C,x)$ where C measurable set in the collection subsets $\mathscr{C}$ by:

$$\mathbb{G}_n(x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_i\in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{3.8}$$

where K(.) is a real-valued kernel function on $\mathbb{R}$ and h_n is a smoothing parameter The estimation of the conditional distribution function in a functional framework was introduced by Ferraty *et al.* (2006) . They constructed a double kernel estimator for the conditional distribution function and specified the almost complete convergence rate of this estimator when the observations are independent and identically distributed, The case of α-mixing observations was studied by Ferraty*et al.* (2005).
The Estimator of the Conditional Density Function f given by

$$\widehat{f_n}(x) = \frac{1}{n\phi(h_n)} \sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right) \tag{3.9}$$

where K(.) is a real-valued kernel function on $\mathbb{R}$ and h_n is a smoothing parameter. The estimation of the conditional density function and its derivatives, in functional statistics, was introduced by Ferraty *et al.* (2006). These authors have obtained almost complete convergence in the case i.i.d. Since this article, an abundant literature has developed on the estimation of conditional density and its derivatives, in particular to use it to estimate the conditional mode. Indeed, by considering α-mixing observations, Ferraty *et al.* (2005) established the almost complete convergence of a kernel estimator of the conditional mode defined by the maximizing random variable conditional density.

CHAPTER 4

SOME ASYMPTOTIC PROPERTIES OF THE CONDITIONAL SET-INDEXED EMPIRICAL PROCESS BASED ON DEPENDENT FUNCTIONAL DATA

This chapter is the subject of publication to appear in :International journal of mathematics and statistics.

The purpose of this chapter is to establish the invariance principle for the conditional set-indexed empirical process formed by strong mixing random variables when the covariates are functional. We establish our results under some assumptions on the richness of the index class $\mathscr{C}$ of sets in terms of metric entropy with bracketing. We apply our main result for testing the conditional independence, that is, testing whether two random vectors Y_1 and Y_2 are independent, given X. The theoretical results of the present paper are (or will be) key tools for many further developments in functional data analysis.

4.1 Introduction

The theory of empirical processes is one of the major continuing themes in the historical development of mathematical statistics and it has many applications ranging from parameter estimation to hypothesis testing, its history theory dates back to the 1930's and 1940's there has been a great deal research works. The asymptotic properties of empirical processes indexed by functions have been intensively studied during the past decades (see, e.g., Van der Vaart and Wellner (1996) or Dudley (1999) for self-contained, comprehensive books on the topic with various statistical applications). Vapnik and Červonenkis (1971) characterize, modulo measurability, the classes $\mathscr{C}$ of sets for which the Glivenko-Cantelli theorem holds, in the independent framework. In this setting many papers were published, we cite among many others Dudley (1978), Giné and Zinn (1984), Le Cam (1983), Pollard (1982) and Bass and Pyke (1984). Dudley (1978) studied the empirical process indexed by a class of measurable sets, that is, he considered $\mathscr{F} = \{\mathbb{1}_A(\cdot) : A \in \mathcal{A}\}$, where A is a suitable subset of the Borel σ-algebra. He obtained several very useful results that go far beyond Donsker's theorem, more precisely, he stated different assumptions under which weak convergence to a Gaussian process holds, including a so-called metric entropy with inclusion. Generalizing this idea, Ossiander (1987) introduced L_2-brackets to approximate the elements of $\mathscr{F}$. These brackets allow to study larger classes of functions as long as a metric entropy integrability condition is satisfied, see Ossiander (1987), Theorem 3.1. To deal with random variables such as time series that are dependent, one naturally asks whether results obtained under the independence assumption remain valid. However, a bracketing condition under strong mixing was stated by Andrews and Pollard (1994). Doukhan (1995) studied the function-indexed empirical process for β-mixing sequences. The case of Gaussian long-range dependent random vectors was already handled by Arcones (1994), Theorem 9. The assumption on the bracketing number therein is very restrictive and was considerably improved later. In this lines of research in different type of mixing, we may cite Eberlein (1984), Nobel and Dembo (1993) and Yu (1994). The extension of the above exploration to conditional empirical processes is practically useful and technically more challenging, we may refer to Stute (1986a), Stute (1986b), Horváth and Yandell (1988) for the case of independent observations, other authors were interested to the dependent case, for example Yoshihara (1990) established the asymptotic normality when the sequences are ϕ-mixing. Polonik and Yao (2002) have established uniform convergence and asymptotic normality of set-indexed conditional empirical process in a strictly stationary and strong mixing framework. The results of Polonik and Yao (2002) were extended by Poryvaĭ (2005). In the present paper, we are interested in the limiting behavior of he conditional set-indexed empirical process when the covariates are functional. Functional data analysis is a field that has been really popularized with the book by Ramsay and Silverman (2005a) and that received a lot of attention in the last 20 years with a general aim of adapting existing multivariate ideas to the functional framework. For good sources of references to research literature in this area along with statistical applications consult Ramsay and Silverman (2005a), Bosq (2000), Ramsay and Silverman (2005b), Ferraty and Vieu (2006), Bosq and Blanke (2007), Shi and Choi (2011), Horváth and Kokoszka (2012), Zhang (2014), Bongiorno *et al.* (2014), Hsing and Eubank (2015) and Aneiros *et al.* (2017). Dimensionality effects have tended to slow down the development of nonparametric modelling ideas in infinite-dimensional setting. However, this field has been investigated many years ago by Ferraty and Vieu (2006) and caused up considerable interest since several hundreds of papers have been published in the last decade. More precisely, dimensionality problem links with probability theory in infinite-dimensional space by means of the small ball probability function of the underlying process and with the topological structure on the infinite-dimensional space. More precisely the interest of using a semi-metric-type topology are discussed in details in the book of Ferraty and Vieu (2006), we may refer for recent references to Bouzebda (2020); Bouzebda and Nezzal (2021).

This paper extends asymptotic results for multivariate statistics of set-indexed conditional empirical process to the context of functional statistical samples. We establish the uniform convergence and asymptotic normality when the observations assumed are strong mixing tacking its values in semi-metric space. It should be noted that even for i.i.d. functional data, no weak convergence has so far been established. To the best of our knowledge, the results presented here, respond to a problem that has not been studied systematically up to the present, which was the basic motivation of the paper.
The remainder of this paper is organized as follows. Section 5.2, we present the notation and definitions together with the conditional empirical process. Section 5.2.1, we give our main results. An application of our main result to the test of the conditional independence is given in Section 5.4. Some concluding remarks and possible future developments are relegated to Section 5.5. To prevent from interrupting the flow of the presentation, all proofs are gathered in Section 5.6.

4.2 The set indexed conditional empirical process

We consider a sample of random elements $(X_1, Y_1), \ldots, (X_n, Y_n)$ copies of (X, Y) that takes its value in a space $\mathcal{E} \times \mathbb{R}^d$. The functional space $\mathcal{E}$ is equipped with a semi-metric $d_{\mathcal{E}}(\cdot, \cdot)$[1]. We aim to study the links between X and Y, by estimating functional operators associated to the conditional distribution of Y given X such as the regression operator, for some measurable set C in a class of sets $\mathscr{C}$,

$$\mathbb{G}(C \mid x) = \mathbb{E}\left(\mathbb{1}_{\{Y \in C\}} \mid X = x\right).$$

This regression relationship suggests to consider the following Nadaraya Watson-type (Nadaraya (1964) and Watson (1964)) conditional empirical distribution:

$$\mathbb{G}_n(C, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{4.1}$$

where $K(\cdot)$ is a real-valued kernel function from $[0, \infty)$ into $[0, \infty)$ and h_n is a smoothing parameter satisfying $h_n \to 0$ as $n \to \infty$, C is a measurable set, and $x \in \mathcal{E}$. By choosing $C = (-\infty, z]$, $z \in \mathbb{R}^d$, it reduces to the conditional empirical distribution function $F_n(z|x) = \mathbb{G}_n((-\infty, z], x)$, refer to Stute (1986a), Stute (1986b), Horváth and Yandell (1988). However, the corresponding class $\mathscr{C} = \left\{(-\infty, z], z \in \mathbb{R}^d\right\}$. Concerning the semi-metric topology defined on $\mathcal{E}$, we will use the notation

$$B(x, t) = \{x_1 \in \mathcal{E} : d_{\mathcal{E}}(x_1, x) \leq t\},$$

[1] A semi-metric (sometimes called pseudo-metric) $d(\cdot, \cdot)$ is a metric which allows $d(x_1, x_2) = 0$ for some $x_1 \neq x_2$.

for the ball in $\mathcal{E}$ with center x and radius t. We denote

$$F(t;x)=\mathbb{P}(d_{\mathcal{E}}(x,X_i)\leq t)=\mathbb{P}(X_i\in B(x,t))=\mathbb{P}(D_i\leq t),$$

which is usually called in the literature the small ball probability function when t is decreasing to zero. One is interested in the behavior of $F(u;x)$ as $u\to 0$. Gasser *et al.* (1998) assume that $F(h;x)=\phi(h_n)f_1(x)$ as $h\to 0$ and refer to $f_1(x)$ as the probability density (functional). When $\mathscr{H}=\mathbb{R}^m$, then $F(h;x)=P\left[\|x-X_i\|\leqslant h\right]$ and it can be seen that in this case $\phi(h_n)=C(m)h^m$ ($C(m)$ is the volume of a unit ball in $\mathbb{R}^m$) and $f_1(x)$ is the probability density of the random variable X_1. Indeed, it can be shown directly that $\lim_{h\to 0}(1/h^m)\,F(h;x)=C(m)f_1(x)$. Motivated by the work of Gasser *et al.* (1998) and the above argument we make the assumption **(H4)**(i)-(ii), refer to this discussion and details to Masry (2005).

Often statistical observations are not independent but are not far from being independent. If not taken into account, dependence can have disastrous effects on statistical inference. The notion of mixing quantifies how close to independence a sequence of random variables is, and it can help us to extend classical results for independent sequences to weakly dependent or mixing sequences, refer to Bradley (2007) for more details. Let us specify the dependence that we will consider in the present paper.

Definition 3. *A sequence $\{\zeta_k, k\geq 1\}$ is said to be α-mixing if the αmixing coefficient*

$$\alpha(n)\overset{def}{=}\sup_{k\geq 1}\sup\left\{|\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)|:A\in\mathscr{F}_{n+k}^{\infty},B\in\mathscr{F}_1^k\right\}$$

converges to zero as $n\to\infty$, where $\mathscr{F}_l^m=\sigma\{\zeta_l,\zeta_{l+1},\ldots,\zeta_m\}$ denotes the σ-algebra generated by $\zeta_l,\zeta_{l+1},\ldots,\zeta_m$ with $l\leq m$. We use the term geometrically strong mixing if, for some $a>0$ and $\beta>1$,

$$\alpha(j)\leq aj^{-\beta},$$

and exponentially strong mixing if, for some $b>0$ and $0<\gamma<1$,

$$\alpha(k)\leq b\gamma^k.$$

Throughout the sequel, we assume tacitly that sequence of random elements $\{(X_i,Y_i), i=1,\ldots,n\}$ is strongly mixing.

4.2.1 Assumptions and notation

Throughout this paper x is a fixed element of the functional space $\mathcal{E}$. We define metric entropy with inclusion which provides a measure of richness(or complexity) of class of sets $\mathscr{C}$. For each $\varepsilon>0$, the covering number is defined as :

$$\begin{aligned}&\mathcal{N}(\varepsilon,\mathscr{C},\mathbb{G}(\cdot\mid x))\\&\quad=\inf\{n\in\mathbb{N}:\ \exists C_1,\ldots,C_n\in\mathscr{C}\text{ such that }\forall C\in\mathscr{C}\ \exists\ 1\leq i,j\leq n\\&\qquad\text{with }C_i\subset C\subset C_j\text{ and }\mathbb{G}(C_j\setminus C_i\mid x)<\varepsilon\},\end{aligned}$$

the quantity $\log(\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)))$ is called metric entropy with inclusion of $\mathscr{C}$ with respect to $\mathbb{G}(\cdot \mid x)$. Estimates for such covering numbers are known for many classes; see, e.g., Dudley (1984). We will often assume below that either $\log \mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x))$ or $\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x))$ behave like powers of ε^{-1}. We say that the condition (R_γ) holds if

$$\log \mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)) \leq H_\gamma(\varepsilon), \text{ for all } \varepsilon > 0, \tag{4.2}$$

where

$$H_\gamma(\varepsilon) = \begin{cases} \log(A\varepsilon) & \text{if } \gamma = 0, \\ A\varepsilon^{-\gamma} & \text{if } \gamma > 0, \end{cases}$$

for some constants $A, r > 0$. As in Polonik and Yao (2002), it is worth noticing that the condition (5.3), $\gamma = 0$, holds for intervals, rectangles, balls, ellipsoids, and for classes which are constructed from the above by performing set operations union, intersection and complement finitely many times. The classes of convex sets in $\mathbb{R}^d$ $(d \geq 2)$ fulfil the condition (5.3), $\gamma = (d-1)/2$. This and other classes of sets satisfying (5.3) with $\gamma > 0$, can be found in Dudley (1984). In this section, we establish the weak convergence of the process $\{\widetilde{\nu}_n(C, x) : C \in \mathscr{C}\}$ defined by

$$\widetilde{\nu}_n(C, x) := \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C, x) - \mathbb{E}\mathbb{G}_n(C, x)\right). \tag{4.3}$$

In our analysis, we will make use of the following assumptions.

(H1) For all $t > 0$, we have $\phi(t) > 0$. For all $t \in (0, 1)$, $\tau_0(t)$ exists, where

$$\tau_0(t) = \lim_{r \to 0} \frac{\phi(rt)}{\phi(r)} = \lim_{r \to 0} \mathbb{P}(d_{\mathcal{E}}(x, X) \leq rt) \mid (\mathbb{P}(d_{\mathcal{E}}(x, X) \leq t)) < \infty;$$

(H2) There exist $\beta > 0$ and $\eta_1 > 0$, such that for all $x_1, x_2 \in N_x$, a neighborhood of x, we have

$$|\mathbb{G}(C \mid x_1) - \mathbb{G}(C \mid x_2)| \leq \eta_1 d_{\mathcal{E}}^{\beta}(x_1, x_2);$$

(i) Let $g_2(u) = \operatorname{Var}\left(\mathbb{1}_{\{Y_j \in C\}} \mid X_j = u\right)$ for $u \in \mathcal{E}$. Assume that $g_2(u)$ is independent of j and is continuous in some neighborhood of x, as $h \to 0$,

$$\sup_{\{u : d(x,u) \leq h\}} |g_2(u) - g_2(x)| = o(1),$$

Assume

$$g_\nu(u) = \mathbb{E}(|\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(C \mid x)|^\nu \mid X_i = u), u \in \mathcal{E},$$

is continuous in some neighborhood of x,

(ii) Define, for $i \neq j, u, v \in \mathcal{E}$,

$$g(u, v; x) = \mathbb{E}((\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(C \mid x))(\mathbb{1}_{\{Y_j \in C\}} - \mathbb{G}(C \mid x)) \mid X_i = u, X_j = v).$$

Assume that $g(u, v; x)$ does not depend on i, j and is continuous in some neighborhood of (x, x);

(H3) There exist $m \geq 2$ and $\eta_2 > 0$, such that, we have, almost surely

$$\mathbb{E}(|Y|^m|X) \leq \eta_2 < \infty;$$

(H4) For all $i \geq 1$,

$$0 < c_5\phi(h_n)f_1(x) \leq \mathbb{P}(X_i \in B(x,h)) = F(h;x) \leq c_6\phi(h_n)f_1(x),$$

where $\phi(h_n) \to 0$ as $h \to 0$ and $f_1(x)$ is a nonnegative functional in $x \in \mathcal{E}$,

(H5) For all $(y_1, y_2) \in \mathbb{R}^{2d}$ and constants $b_3 > 0, \eta_4 > 0$, we have for the conditional density $f(\cdot)$ of Y given $X = x$ the following

$$|f(y_1) - f(y_2)| \leq \eta_4 \| y_1 - y_2 \|^{b_3};$$

(i) $F(u;x) = \phi(u)f_1(x)$ as $u \to 0$, where $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin,

(ii) We have

$$\sup_{i \neq j} \mathbb{P}\left(D_i \leq u, D_j \leq u\right) \leq \psi(u)f_2(x),$$

as $u \to 0$, where $\psi(u) \to 0$ as $u \to 0$. We assume that the ratio $\psi(h)/\phi^2(h)$ is bounded;

(H6) The kernel function $K(\cdot)$ is supported within $(0, 1/2)$ and has a continuous first derivative on $(0, 1/2)$. Moreover, there exist constants $0 < \eta_5 \leq \eta_6 < \infty$ such that:

$$0 < \eta_5 \mathbb{1}_{(0,1/2)}(\cdot) \leq K(\cdot) \leq \eta_6 \mathbb{1}_{(0,1/2)}(\cdot),$$

and

$$K(1/2) - \int_0^{1/2} K^{'}(s)\tau_0(s)ds > 0,\ K^2(1/2) - \int_0^{1/2} (K^2)^{'}(s)\tau_0(s)ds > 0;$$

(H7) Assume the class of sets $\mathscr{C}$ satisfies the condition (5.3);

(H8) (Mixing): for some $v > 2$ and $\delta > 1 - \frac{2}{v}$, we have

$$\sum_{\ell=1}^{\infty} \ell^{\delta}[\alpha(\ell)]^{1-\frac{2}{v}} < \infty;$$

(H9) The smoothing parameter (h_n) satisfies:

$$\frac{\log n}{n \min(a_n, \phi(h_n))} \longrightarrow 0,$$

(i) Let $h_n \to 0$ and $n\phi(h_n) \to \infty$ as $n \to \infty$.
Let v_n be a sequence of positive integers satisfying $v_n \to \infty$ such that $v_n =$

$o((n\phi(h_n))^{1/2})$ and

$$(n/\phi(h_n))^{1/2}\alpha(v_n) \to 0 \quad \text{as} \quad n \to \infty.$$

4.2.2 Comments on the assumptions

The Condition **(H1)** is related to the small ball probabilities, which plays a major role both from theoretical and practical points of view, because the notion of ball is strongly linked with the semi-metric $d(\cdot,\cdot)$, the choice of this semi-metric will become an important stage when the data are tacking its values in some infinite dimensional space. The second part of **(H1)** will be used to control the bias of nonparametric estimators, one needs to have some information on the variability of the small-ball probability. Indeed, in many examples, the small ball probability function can be written approximately as the product of two independent functions in terms of x and h, as in the following examples, which can be found in Proposition 1 of Ferraty *et al.* (2007):

1. $\phi(h_n) = Ch_n^{\upsilon}$ for some $\upsilon > 0$ with $\tau_0(s) = s^{\upsilon}$;
2. $\phi(h_n) = Ch_n^{\upsilon}\exp(-Ch_n^{-p})$ for some $\upsilon > 0$ and $p > 0$ with $\tau_0(s)$ is the Dirac's function;
3. $\phi(h_n) = C\,|\ln(h_n)|^{-1}$ with $\tau_0(s) =_{]0,1]}(s)$ the indicator function in $]0,1]$.

The conditions **(H2)**-**(H3)** are classical in the nonparametric regression estimation. **(H4)** is similar to those in Masry (2005). **(H5)**: About the conditions on the density $f(\cdot)$ is classical Lipschitz-type nonparametric functional model. The conditions on the kernel are not very restrictive. The first part of condition **(H6)** appears in many kernel functional studies and is easily satisfied for wide classes of kernel functions, the interested reader can refer to H_4 in Ferraty *et al.* (2007). The second part of this condition, which is added in this paper as a necessary tool to get uniform results, is linked to the function $\tau_0(\cdot)$ and is also rather general. For example, when $\tau_0(\cdot)$ is identified to be the Dirac mass at $1/2$, the second part of $\tau_0(\cdot)$ is true as long as $K'(s) \leq 0$ and $K(1/2) > 0$. Other examples can be derived from Proposition 2 in Ferraty *et al.* (2007). Condition **(H8)** rules out too large or too small bandwidths for which consistency could not be obtained. It is satisfied with $h_n = \mathcal{O}(\log n)^{-\nu_1}$ (for some suitable $\nu_1 > 0$) as long as the process X is of the exponential type (that is when the small-ball probability function is exponentially decaying). It is also satisfied with $h_n = \mathcal{O}(n/\log n)^{-\nu_2}$ (for some suitable $\nu_2 > 0$) for fractal processes (that is, when the small-ball probability is of polynomial decaying). More details can be found in Ferraty and Vieu (2006).

4.3 Main results

Below, we write $Z \stackrel{\mathcal{D}}{=} \mathcal{N}(\mu,\sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2, $\stackrel{\mathcal{D}}{\to}$ denotes the convergence in distribution and $\stackrel{\mathbb{P}}{\to}$ the convergence in probability.

Theorem 14. *[Uniform Consistency] Suppose that the hypotheses **(H1)**-**(H8)** hold and that (X_t, Y_t) is geometrically strong mixing with $\beta > 2$. Let $\mathscr{C}$ be a class of measurable sets for which*

$$\mathcal{N}\left(\varepsilon, \mathscr{C}, \mathbb{G}\left(\cdot \mid x\right)\right) < \infty$$

for any $\varepsilon > 0$. Suppose further that $\forall C \in \mathscr{C}$

$$|\mathbb{G}(C, y)f(y) - \mathbb{G}(C, x)f(x)| \longrightarrow 0, \quad \textit{as} \quad y \to x.$$

If $n\phi(h_n) \to \infty$ and $h_n \to 0$ as $n \to \infty$, then

$$\sup_{C \in \mathscr{C}} |\mathbb{G}_n(C, x) - \mathbb{E}\left(\mathbb{G}_n(C, x)\right)| \xrightarrow{\mathbb{P}} 0.$$

The proof of this theorem is based on the following relations. Remark that, the proof of Theorem 14 is a direct consequence of the decomposition:

$$\begin{aligned} \mathbb{G}_n(C, x) - \mathbb{E}\left(\mathbb{G}_n(C, x)\right) &= \frac{1}{\mathbb{E}(\widehat{f_n}(x))}\left[\widehat{F_n}(C, x) - \mathbb{E}\left(\widehat{F_n}(C, x)\right)\right] \\ &\quad - \frac{\mathbb{G}_n(C, x)}{\mathbb{E}(\widehat{f_n}(x))}\left[\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n}(x))\right], \end{aligned}$$

where

$$\begin{aligned} \widehat{F_n}(C, x) &= \frac{1}{n\phi(h_n)} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right), \\ \widehat{f_n}(x) &= \frac{1}{n\phi(h_n)} \sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right), \end{aligned}$$

and of the Lemmas 1 and 2 below, for which the proofs are given in the Appendix.

Lemma 1. *Suppose that the hypotheses **(H1)**-**(H8)** hold and for every fixed $C \in \mathscr{C}$ as $n \to \infty$ we have :*

$$\sup_{C \in \mathscr{C}} \left|\widehat{F_n}(C, x) - \mathbb{E}\left(\widehat{F_n}(C, x)\right)\right| = o_{\mathbb{P}}(1)$$

Lemma 2. *Suppose that the hypotheses **(H1)**-**(H8)** hold and for every fixed $N_{\mathcal{E}}$ neighborhood of x in the functional space $\mathcal{E}$ as $n \to \infty$, we have*

$$\sup_{x \in N_{\mathcal{E}}} \left|\widehat{f_n}(x) - \mathbb{E}\left(\widehat{f_n}(x)\right)\right| = o_{\mathbb{P}}(1).$$

Before to establishing the asymptotic normality define the "bias" term by

$$B_n(x) = \frac{\mathbb{E}\left(\widehat{f_n}(x)\right) - \mathbb{G}_n(C, x)\mathbb{E}\left(\widehat{F_n}(C, x)\right)}{\mathbb{E}\left(\widehat{F_n}(C, x)\right)}.$$

By stationarity of order one of the (X_i)'s, we have

$$\mathbb{E}(\widehat{f_n}(x)) = 1.$$

The following result give the weak convergence of our estimators. Keep in mind that $f_1(x)$ is given in **(H5)**.

Theorem 15 (Asymptotic normality). *Let **(H2)-(H5)(i)(ii)-(H6)-(H8)-(H9)(i)** hold and (X_i, Y_j) is geometrically strong mixing with $\beta > 2$, then $n\phi(h_n) \to \infty$ as $n \to \infty$. For $m \geq 1$ and $C_1, \ldots, C_m \in \mathscr{C}$,*

$$\{\widetilde{\nu}_n(C_i, x)_{i=1,\ldots,m}\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = \sigma_{ij}(x), i, j = 1, \ldots, m$ and

$$\sigma_{ij}(x) = \frac{C_2}{C_1^2 f_1(x)} \Big(\mathbb{E}(\mathbb{1}_{\{Y \in C_i \cap C_j\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y \in C_i\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y \in C_j\}} \mid X = x)\Big),$$

whenever $f_1(x) > 0$ and

$$C_1 = K(1/2) - \int_0^{1/2} K^{'}(s)\tau_0(s)ds, \;\; C_2 = K^2(1/2) - \int_0^{1/2} (K^2)^{'}(s)\tau_0(s)ds.$$

To establish the density of the process, we need to introduce the following function which provides the information on the asymptotic behaviour of the modulus of continuity

$$\Lambda_\gamma(\sigma^2, n) = \begin{cases} \sqrt{\sigma^2 \log \dfrac{1}{\sigma^2}}, & \text{if} \quad \gamma = 0; \\ \max\Big((\sigma^2)^{(1-\gamma)/2}, n\phi(h_n)^{(3\gamma-1)/(2(3\gamma+1))}\Big), & \text{if} \quad \gamma > 0. \end{cases}$$

Theorem 16. *Suppose that **(H1)-(H9)** hold and the process (X_i, Y_i) is exponentially strong mixing for each $\sigma^2 > 0$, let $\mathscr{C}_\sigma \subset \mathscr{C}$ be a class of measurable sets with*

$$\sup_{C \in \mathscr{C}_\sigma} \mathbb{G}(C, x) \leq \sigma^2 \leq 1,$$

and suppose that $\mathscr{C}$ fulfils (R_γ) with $\gamma \geq 0$. Further, we assume that $\phi(h_n) \to 0$ and $n\phi(h_n) \to +\infty$ as $n \to +\infty$, such that

$$n\phi(h_n) \leq \Big(\Lambda_\gamma(\sigma^2, n)\Big)^2,$$

and as $n \to +\infty$, we have

$$\frac{n\phi\left(\sigma^2 \log\left(\dfrac{1}{\sigma^2}\right)\right)^{1+\gamma}}{\log(n)} \to \infty.$$

Further we assume that $\sigma^2 \geq h^2$. For $\gamma > 0$ and $d = 1, 2$, the later has to be replaced by $\sigma^2 \geq \phi(h_n) \log\left(\dfrac{1}{\phi(h_n)}\right)$ then for every $\epsilon > 0$ there exist a constant $M > 0$ such that

$$\mathbb{P}\left(\sup_{C \in \mathscr{C}_\sigma} |\widetilde{\nu}_n(C, x)| \geq M\Lambda_\gamma(\sigma^2, n)\right) \leq \epsilon,$$

for all sufficiently large n.

By combining Theorem 15 and Theorem 16 we have the following result.

Theorem 17. *Under conditions of Theorem 15 and Theorem 16, then the process:*

$$\{\widetilde{\nu}_n(C,x) : C \in \mathscr{C}\},$$

converges in law to a Gaussian process $\{\widetilde{\nu}(C,x) : C \in \mathscr{C}\}$, *that admits a version with uniformly bounded and uniformly continuous paths with respect to* $\|\cdot\|_2-$*norm with covariance* $\sigma_{ij}(x)$ *given in Theorem 15.*

Remark 9. *Central limit theorems are usually used to establish confidence intervals for the target to be estimated. In the context of non-parametric estimation the asymptotic variance* $\Sigma := \sigma_{i,j}(x)$ *in the central limit depends on certain functions, including the ones that are estimated. This situation is classic regardless of whether the data is i.i.d. or dependent. As a result, only approximate confidence intervals can be obtained in practice, even when* Σ *functionally specified. To be more precise let us consider the following particular case of Theorem 15, where* $m = 1$. *In the this situation,* Σ *is reduced to, for* $A \in \mathscr{C}$,

$$\sigma^2(x) = \frac{C_2}{C_1^2 f_1(x)}\left(\mathbb{E}(\mathbb{1}_{\{Y\in A\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y\in A\}} \mid X = x)^2\right) = \frac{C_2}{C_1^2 f_1(x)} W_2(x).$$

Observe that the limiting variance contains the unknown function $f_1(\cdot)$ *and that the normalization depends on the function* $\phi(h_n)$ *which is not identifiable explicitly. Let us introduce the following estimate*

$$\mathcal{F}_{x,n}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{d(x,X_i)\leq t\}},$$

One may estimate $\tau_0(\cdot)$ *by*

$$\tau_n(t) = \frac{\mathcal{F}_{x,n}(th)}{\mathcal{F}_{x,n}(h)}.$$

This can use to give the following estimates

$$C_{1,n} = K(1/2) - \int_0^{1/2} K^{'}(s)\tau_n(s)ds, \;\; C2,n = K^2(1/2) - \int_0^{1/2} (K^2)^{'}(s)\tau_n(s)ds.$$

One can estimate $W_{2,n}(x)$ *by*

$$W_{2,n}(x) = (\mathbb{G}_n(C,x) - \mathbb{G}_n^2(C,x)),$$

The use of Theorem 15, in connection with Slutsky's theorem, gives

$$\frac{C_{1,n}}{\sqrt{C_{2,n}}}\sqrt{\frac{n\mathcal{F}_{x,n}(h_n)}{W_{2,n}(x)}}\left(\mathbb{G}_n(C,x) - \mathbb{G}(C,x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This result can be used in the construction of the confident interval in the usual way, we omit the details.

4.3.1 The bandwidth selection criterion

Many methods have been established and developed to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators especially for Nadaraya-Watson regression estimator we quote among them Hall (1984), Härdle (1985), Rachdi and Vieu (2007), Bouzebda and El-hadjali (2020) and Bouzebda (2020). This parameter has to be selected suitably, either in the standard finite dimensional case, or in the infinite dimensional framework for insuring good practical performances. Let us define the leave-out-(X_i, Y_i) estimator for regression function

$$\mathbb{G}_{n,j}(C,x) = \frac{\sum_{i=1, i\neq j}^{n} \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}. \tag{4.4}$$

In order to minimize the quadratic loss function, we introduce the following criterion, we have for some (known) non-negative weight function $\mathcal{W}(\cdot)$:

$$CV\left(C, h\right) := \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{1}_{\{Y_j \in C\}} - \mathbb{G}_{n,j}(C, X_j)\right)^2 \mathcal{W}\left(X_j\right). \tag{4.5}$$

Following the ideas developed by Rachdi and Vieu (2007), a natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose $\widehat{h}_n \in [a_n, b_n]$ minimizing among $h \in [a_n, b_n]$:

$$\sup_{C \in \mathscr{C}} CV\left(\Psi, h\right).$$

The main interest of our results is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable, like in the last equation. One can replace (5.8) by

$$CV\left(C, h_n\right) := \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{1}_{\{Y_j \in C\}} - \mathbb{G}_{n,j}(C, X_j)\right)^2 \widehat{\mathcal{W}}\left(X_j, x\right). \tag{4.6}$$

In practice, one takes, for $j = 1, \ldots, n$, the uniform global weights $\mathcal{W}\left(X_j\right) = 1$, and the local weights

$$\widehat{W}(X_j, x) = \begin{cases} 1 & \text{if} \quad d(X_j, x) \leq h_n, \\ 0 & \quad \text{otherwise.} \end{cases}$$

For sake of brevity, we have just considered the most popular method, that is, the cross-validated selected bandwidth. This may be extended to any other bandwidth selector such the bandwidth based on Bayesian ideas Shang (2014).

4.4 Testing the independence

Concepts of conditional independence play an important role in unifying many seemingly unrelated ideas of statistical inference, see Dawid (1980). Measuring and testing conditional dependence are fundamental problems in statistics, which form the basis of limit theorems, Markov chain, sufficiency and causality Dawid (1979), among others. Conditional independence also plays a central role in graphical modeling Koller and Friedman (2009), causal inference Pearl (2009) and artificial intelligence Zhang *et al.* (2011), refer also to Zhou *et al.* (2020) for recent references. The idea of treating conditional independence as an abstract concept with its own calculus was introduced by Dawid (1979), who showed that many results and theorems concerning statistical concepts such as ancillarity, sufficiency, causality, etc., are just applications of general properties of conditional independence-extended to encompass stochastic and non-stochastic variables together. Let $\mathscr{C}_1, \mathscr{C}_2$ be some classes of sets. In this section, we consider a sample of random elements $(X_1, Y_{1,1}, Y_{1,2}), \ldots, (X_n, Y_{n,1}, Y_{n,2})$ copies of (X, Y_1, Y_2) that takes its value in a space $\mathcal{E} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and define, for $(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\mathbb{G}_n(C_1 \times C_2, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_{i,1} \in C_1\}} \mathbb{1}_{\{Y_{i,2} \in C_2\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{4.7}$$

$$\mathbb{G}_{n,1}(C_1, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_{i,1} \in C_1\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{4.8}$$

$$\mathbb{G}_{n,2}(C_2, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_{i,2} \in C_2\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}. \tag{4.9}$$

We will investigate the following processes, for $(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\widehat{\nu}_n(C_1, C_2, x) = \sqrt{n\phi(h_n)} \left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\mathbb{E}(\mathbb{G}_n(C_2, x))\right), \tag{4.10}$$

$$\breve{\nu}_n(C_1, C_2, x) = \sqrt{n\phi(h_n)} \left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{G}_{n,1}(C_1, x)\mathbb{G}_{n,2}(C_2, x)\right). \tag{4.11}$$

Notice that we have

$$\begin{aligned}
\breve{\nu}_n(C_1, C_2, x) \quad &= \quad \sqrt{n\phi(h_n)} \left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\mathbb{E}(\mathbb{G}_n(C_2, x))\right) \\
&\quad + \sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_2, x)) \left(\mathbb{G}_n(C_1, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\right) \\
&\quad - \sqrt{n\phi(h_n)}(\mathbb{G}_n(C_1, x)) \left(\mathbb{G}_n(C_2, x) - \mathbb{E}(\mathbb{G}_n(C_2, x))\right).
\end{aligned}$$

Hence we have

$$\begin{aligned}
\breve{\nu}_n(C_1, C_2, x) &\overset{d}{=} \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1 \times C_2, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\mathbb{E}(\mathbb{G}_n(C_2, x))\right) \\
&\quad + \sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_2, x))\left(\mathbb{G}_n(C_1, x) - \mathbb{E}(\mathbb{G}_n(C_1, x))\right) \\
&\quad - \sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_1, x))\left(\mathbb{G}_n(C_2, x) - \mathbb{E}(\mathbb{G}_n(C_2, x))\right) \\
&= \widehat{\nu}_n(C_1, C_2, x) + \mathbb{E}(\mathbb{G}_n(C_2, x))\widetilde{\nu}_n(C_1, x) - \mathbb{E}(\mathbb{G}_n(C_1, x)) \qquad (4.12) \\
&\quad \times \widetilde{\nu}_n(C_2, x).
\end{aligned}$$

One can show that, for $(A_1, B_1), (A_2, B_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\begin{aligned}
&\operatorname{cov}(\widehat{\nu}_n(A_1, B_1, x), \widehat{\nu}_n(A_2, B_2, x)) \\
&= \frac{C_2}{C_1^2 f_1(x)}\Big(\mathbb{E}(\mathbb{1}_{\{Y \in A_1 \cap A_2\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y \in A_1\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y \in A_2\}} \mid X = x)\Big) \\
&\quad \times \Big(\mathbb{E}(\mathbb{1}_{\{Y \in B_1 \cap B_2\}} \mid X = x) - \mathbb{E}(\mathbb{1}_{\{Y \in B_1\}} \mid X = x)\mathbb{E}(\mathbb{1}_{\{Y \in B_2\}} \mid X = x)\Big), \qquad (4.13)
\end{aligned}$$

whenever $f_1(x) > 0$. The decomposition in (5.15) give an idea on the process $\breve{\nu}_n(C_1, C_2, x)$ and its structure, however the calculation of the associated covariance more involved. Let $\{\widehat{\nu}(C_1, C_2, x) : (C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2\}$ be a Gaussian process with covariance given in (4.13). Let us introduce the following limiting process, for $(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2$,

$$\breve{\nu}(C_1, C_2, x) = \widehat{\nu}(C_1, C_2, x) + \mathbb{G}(C_2, x)\widetilde{\nu}(C_1, x) - \mathbb{G}(C_1, x)\widetilde{\nu}(C_2, x).$$

We would test the following null hypothesis

$$\mathcal{H}_0 : Y_1 \text{ and } Y_2 \text{ are conditionally independent given } X = x.$$

Against the alternative

$$\mathcal{H}_1 : Y_1 \text{ and } Y_2 \text{ are conditionally dependent.}$$

Statistics of independence those can be used are

$$S_{1,n} = \sup_{(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2} |\widehat{\nu}_n(C_1, C_2, x)|, \qquad (4.14)$$

$$S_{2,n} = \sup_{(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2} |\breve{\nu}_n(C_1, C_2, x)|. \qquad (4.15)$$

A combination of Theorem 17 with continuous mapping theorem we obtain the following result.

Theorem 18. *We have under condition of Theorem 17, as $n \to \infty$,*

$$S_{1,n} \to \sup_{(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2} |\widehat{\nu}(C_1, C_2, x)|, \qquad (4.16)$$

$$S_{2,n} \to \sup_{(C_1, C_2) \in \mathscr{C}_1 \times \mathscr{C}_2} |\breve{\nu}(C_1, C_2, x)|. \qquad (4.17)$$

Remark 10. *It is well known that Theorem 18can be used easily through routine bootstrap sampling as in Bouzebda (2012), Bouzebda and Cherfi (2012) and Bouzebda et al. (2018), which we describe briefly as follows. Let N be a large integer. Let $S_{j,n}^{(1)}, \ldots, S_{j,n}^{(N)}$ be the bootstrapped versions of $S_{j,n}$, for $j = 1, 2$. With the convention that large values of $S_{j,n}$, $j = 1, 2$, lead to the rejection of the null hypothesis $\mathcal{H}_0$, under some regularity conditions, a valid approximation to the P-value for the test based on $S_{j,n}$, $j = 1, 2$, for N large enough, is given by*

$$\frac{1}{N} \sum_{k=1}^{N} 1\!\mathrm{I}\{S_{j,n}^{(k)} \geq S_{j,n}\}.$$

The investigation of the bootstrap should require a different methodology than that used in the present paper, and we leave this problem open for future research.

4.5 Concluding remarks

In the present work, we have established the invariance principle for the conditional set-indexed empirical process formed by strong mixing random variables when the covariates are functional. Our result results are obtained under assumptions on the richness of the index class $\mathscr{C}$ of sets in terms of metric entropy with bracketing in the framework of mixing data. An application of testing the conditional independence is proposed. Notice that mixing is some kind of asymptotic independence assumption which is commonly used for seek of simplicity but which can be unrealistic in situations where there is strong dependence between the data. Extending non-parametric functional ideas to general dependence structure is a rather underdeveloped field. Note that the ergodic framework avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies. It would be interesting to extend our work to the case of the functional ergodic data, which requires non trivial mathematics, this would go well beyond the scope of the present paper.

4.6 Appendix

This section is devoted to the proofs of our results. The aforementioned notation is also used in what follows.

Proof of Lemma 3

Use finite metric entropy with inclusion, fix $\epsilon > 0$ for $C \in \mathscr{C}$. Let C_*, C^* be a bracket for C, i.e., $C_* \subset C \subset C^*$, such that

$$\mathbb{G}\left(C_* \triangle C^* \mid x\right) < \epsilon.$$

Since for $A \subset B$ we have $\mathbb{G}_n(A,x) \leq \mathbb{G}_n(B,x)$ and $\mathbb{G}(A \mid x) \leq \mathbb{G}(B \mid x)$, it follows:

$$\begin{aligned}
&\sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C,x) - \mathbb{E}\left(\mathbb{G}_n(C,x)\right)\right] \\
&\quad\leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C_*,x)\right)\right] \\
&\quad\leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C^*,x)\right)\right] \\
&\qquad + \sup_{C\in\mathscr{C}} \left[\mathbb{E}\left(\mathbb{G}_n(C^*,x)\right) - \mathbb{E}\left(\mathbb{G}_n(C_*,x)\right)\right] \\
&\quad\leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C^*,x)\right)\right] \\
&\qquad + \sup_{C\in\mathscr{C}} \mathbb{G}\left(C_*\triangle C^* \mid x\right) \\
&\quad\leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C^*,x)\right)\right] + \epsilon. \qquad (4.18)
\end{aligned}$$

An analogous lower bound holds with C^* replaced by C_*. Since the first term in the last line is a supremum over finitely sets (for fixed $\epsilon > 0$) it follows pointwise consistency of $\mathbb{G}_n(\cdot,\cdot)$ that the term is $o_{\mathbb{P}}(1)$ and hence we finally obtained. □

Proof of Lemmas 4

We shall proof that

$$\mathbb{P}\left(\left|\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n})\right| > \epsilon\right) \to 0.$$

For that purpose we apply inequality (1.26) of Bosq (1998) to the random variable

$$Y_{in} = K_h\left(d_{\mathcal{E}}(x - X_i)\right) - \mathbb{E}\left(K_h(d_{\mathcal{E}}\left(x - X_i\right))\right).$$

Note that, for $1 \leq i \leq n$, we have

$$\mid Y_{tn} \mid \leq \|K\|_\infty \phi(h_n).$$

By choosing $q = q_n = \sqrt{n}\phi\left(h_n\right)$, then by using the geometrically strong mixing assumption and Billingsley inequality (1.11) in Bosq (1998), this implies that

$$\begin{aligned}
\mathbb{P}(|\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n})| > \epsilon) \quad &\leq \quad 4\exp\left(-\frac{\epsilon^2}{8\|K\|_\infty} q\phi(h_n)\right) \\
&\qquad +22\left(1 + \frac{4\|K\|_\infty \phi(h_n)}{\epsilon}\right)^{\frac{1}{2}} q c_0 \rho^{\left[\frac{n}{2q}\right]}.
\end{aligned}$$

This gives that

$$\mathbb{P}(|\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n})| > \epsilon) \leq \beta\exp\left(-\gamma\sqrt{n\phi(h_n)}\right),$$

where $\beta = \beta(\epsilon, k)$ and $\gamma = (\epsilon, k)$ are strictly positive. We let

$$U_n = \frac{\phi(h_n)}{(\log n)^2}.$$

We finally obtain that

$$\mathbb{P}(|\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n})| > \epsilon) \leq \frac{\beta}{n^{\gamma\sqrt{U_n}}}.$$

Thus the proof is complete. □

Proof of Theorem 20

We will use similar arguments to those used in the paper by Masry (2005). Let

$$\Delta_i(x) = K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right).$$

By conditioning on X_i

$$\mu_n(x) = \mathbb{E}(\mathbb{G}(C \mid X_i) - \mathbb{G}(C \mid x)\Delta_i(x)) \tag{4.19}$$

and making use of the first part of **(H2)**, we have

$$\mu_n(x) \leq \sup_{u \in B(x,h)} |r(u) - r(x)|\mathbb{E}(\Delta_1(x)) \leq const.h^{\beta} X_i \mathbb{E}(\Delta_1(x)). \tag{4.20}$$

Introduce

$$Z_{n,i}(x) = (\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(C \mid x))\Delta_i(x) - \mu_n(x).$$

With this notation, we have

$$Q_n(x) = \frac{1}{n}\sum_{i=1}^{n} Z_{n,i}.$$

We have

$$\sigma^2_{n,0}(x) = \frac{1}{\mathbb{E}^2(\Delta_1(x))}\text{var}(Z_{n,1}).$$

By using the first part of condition **(H6)**, we readily infer that

$$c_1^j c_5 f_1(x)\phi(h_n) \leq \mathbb{E}^j(\Delta_1(x)) \leq c_2^j c_6 f_1(x)\phi(h_n), \quad \text{for} \;\; j = 1, 2. \tag{4.21}$$

Remark that

$$\begin{aligned} n\text{var}(Q_n(x)) &= \frac{1}{\mathbb{E}^2\left(\Delta_1(x)\right)}\text{var}(Z_{n,i}(x)) \\ &\quad + \frac{1}{\mathbb{E}^2(\Delta_1(x))}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{cov}\left(Z_{n,i}(x), Z_{n,j}(x)\right) \\ &= I_1 + I_2. \end{aligned}$$

Note that $I_1 = \sigma^2_{n,0}(x)$ and using (4.20), we infer that

$$\sigma^2_{n,0}(x) = \frac{1}{\mathbb{E}^2(\Delta_1(x))}\mathbb{E}((\mathbb{1}_{\{Y_1 \in C\}} - \mathbb{G}(x))^2\Delta_1^2(x)) + \mathcal{O}\left(h_n^{2\beta}\right).$$

By conditioning on X_1, we obtain that

$$\sigma_{n,0}^2(x) = \frac{1}{\mathbb{E}^2(\Delta_1(x))}\mathbb{E}(g_2(X_1)\Delta_1^2(x)) + \frac{\mathbb{E}((\mathbb{G}(X_1) - \mathbb{G}(x))^2\Delta_1^2(x))}{\mathbb{E}^2(\Delta_1(x))} + \mathcal{O}\left(h_n^{2\beta}\right).$$

Making use of the condition **(H2)** we obtain that the second term is $\mathcal{O}\left(h_n^{2\beta}\right)$. We now establish the upper and the lower bounds of $\sigma_{n,0}^2(x)$. We write

$$\begin{aligned}\mathbb{E}(g_2(X)\Delta_1^2(x)) &= g_2(x)\mathbb{E}(\Delta_1^2(x)) + \mathbb{E}((g_2(X_1)) - g_2(x)\Delta_1^2(x)) \\ &= I_{1,1} + I_{1,2}.\end{aligned}$$

Using assumption **(H2)**(i), we have

$$|I_{1,2}| \leq \sup_{u:d(x,u)\leq h} |g_2(u) - g_2(x)|\mathbb{E}(\Delta_1^2(x)) = o(1)\mathbb{E}(\Delta_1^2(x)),$$

whereas

$$I_{1,1} = g_2(u)\mathbb{E}(\Delta_1^2).$$

Thus we can see that

$$\mathbb{E}(g_2(x)\mathbb{E}(\Delta_1^2(x))) = g_2(1 + o(1))\mathbb{E}(\Delta_1^2(x)).$$

It follows that we have

$$\sigma_{n,0}^2(x) = g_2(x)(1 + o(1))\frac{\mathbb{E}(\Delta_1^2(x))}{\mathbb{E}^2(\Delta_1(x))} + \mathcal{O}\left(h_n^{2\beta}\right).$$

By using (4.21), there exist positive constants c_8 and c_9 in such a way that

$$c_8\frac{g_2}{f_1(x)} + \mathcal{O}\left(\phi(h_n)h_n^{2\beta}\right) \leq \phi(h_n)\sigma_{n,0}^2(x) \leq c_9\frac{g_2(x)}{f_1(x)} + \mathcal{O}\left(\phi(h_n)h_n^{2\beta}\right). \tag{4.22}$$

Then, we readily infer that

$$\begin{aligned}I_2 &= \frac{1}{n\mathbb{E}^2(\Delta_1(x))}\sum_{i=1,1\leq|i-j|\leq a_n}^{n}\sum_{j=1}^{n}\operatorname{cov}\left(Z_{n,i}(x), Z_{n,j}(x)\right) \\ &\quad + \sum_{i=1,|i-j|>a_n}^{n}\sum_{j=1}\operatorname{cov}(Z_{n,i}(x), Z_{n,j}(x)) \\ &= I_{2,1} + I_{2,2},\end{aligned} \tag{4.23}$$

where $a_n = o(n)$ at a specified rate specified in the sequel. For $I_{2,1}$, we have

$$\operatorname{cov}(Z_{n,i}(x), Z_{n,j}(x)) = \mathbb{E}\left(\left(\mathbb{1}_{\{Y_i\in C\}} - \mathbb{G}(C \mid x)\right)\left(\mathbb{1}_{\{Y_j\in C\}} - \mathbb{G}(C \mid x)\right)\right) - \mu_n^2(x).$$

Conditioning on (X_i, X_j) and using condition **(H2)**(ii), we obtain that

$$\operatorname{cov}(Z_{n,i}(x), Z_{n,j}(x)) = \mathbb{E}\left(g(X_i, X_j; x)\Delta_i(x)\Delta_j(x)\right) - \mu_n^2(x).$$

By combining condition **(H6)** (upper bound) and **(H2)**(ii), we show that there exists a finite constant such that

$$|cov(Z_{n,i}(x), Z_{n,j}(x))| \leq const. \sup_{i \neq j} \mathbb{P}\left((X_i, X_j) \in B(x,h) \times B(x,h)\right) + \mu_n^2(x).$$

Making use of the condition **(H4)**(ii) in combination with (4.20), we readily obtain

$$|\text{cov}(Z_{n,i}(x), Z_{n,j}(x))| \leq const. f_2(x)\psi(h_n) + \mathcal{O}\left(h_n^{2\beta}\right) \mathbb{E}^2(\Delta_1(x)).$$

Thus, the use (4.22) implies that

$$|I_{2,1}| \leq \frac{const. f_2(x)\psi(h_n) + \mathcal{O}\left(h_n^{2\beta}\right) \mathbb{E}^2(\Delta_1(x))}{n\mathbb{E}^2(\Delta_1(x))} \sum_{i=1, 1 \leq |i-j| \leq a_n}^{n} \sum_{j=1}^{n} 1 \tag{4.24}$$

$$\leq \frac{const. f_2(x)\psi(h_n)a_n}{\mathbb{E}^2(\Delta_1(x))} + \mathcal{O}\left(h_n^{2\beta}\right) a_n. \tag{4.25}$$

Finally, using the lower bound in (4.21), we obtain

$$|I_{2,1}| \leq \frac{const. f_2(x)\psi(h_n)a_n}{f_1^2 \phi^2(h_n)} + \mathcal{O}\left(h_n^{2\beta}\right) a_n. \tag{4.26}$$

It now follows from the lower bound on $\sigma_{n,0}^2$ in (4.26) that we have

$$\frac{|I_{2,1}(x)|}{\sigma_{n,0}^2(x)} \leq const. \frac{f_2(x)}{f_1(x)g_2(x)} \frac{\psi(h_n)a_n}{\phi(h_n)} + const. \frac{f_1(x)}{g_2(x)} \mathcal{O}\left(h_n^{2\beta}\right) \phi(h_n)a_n. \tag{4.27}$$

We shall subsequently select a_n to make the right side of (4.27) tend to zero as $n \to \infty$. Now consider the contribution of $I_{2,2}$ of (4.23). By Davydov's lemma (Hall (1984), Corollary A.2), we have

$$|\text{cov}(Z_{n,i}(x), Z_{n,j}(x))| \leq 8 \left(\mathbb{E}|(\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(C \mid x))\Delta_i(x)|^{\nu}\right)^{2/\nu} (\alpha(|i-j|))^{1-2/\nu}.$$

By the first part of **(H6)** (upper bound) and the continuity of $g_\nu(\cdot)$ in condition **(H2)**(i), we obtain that

$$\mathbb{E}|(\mathbb{1}_{Y_i \in C}) - \mathbb{G}(C \mid x)\Delta_i(x)|^{\nu} = \mathbb{E}|g_\nu(X_i)\Delta_i(x)|^{\nu} \leq const. \mathbb{P}\left(X_i \in B(x,h)\right).$$

By condition **(H3)**(i) (upper bound), we obtain

$$|\text{cov}(Z_{n,i}(x), Z_{n,j}(x))| \leq const. \left(f_1(x)\phi(h_n)^{2/\nu}\right) (\alpha(|i-j|))^{1-2/\nu}.$$

It then follows from the last relation that

$$I_{2,2} \leq \frac{const f_1^{2/\nu}(x)(\phi(h_n))^{2/\nu}}{n\mathbb{E}^2\left(\Delta_1(x)\right)} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha(|i-j|))^{1-2/\nu}.$$

Using the lower bound (4.21) for $\mathbb{E}(\Delta_1(x))$ and reducing the double sum above into a single sum, we find that

$$I_{2,2} \leq \frac{const}{a_n^\delta f_1^{2(1-1/\nu)}(x)(\phi(h_n))^{2(1-1/\nu)}} \sum_{l=a_n+1}^{\infty} l^\delta (\alpha(l))^{1-2/\nu}.$$

Now, using the lower bound on $\sigma_{n,0}^2$, we obtain

$$\frac{I_{22}}{\sigma_{n,0}^2} \leq \frac{const}{a_n^\delta g_2(x) f_1^{1-2/\nu}(x)(\phi(h_n))^{(1-2/\nu)}} \sum_{l=a_n+1}^{\infty} l^\delta (\alpha(l))^{1-2/\nu}.$$

Now select a_n as $a_n = \frac{1}{(\phi(h_n))^{(1-2/\nu)/\delta}}$, then by condition **(H8)**, we have, as $n \to \infty$,

$$\frac{I_{2,2}}{\sigma_{n,0}^2} \to 0.$$

Now relation (4.27) can be written as follows

$$\frac{I_{2,1}}{\sigma_{n,0}^2} \leq const. \frac{f_2(x)}{f_1(x) g_2(x)} \frac{\psi(h_n)}{\phi^2(h_n)} \phi(h_n) a_n + \frac{f_1(x)}{g_2(x)} \mathcal{O}\left(h_n^{2\beta}\right) \phi(h_n) a_n.$$

The first term on the right side tends to zero since $\psi(h)/\phi^2(h)$ is assumed bounded and $\phi^2(h_n)a_n \to 0$ with the above choice of a_n. The second term clearly tends to zero as $n \to \infty$. It is seen from the proof that the dominating term for $\sigma_{n,0}^2$ is given by

$$g_2(x) \frac{\mathbb{E}(\Delta_1^2(x))}{\mathbb{E}^2(\Delta_1(x))}.$$

Recall $F(u, x) = \mathbb{P}(D_i \leq u)$. Under the assumption **(H1)** and **(H3)**(i), we have, for $j = 1, 2$,

$$\begin{aligned} \frac{1}{\phi(h_n)} \mathbb{E}(\Delta_1^j(x)) &= \frac{1}{\phi(h_n)} \int_0^{h_n} K^j(u/h_n) dF(u; x) \\ &\sim f_1(x) \frac{1}{\phi(h_n)} \int_0^{h_n} K^j(u/h_n) \mathcal{F'}_x(h_n) du \\ &\longrightarrow C_j f_1(x). \end{aligned}$$

It finally follows that we have

$$\sigma_{n,0}^2(x) \longrightarrow \frac{C_2 g_2(x)}{C_1^2 f_1(x)}.$$

Thus the proof is complete. $\square$

Lemma 4.6.1. *Under the present assumptions for each $\epsilon > 0$ and each integer $r \in [1, n/2]$*

there exist positive constants c, c_1, c_2 such that for $C \in \mathscr{C}_\sigma$ and large enough n

$$\begin{aligned} \mathbb{P}\left(|\tilde{\nu}_n(C)| > M\right) &\leq 4\exp\left(-\frac{\epsilon^2}{c(\sigma^2 + \sqrt{\frac{n}{\phi(h_n)}\frac{\epsilon}{r}})}\right) \\ &+ \exp\left[-c_1\frac{n}{r} + c_2\left(\log r + (0 \vee \log\frac{n}{\phi(h_n)\epsilon^2})\right)\right]. \end{aligned} \tag{4.28}$$

The reader can see Polonik and Yao (2002) for the proof. □

Lemma 4.6.2. *Any class $\mathscr{C}$ of set admits a non-increasing upper bound H of the entropy function $\mathbb{H}(\mathscr{C}, \cdot, d_p(\cdot))$ such that $x \to x^4 H(x)$ is non nondecreasing and, for every $\delta \in [0, 1]$,*

$$\psi(\delta) = \int_0^\delta \sqrt{1 \vee H(u)}d(u^t) \leq 4\int_0^\delta \sqrt{1 \vee \mathbb{H}((\mathscr{C}, u, d_p(\cdot)))}d(u^t).$$

The reader is referred to Doukhan (1995) for the proof and the notation. □

Proof of Theorem 16

We recall that the family $\{\tilde{\nu}_n(C) : C \in \mathscr{C}\}$ is dense in the space $(\mathcal{L}^\infty(\mathscr{C}), \|\cdot\|_{\mathscr{C}})$ if for every $M > 0$, we have

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup\mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta}|\tilde{\nu}_n(C)| > M\right) = 0,$$

where

$$\mathscr{C}_\sigma = \{C_1, C_2 \in \mathscr{C} : \mathbb{G}(C_1 \triangle C_2 \mid X = x) \leq \sigma\}.$$

Let $\mathfrak{B}(\eta)$ denote a collection of brackets with (finite) minimal number of sets such that $|\mathfrak{B}(\eta)| = N_I\left(\eta, \mathscr{C}_\sigma, F(\cdot \mid x)\right)$. By definition of H_γ we trivially have under $R(\gamma)$ that

$$\log|\mathfrak{B}(\eta)| \leq H_\gamma(\eta).$$

Now, let $\delta_0 \geq \delta_1 \geq \cdots \geq \delta_N$ and $\eta_0, \eta_1, \ldots, \eta_N$ be positive real numbers defined below. For δ_j, let $C_{1,j}, C_{2,j}$ denote the brackets for $C \in \mathscr{C}$ at the level δ_j^2, which means $C_{1,j} \subset C \subset C_{2,j}$ and

$$\mathbb{G}(C_{1,j} \triangle C_{2,j} \mid X = x) \leq \delta_j^2.$$

Let further $\epsilon, M > 0$ such that

$$\sum_{j=0}^N \eta_j \leq \frac{\epsilon M}{8}, \tag{4.29}$$

then it is easy to see that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta}|\widetilde{\nu}_n(C)|>M\right)\\
\leq\;& |\mathfrak{B}(\eta_0^2)|\mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta}|\widetilde{\nu}_n(C)|>(1-\frac{\epsilon}{4})M\right)\\
&+\sum_{j=0}^{N-1}|\mathfrak{B}(\eta_j^2)||\mathfrak{B}(\eta_{j+1}^2)|\mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta}|\widetilde{\nu}_n(C_{1,j})-\widetilde{\nu}_n(C_{1,j+1})|>\eta_j\right)\\
&+\mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta}|\widetilde{\nu}_n(C_{1,N})-\widetilde{\nu}_n(C)|>\frac{\epsilon}{8}M\eta N\right)\\
=\;& \text{(I)}+\text{(II)}+\text{(III)}.
\end{aligned} \tag{4.30}$$

Expressions (I)-(III) are now estimated separately. As for (I), we choose δ_0 to satisfy

$$H_\gamma(\delta_0^2)=\frac{1}{2c}\left(\frac{(1-\frac{\epsilon}{4})^2B^2}{\sigma^2+\sqrt{\frac{n}{\phi(h_n)}}\frac{(1-\frac{\epsilon}{4})M}{r_0}}\right),$$

with $r_0=\frac{1}{\sigma^2}\sqrt{\frac{n}{\phi(h_n)}}(1-\frac{\epsilon}{4})M$ such that

$$H_\gamma(\delta_0^2)=\frac{(1-\epsilon/4)^2M^2}{2c\sigma^2}.$$

Using the exponential inequality, with $r=r_0$, leads to

$$\begin{aligned}
&(I)\\
=\;& 4\exp\left[-\frac{(1-\frac{\epsilon}{4})^2M^2}{4c\sigma^2}\right]\\
&+\exp\left[\frac{(1-\frac{\epsilon}{4})^2M^2}{4c\sigma^2}-c_1\frac{\sqrt{n\phi(h_n)}\sigma^2}{(1-\frac{\epsilon}{4})^2M^2}+c_2\left(\log n+\left(0\vee\log\frac{n}{\phi(h_n)(1-\frac{\epsilon}{4})^2M^2}\right)\right)\right].
\end{aligned}$$

Since r_0 is between 1 and $\frac{n}{2}$, we obtain the following two conditions

$$M \geq (1-\epsilon/4)^{-1}\sigma^2\sqrt{\frac{\phi(h_n)}{n}}, \tag{4.31}$$

$$M \leq \frac{1}{2}\sigma^2\sqrt{n\phi(h_n)}. \tag{4.32}$$

Now, (4.31) becomes small if B/σ becomes large. We need that for some $B>0$, large enough

$$\frac{M^2}{\sigma^2}-\frac{\sqrt{n\phi(h_n)}\sigma^2}{M^2}\leq -B\log n.$$

This is equivalent to the condition

$$M^4+BM^2\sigma^2\log n\leq\sqrt{n\phi(h_n)}\sigma^4. \tag{4.33}$$

As for the estimation of (II) define $s = \sqrt{\frac{cM}{4\sqrt{\phi(h_n)}}}$ and with δ_0 from above choose N and $\delta_j, j \geq 1$ as follows

$$\delta_{j+1} = s \vee \sup \left\{ x \leq \frac{\delta_j}{2} : H_\gamma(\delta_j^2) \right\},$$

and $N = \min \{j : \delta_j = s\}$. We only consider the case

$$s < \delta_0, \tag{4.34}$$

such that $N \geq 1$. This is the more difficult case the case $s > \delta_0$ follows from Alexander (1984). To make (I) small we take $\delta_0 = \eta_0$ using

$$\begin{aligned} (I) &\leq 2\exp\left((H(t_0)\right)\exp(-N_I(1-\epsilon/4)M, n, \alpha)) \\ &\leq 2\exp\left(-(1-\epsilon)N_I(M, n, \alpha)\right). \end{aligned}$$

To handle (III), one can see that

$$|\widetilde{\nu}_n(C)| \leq 2n^{1/2} \|\mathbb{G}(C_{1,j} \triangle C_{2,j} \mid X = x) \leq \|_\infty \leq \epsilon M/8.$$

Hence, we have $(III) = 0$. We choose for $j = 0, \ldots, N$,

$$\eta_j = \sqrt{20} c \delta_j \sqrt{H_\gamma(\delta_{j+1}^2)},$$

with this choice it easy to see that

$$\sum_{j=1}^{N} \eta_j \leq \sqrt{20} c 2^{3/2} \int_s^{\delta_0} \sqrt{H_\gamma(x^2)} dx.$$

Hence, in view of condition (4.29), we require

$$M \geq B \int_s^{\delta_0} \sqrt{H_\gamma(x^2)} dx, \tag{4.35}$$

for $B \geq B_0 > 0$, (4.28) is now applied to each summand of (II) separately. To that end, we choose quantities $r_j, j = 0, \ldots, N-1$, analogously to r_0. Observing that

$$\mathbb{G}\left(C_{1,j} \triangle C_{1,j+1} \mid x\right) \leq 2\delta_{j+1}^2,$$

we choose

$$r_j = \frac{1}{2\delta_{j+1}^2} \sqrt{\frac{n}{\phi(h_n)}} \eta_j. \tag{4.36}$$

To apply (4.28) with $r = r_j$, we need $1 \leq r_j \leq n/2$, that $r_j \geq 1$ for large enough n can be seen easily. Since r_j is increasing in j, it remains to assure that $1 \leq r_N \leq n/2$. This leads to the conditions

$$M \geq \frac{H_\gamma(s^2)}{\sqrt{\phi(h_n)}}, \tag{4.37}$$

$$M \leq \frac{1}{2} n^{\frac{3}{2}} h^{-\frac{d}{2}} H(s^2). \tag{4.38}$$

Now plugging the above quantities into (4.28), we obtain

$$\begin{aligned}
(II) \quad &= \quad 4 \sum_{j=0}^{N-1} \exp \left[4H_\gamma(\delta_{j+1}^2 - \frac{\eta_j^2}{4c\delta_{j+1}^2}) \right] \\
&\quad + \sum_{j=0}^{N-1} \exp \left[4H_\gamma(\delta_{j+1}^2) - c_1 \frac{\sqrt{\phi(h_n)}\delta_j}{\sqrt{H_\gamma(\delta_{j+1}^2)}} \right. \\
&\quad \left. + c_2 \left(\log n + (0 \vee \log \frac{n}{\phi(h_n) s^2 H_\gamma(s^2)} \right) \right] \\
&\leq \quad 4 \sum_{j=0}^{N-1} \exp \left[-H_\gamma(\delta_{j+1}^2) \right] \\
&\quad + \sum_{j=0}^{N-1} \exp \left[4H_\gamma(\delta_{j+1}^2) - c_1 \frac{\sqrt{\phi(h_n)}\delta_j}{\sqrt{H_\gamma(\delta_{j+1}^2)}} \right. \\
&\quad \left. + c_2 \left(\log n + (0 \vee \log \frac{n}{\phi(h_n) s^2 H_\gamma(s^2)}) \right) \right].
\end{aligned} \tag{4.39}$$

Using the fact that $H_\gamma(\delta_{j+1}^2) \geq 2H_\gamma(\delta_j^2)$, the term in (4.39) can be shown to be (at least) of the same order as (4.31). As for the term (4.39) give

$$\log \frac{n}{\phi(h_n) s^2 H_\gamma(s^2)} = \mathcal{O}(\log n).$$

In order to get (4.39) small, we need that for all $j = 0, 1, \ldots, N-1$

$$4H_\gamma(\delta_{j+1}^2) c_1 \frac{\sqrt{n\phi(h_n)}\delta_j}{\sqrt{H_\gamma(\delta_{j+1}^2)}} - c_1 \log n \leq -A_j(n), \tag{4.40}$$

for some real valued function A_j such that

$$\sum_{j=1}^{N} \exp(-A_j(n)) < \infty.$$

Since the left hand side of (4.31) is increasing in j, it suffices to choose $A_N(n)$ satisfying (4.31), which means that we need

$$4\sqrt{H(s^2)} \left(H(s^2) + c_3 \log n + A_N(n) \right) \leq c_1 \left(n\phi(h_n) \right)^{\frac{1}{4}} \sqrt{M}, \tag{4.41}$$

satisfying in addition

$$N \exp\left(-A_N(n)\right) < \infty. \tag{4.42}$$

It remains to consider (III). Using Lemma 5.2 of Polonik and Yao (2002) and arguments as in (5.6), we obtain

$$\begin{aligned}\widetilde{\nu}_n(C) &\leq \widetilde{\nu}_n(C_2) + \sqrt{\phi(h_n)}\left(\mathbb{E}\left(\mathbb{G}_n(C_2,x) - \mathbb{E}\mathbb{G}_n(C_1,x)\right)\right) \\ &\leq \phi(h_n)\left(\left(\mathbb{G}_n(C_2,x) - \mathbb{E}\mathbb{G}_n(C_2,x)\right)\right) + \mathcal{O}\left(\sqrt{n\phi(h_n)}h_n^2\right) \\ &\quad + \sqrt{n\phi(h_n)}f(x)\eta_n. \qquad (4.43)\end{aligned}$$

Analogously we have an estimate of $\widetilde{\nu}_n(C)$ from below by replacing C_2 by C_1 in (4.43) and (4.43). Hence we obtain

$$\begin{aligned}(III) &= \mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta} |\widetilde{\nu}_n(C_{1,N}) - \widetilde{\nu}_n(C_{2,N})|\right. \\ &\qquad \left. > \frac{\epsilon}{8}M + \eta_N - \sqrt{n\phi(h_n)}f(x)\delta_N - c_3\sqrt{n\phi(h_n)}h^2\right) \\ &\leq \mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta} |\widetilde{\nu}_n(C_{1,N}) - \widetilde{\nu}_n(C_{2,N})|\right. \\ &\qquad \left. > \frac{\epsilon}{8}M + \eta_N - c_4\sqrt{n\phi(h_n)}f(x)\delta_N\right) \\ &\leq \mathbb{P}\left(\sup_{C\in\mathscr{C}_\delta} |\widetilde{\nu}_n(C_{1,N}) - \widetilde{\nu}_n(C_{2,N})| > \eta_N\right). \qquad (4.44)\end{aligned}$$

For the second inequality, we used the fact the fact that $h^2 = \mathcal{O}(\delta_N)$ or equivalently

$$M^2 \geq Bn\phi(h_n)^s, \qquad (4.45)$$

for some $B > 0$. Hence (III) can be treated as (II) above. Now we consider the different cases of γ and check the above conditions on M below. We frequently use B to denote a positive constant which has to be chosen appropriately (usually large enough) and which usually is different at different places. As for $\gamma = 0$, we have

$$\int_s^{\delta_0} \sqrt{H_\gamma(x^2)}dx = \mathcal{O}\left(\sqrt{\delta_0^2 \log \frac{1}{\delta_0^2}}\right).$$

In view of (4.31) $M^2 = \sigma^2 D(\sigma^2)$ with $D(\sigma^2) \to \infty$ as $\sigma^2 \to 0$ using (4.35) leads to the choice

$$M = B\sqrt{\sigma^2 \log \frac{1}{\sigma^2}}.$$

Note that here N can be chosen as $N = \mathcal{O}(\log\log n)$, such that $A_N(n) = \log n$ is a valid choice. With these choices, condition (4.31) is satisfied automatically for large enough n, and (4.31) holds if

$$\frac{\sigma^2}{\log \frac{1}{\sigma^2}} \geq \frac{4B}{\phi(h_n)}.$$

Further (4.33) holds for large enough n if $\frac{\sqrt{\phi(h_n)}}{\log\frac{1}{\sigma^2}\log n}$ as $n\to\infty$ (4.30) follows automatically, and (4.37) holds if

$$\frac{\phi(h_n)\sigma^2\log\frac{1}{\sigma^2}}{(\log n)^2} > B > 0.$$

Inequality (4.45) follows from the assumption that

$$\sqrt{\phi(h_n)^4} = \mathcal{O}\left(\Lambda_0(\sigma^2), n\right).$$

Finally (4.41) follows from

$$\frac{\phi(h_n)\sigma^2\log\frac{1}{\sigma^2}}{(\log n)^6} \quad as \quad n\to\infty,$$

which is the strongest condition. As for $\gamma > 0$ a crucial condition again is (4.41). This condition can be seen to holds if

$$M \geq B(\phi(h_n))^{\frac{3\gamma-1}{2(3\gamma+1)}}.$$

For $0<\gamma<1$, we have

$$\int_s^{\delta_0}\sqrt{H_\gamma(x^2)}dx = \mathcal{O}(\delta_0^{1-\gamma}).$$

Hence we choose

$$M = B\left(\max\left(\left(\sigma^2\right)^{\frac{1-\gamma}{2}}, (\phi(h_n))^{\frac{3\gamma-1}{2(3\gamma+1)}}\right)\right). \tag{4.46}$$

Note further that with this choice of M, we may assume $N=\mathcal{O}(\log n)$ such that again $A_N = \log n$ is a valid choice, satisfying (4.42). With these choices all the above conditions on M are satisfied under the present assumptions. To see this first assume that

$$(\sigma^2)^{\frac{1-\gamma}{2}} \geq \left((n\phi(h_n))^{\frac{3\gamma-1}{2(3\gamma+1)}}\right),$$

such that $M = B(\sigma^2)^{\frac{1-\gamma}{2}}$. In this case, (4.31) hold automatically for large n, and (4.31) follows from

$$\sigma^2 \geq B(\phi(h_n))^{-\frac{1}{\gamma+1}}.$$

Further (4.33) holds if

$$\sqrt{n\phi(h_n)}(\sigma^2)^{2\gamma} \geq B\log n,$$

the equation (4.37) reads as

$$M \geq (n\phi(h_n))^{-\frac{1-\gamma}{2(\gamma+1)}},$$

which means here

$$\sigma^2 \geq B(n\phi(h_n))^{-\frac{1}{\gamma+1}},$$

equation (4.30) reads as

$$M \leq n^{\frac{3+\gamma}{2(\gamma+1)}}h^{-\frac{d(1-\gamma)}{2(1+\gamma)}},$$

which of course is satisfied here since $\sigma^2 \leq 1$. Last but not least, the assumption $s \leq \delta_0$ means,

$$M \leq (\phi(h_n))^{\frac{\gamma}{2(\gamma+2)}}(\sigma^2)^{\frac{1}{\gamma+2}}.$$

Plug in our choice of M again leads to

$$\sigma^2 \geq B\,(n\phi(h_n))^{-\frac{3\gamma-1}{2(3\gamma+1)}}.$$

It remains to consider the case

$$(\sigma^2)^{\frac{1-\gamma}{2}} < (n\phi(h_n))^{\frac{3\gamma-1}{2(3\gamma+1)}}, \tag{4.47}$$

such that

$$M = B\,(n\phi(h_n))^{\frac{3\gamma-1}{2(3\gamma+1)}}.$$

Here (4.31) holds automatically for large n because $\sigma^2 \leq 1$, conditions (4.31) and (4.33) lead to lower estimates for σ^2 which do not conflict with (4.47) and (4.37) and (4.30) are also seen easily to be satisfied. Finally inequality (4.45) follows for the same reasons as given above in the case $\gamma = 0$. □

Acknowledgement

The authors are indebted to the Editor-in-Chief, Associate Editor and the referee for their very valuable comments, suggestions careful reading of the article which led to a considerable improvement of the manuscript.

CRediT author statement

Youssouf Souddi : Conceptualization, Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing.
Salim Bouzebda and Fethi Madani : Conceptualization, Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing.
All authors contributed equally to this work.

CHAPTER 5

SOME CHARACTERISTICS OF THE CONDITIONAL SET-INDEXED EMPIRICAL PROCESS INVOLVING FUNCTIONAL ERGODIC DATA

This chapter is the subject of publication : Bulletin of the Institute of Mathematics, Academia Sinica new series.
DOI: 10.21915/BIMAS.2021405

The purpose of this chapter is to establish the invariance principle for the conditional set-indexed empirical process formed by functional ergodic random variables. The limit theorems, discussed in this paper, are key tools for many further developments in functional data analysis involving empirical process techniques. These results are proved under some standard structural conditions on the Vapnik-Chervonenkis classes of functions and some mild conditions on the model.

5.1 Introduction

The theory of empirical process is branch of statistics that is play fundamental role in its various applications especially important in estimation theory there has been a great deal research works. The asymptotic properties of empirical processes indexed by functions have been intensively studied during the past decades (see, e.g., Van der Vaart and Wellner (1996) or Dudley (1999) for self-contained, comprehensive books on the topic with various statistical applications). Many authors have studied it in the last century in finite framework, so that it developed rapidly due to its role in solving problems of statistics ,modulo measurability, the classes $\mathscr{C}$ of sets for which the Glivenko-Cantelli theorem holds characterize by Vapnik and Červonenkis (1971) in the setting of independents variables and in this framework many results had been obtained we cite Dudley (1978), Giné and Zinn (1984), Le Cam (1983), Pollard (1982) and Bass and Pyke (1984). Empirical processes based on dependent data have been studied under various mixing conditions, for example, Yoshihara (1990) established the asymptotic normality when the sequences are ϕ-mixing, in these lines of research in different type of mixing, we may cite Eberlein (1984), Nobel and Dembo (1993) and Yu (1994). However, a bracketing condition under strong mixing was stated by Andrews and Pollard (1994). Doukhan (1995) studied the function-indexed empirical process for β-mixing sequences, where Arcones (1994) was given results the case of Gaussian long-range dependent random vectors, Polonik and Yao (2002) have established uniform convergence and asymptotic normality of set-indexed conditional empirical process in a strictly stationary and strong mixing framework and derived the Bahadur Kiefer approximations of conditional quantile in this framework Poryvaĭ (2005) extended the work of Polonik and Yao (2002). On the other hand, the modelization of functional variables that taking values in infinite dimensional spaces had received a lot of attention in the last few years, there are an increasing number of situation coming from different fields of applied sciences (environment, chemometrics, biometrics, medicine, econometrics,....) in which the collected data are curves, the study of statistical models adapted to such type of infinite dimensional data has been the subject of several works in the recent statistical literature good overviews about this literature can be found in Ramsay and Silverman (2005a), Bosq (2000), Ramsay and Silverman (2005b), Ferraty and Vieu (2006), Bosq and Blanke (2007), Shi and Choi (2011), Horváth and Kokoszka (2012), Zhang (2014), Bongiorno *et al.* (2014), Hsing and Eubank (2015) and Aneiros *et al.* (2017) and hundreds of papers and books have been published in this framework last decade.

However, there are a few results for the empirical process considered functional framework, we may refer for recent references to Bouzebda (2020,1); Bouzebda and Nezzal (2021), Bouzebda and Chaouch (2022). Bouzebda *et al.* (2021) obtained several very useful results for set-indexed conditional empirical processes in functional setting the strong mixing dependence. Notice that mixing is some kind of asymptotic independence assumption which is commonly used for seeking simplicity but which can be unrealistic in situations where there is strong dependence between the data. Extending non-parametric functional ideas to general dependence structure is a rather underdeveloped field, the ergodic framework avoids the widely used strong mixing condition and its variants to measure the dependency which go far beyond the invariance principle that is the basic motivation of the paper. The general framework of ergodic functional data has been initiated by Laib and Louani (2010) who stated consistencies with rates together with the asymptotic normality of the regression function estimate, for recent paper on the subject we refer to Bouzebda and Chaouch (2022), where the authors extended the last reference to a more general framework. For reader convenience, we introduce some details defining the ergodic property of processes and its link with the mixing one. Let $\{X_n, n \in \mathbb{Z}\}$ be a stationary sequence. Consider the backward field $\mathscr{A}_n = \sigma(X_k : k \leq n)$ and the forward field $\mathscr{B}_m = \sigma(X_k : k \geq m)$. The sequence is strongly mixing if, as $n \to \infty$,

$$\sup_{A \in \mathscr{A}_0, B \in \mathscr{B}_n} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \alpha(n) \to 0.$$

The sequence is ergodic if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| \mathbb{P}\left(A \cap \tau^{-k} B\right) - \mathbb{P}(A)\mathbb{P}(B) \right| = 0,$$

where τ is the time-evolution or shift transformation. The naming of strong mixing in the above definition is more stringent than what is ordinarily referred (when using the vocabulary of measure preserving dynamical systems) as strong mixing, namely to that $\lim_{n\to\infty} \mathbb{P}\left(A \cap \tau^{-n} B\right) = \mathbb{P}(A)\mathbb{P}(B)$ for any two measurable sets A, B, see, for instance Rosenblatt (1972). Hence, strong mixing implies ergodicity, whereas the inverse is not always true (see e.g. Remark 2.6 in page 50 in connection with Proposition 2.8 in page 51 in Bradley (2007)). Some motivations to consider ergodic dependence structure in the data rather than a mixing one are discussed in Laib and Louani (2010); Bouzebda *et al.* (2015); Bouzebda and Didi (2017b,a, 2021) where details on the definition of ergodic property of processes together with illustrating examples of such processes are also given. The aim of the present paper is to extend asymptotic results for set-indexed conditional empirical processes to the context of functional ergodic data. We establish uniform convergence and asymptotic normality when the observations are assumed to be ergodic in nature taking their values in semi-metric space. This paper responds to a problem that has not been studied systematically up to the present The remainder of this paper is organized as follows. Section 5.2, we present the notation and definitions together with the conditional empirical process. Section 5.3, we give our main results. We discuss the bandwidth selection procedure in Section 5.3.1. An application of our main result to the test of the conditional independence is given in Section 5.4. Some concluding remarks and possible future developments are relegated to Section 5.5. To prevent from interrupting the flow of the presentation, all proofs are gathered in Section 5.6. Some examples are collected in Section 5.7.

5.2 The set indexed conditional empirical process

For the sake of clarity, introduce some details defining the ergodic property of processes. Taking a measurable space $(S, \mathscr{J})$ denote by $S^{\mathbb{N}}$ the space of all functions $s : \mathbb{N} \to S$. If s_j is the value the function s takes at $j \in \mathbb{N}$, define H_j as the j-th coordinate map, i.e $H_j(s) = s_j$ and consider $H_j^{-1}(\mathscr{J}), j \in \mathbb{N}$ a random process $Z = \{Z_j : j \in \mathbb{N}\}$ can be considered as random variable defined on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in $(S^{\mathbb{N}}, \mathscr{J}^N)$. Now a set $B \in \mathcal{F}$ is called invariant if there exists some set $\mathcal{A} \in \mathscr{J}^{\mathbb{N}}$ such that $B = \{(Z_n, Z_{n+1}, \dots) \in \mathcal{A}\}$ is true for any $n \geq 1$. The process Z is then said ergodic whenever, for any invariant set B, we have $\mathbb{P}(B) = 0$ or $\mathbb{P}(\Omega \mid B) = 0$. It is well known from the ergodic theorem that, for a stationary ergodic process Z, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Z_i = \mathbb{E}(Z_1) \quad \text{almost surely.} \tag{5.1}$$

Therefore, the ergodic property in our setting is formulated on the basis of the statement (5.1). We consider a sample of random elements $(X_1, Y_1), \ldots, (X_n, Y_n)$ copies of (X, Y) that takes its value in a space $\mathcal{E} \times \mathbb{R}^d$. The functional space $\mathcal{E}$ is equipped with a semi-metric $d_{\mathcal{E}}(\cdot,\cdot)$. We aim to study the links between X and Y, by estimating functional operators associated to the conditional distribution of Y given X such as the regression operator, for some measurable set C in a class of sets $\mathscr{C}$,

$$\mathbb{G}(C \mid x) = \mathbb{E}\left(\mathbb{1}_{\{Y\in C\}} \mid X = x\right).$$

This regression relationship suggests to consider the following Nadaraya Watson-type (Nadaraya (1964) and Watson (1964)) conditional empirical distribution:

$$\mathbb{G}_n(C, x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{Y_i\in C\}} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}, \tag{5.2}$$

where $K(\cdot)$ is a real-valued kernel function from $[0,\infty)$ into $[0,\infty)$ and h_n is a smoothing parameter satisfying $h_n \to 0$ as $n \to \infty$, C is a measurable set, and $x \in \mathcal{E}$. By choosing $C = (-\infty, z]$, $z \in \mathbb{R}^d$, it reduces to the conditional empirical distribution function $F_n(z|x) = \mathbb{G}_n((-\infty, z], x)$, refer to Stute (1986a), Stute (1986b), Horváth and Yandell (1988). However, the corresponding class $\mathscr{C} = \left\{(-\infty, z], z \in \mathbb{R}^d\right\}$. Concerning the semi-metric topology defined on $\mathcal{E}$, we will use the notation

$$B(x, t) = \{x_1 \in \mathcal{E} : d_{\mathcal{E}}(x_1, x) \leq t\},$$

for the ball in $\mathcal{E}$ with center x and radius t, usually called in the literature the small ball probability function when t is decreasing to zero. This notion plays a major role both from theoretical and practical points of view, because the notion of ball is strongly linked with the semi-metric $d(\cdot,\cdot)$, the choice of this semi-metric will become an important stage when the data is taking its values in some infinite dimensional space. Indeed, in many examples, the small ball probability function can be written approximately as the product of two independent functions in terms of x and h, as in the following examples, which can be found in Proposition 1 of Ferraty *et al.* (2007):

1. $\phi(h_n) = Ch_n^{\upsilon}$ for some $\upsilon > 0$ with $\tau_0(s) = s^{\upsilon}$;
2. $\phi(h_n) = Ch_n^{\upsilon}\exp(-Ch_n^{-p})$ for some $\upsilon > 0$ and $p > 0$ with $\tau_0(s)$ is the Dirac's function;
3. $\phi(h_n) = C\,|\ln(h_n)|^{-1}$ with $\tau_0(s) =_{]0,1]}(s)$ the indicator function in $]0, 1]$.

Let $\mathcal{F}_i$ be the σ-filed generated by $((X_1, Y_1), \ldots, (X_i, Y_i))$ and $\mathfrak{G}_i$ that generated by $((X_1, Y_1), \ldots, (X_i, Y_i), X_{i+1})$. Let $B(x, u)$ be a ball centered at $x \in \mathcal{E}$ with radius u. Let $D_i = d(x, X_i)$ so that D_i is a nonnegative real-valued random variables. Working on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let

$$F_x(u) = \mathbb{P}(D_i \leq u) = \mathbb{P}(X_i \in B(x, u)),$$

and $F_x^{\mathcal{F}_{i-1}} = \mathbb{P}(X_i \in B(x, u) \mid \mathcal{F}_{i-1})$ be the distribution function and the conditional distribution function, given the σ-filed $\mathcal{F}_{i-1}$ of $(D_i)_{i\geq 1}$ respectively. Denote by $o_{a.s}(u)$ a real random function l such that $l(u)/u$ converges to zero almost surely as $u \to 0$. Similarly define $\mathcal{O}_{a.s}(u)$ as a real random function l such that $l(u)/u$ is almost surely bounded. Throughout the sequel, we assume tacitly that the sequence of random elements $\{(X_i, Y_i), i = 1, \ldots, n\}$ is ergodic.

5.2.1 Assumptions and notation

Throughout this paper x is a fixed element of the functional space $\mathcal{E}$. We define the metric entropy with inclusion which provides a measure of richness (or complexity) of the class of sets $\mathscr{C}$. For each $\varepsilon > 0$, the covering number is defined as :

$$\begin{aligned}
&\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)) \\
&= \inf\{n \in \mathbb{N} : \exists C_1, \ldots, C_n \in \mathscr{C} \text{ such that } \forall C \in \mathscr{C}\ \exists\ 1 \leq i, j \leq n \\
&\quad \text{with } C_i \subset C \subset C_j \text{ and } \mathbb{G}(C_j \setminus C_i \mid x) < \varepsilon\},
\end{aligned}$$

the quantity $\log(\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)))$ is called metric entropy with inclusion of $\mathscr{C}$ with respect to $\mathbb{G}(\cdot \mid x)$. Estimates for such covering numbers are known for many classes; see, e.g., Dudley (1984). We will often assume below that either $\log \mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x))$ or $\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x))$ behave like powers of ε^{-1}. We say that the condition (R_γ) holds if

$$\log \mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)) \leq H_\gamma(\varepsilon), \text{ for all } \varepsilon > 0, \tag{5.3}$$

where

$$H_\gamma(\varepsilon) = \begin{cases} \log(A\varepsilon) & \text{if} \quad \gamma = 0, \\ A\varepsilon^{-\gamma} & \text{if} \quad \gamma > 0, \end{cases}$$

for some constants $A, r > 0$. As in Polonik and Yao (2002), it is worth noticing that the condition (5.3), $\gamma = 0$, holds for intervals, rectangles, balls, ellipsoids, and for classes which are constructed from the above by performing set operations union, intersection and complement finitely many times. The classes of convex sets in $\mathbb{R}^d$ $(d \geq 2)$ fulfill the condition (5.3), $\gamma = (d-1)/2$. This and other classes of sets satisfying (5.3) with $\gamma > 0$, can be found in Dudley (1984).

Example 5. *Bouzebda et al. (2016) The set $\mathcal{C}$ all indicator functions $\mathbb{1}_{(\infty,t]}$ of cells in $\mathbb{R}$ satisfies*

$$\mathcal{N}\left(\varepsilon, \mathcal{C}, d_\gamma^{(2)}\right) \leq \frac{2}{\epsilon^2}$$

for any probability measure γ *and* $\epsilon \leq 1$.*Notice that*

$$\int_0^1 \sqrt{\log(\frac{1}{\epsilon})}d\epsilon \leq \int_0^\infty u^{1/2}\exp(-u)du \leq 1$$

For more details and discussion on this example refer to example 2.5.4 in Van der Vaart and Wellner (1996)

We give now further notation. For $j \geq 1$ set

$$M_j = K^j(1) - \int_0^1 (K^j)'(u)\tau_0(u)du.$$

In this section, we establish the weak convergence of the process $\{\widetilde{\nu}_n(C,x) : C \in \mathscr{C}\}$ defined by

$$\widetilde{\nu}_n(C,x) := \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C,x) - \mathbb{E}\mathbb{G}_n(C,x)\right). \tag{5.4}$$

In our analysis, we will make use of the following assumptions.

(H1) For $x \in \mathcal{E}$, there exists a sequence of nonnegative bounded random functionals $(f_{i,1})_{i\geq 1}$, a sequence of random functions $(g_{i,x})_{i\geq 1}$ a deterministic nonnegative bounded functional f_1 and a nonnegative real function ϕ where $\phi(h_n) \to 0$ as $h \to 0$ such that

(i) $F_x(u) = \phi(u)f_1(x) + o(\phi(u))$ *as* $u \to 0$.

(ii) For any $i \in \mathbb{N}$, $F_x^{\mathfrak{F}_{i-1}}(u) = \phi(u)f_{i,1}(x) + g_{i,x}(u) = o_{a.s}(\phi(u))$ as $u \to 0$. $g_{i,x}(u)/\phi(u)$ almost surly bounded and $n^{-1}\sum_{i=1}^n g_{i,x}^j(u) = o_{a.s}(\phi^j(u))$ *as* $n \to \infty, j = 1,2$.

(iii) $n^{-1}\sum_{i=1}^n f_{i,1}^j(x) \to f_1^j(x)$ almost surely as$n \to \infty$, for $j = 1,2$.

(iv) There exists nondecreasing bounded function $\tau_0(u)$ such that uniformly for all $u \in (0,1)$,

$$\tau_0(u) + o(1) = \frac{\phi(ru)}{\phi(r)}$$

as $r \downarrow 0$ and $1 \leq j \leq 2+\delta$ *with* $\delta > 0$, $\int_0^1 (K^j(u))'\tau_0(u)du < \infty$.

(H2) (i) There exist $\beta > 0$ and $\eta_1 > 0$, such that for all $x_1, x_2 \in N_x$, a neighborhood of x, we have

$$|\mathbb{G}(C \mid x_1) - \mathbb{G}(C \mid x_2)| \leq \eta_1 d_{\mathcal{E}}^{\beta}(x_1,x_2).$$

(H3) There exist $m \geq 2$ and $\eta_2 > 0$, such that, we have, almost surely

$$\mathbb{E}(|Y|^m|X) \leq \eta_2 < \infty;$$

(i) The conditional mean of $\mathbb{1}_{\{Y_i\in C\}}$ given the σ-field $\mathfrak{G}_{i-1}$ depends only on X_i, i.e., for any $i \geq 1$,$\mathbb{E}\left(\mathbb{1}_{\{Y_1\in C\}} \mid \mathfrak{G}_{i-1}\right) = \mathbb{G}(X_i)$ almost surely.

(ii) The conditional mean of $\mathbb{1}_{\{Y_i\in C\}}$ given the σ-field $\mathfrak{G}_{i-1}$ depends only on X_i, i.e., for any $i \geq 1$, $\mathbb{E}\left(\left(\mathbb{1}_{\{Y_1\in C\}} - \mathbb{G}(X_i)\right)^2 \mid \mathfrak{G}_{i-1}\right) = W_2(X_i)$ almost surely.

Moreover, the function W_2 is continuous in a neighborhood of x, that is,

$$\sup_{\{u:d(x,u)\leq h\}} |W_2(u) - W_2(x)| = o(1) \quad as \quad h \to 0;$$

(H4) For all $(y_1, y_2) \in \mathbb{R}^{2d}$ and constants $b_3 > 0, \eta_4 > 0$, we have for the conditional density $f(\cdot)$ of Y given $X = x$ the following

$$|f(y_1) - f(y_2)| \leq \eta_4 \| y_1 - y_2 \|^{b_3};$$

(i) $F(u; x) = \phi(u) f_1(x)$ as $u \to 0$, where $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin,

(H5) The kernel function $K(\cdot)$ is supported within $(0, 1)$ and has a continuous first derivative on $(0, 1)$ and satisfied the condition $K'(t) < 0 \quad \forall t \in (0, 1)$. Moreover,

$$\left| \int_0^1 (K^j)'(u) du \right| < \infty, \quad \text{for} \quad j = 1, 2.$$

(H6) Assume the class of sets $\mathscr{C}$ satisfies the condition (5.3);

(H7) The smoothing parameter (h_n) satisfies:

(i) $\frac{\log n}{n \min(a_n, \phi(h_n))} \longrightarrow 0,$

(ii) Let $h_n \to 0$ and $n\phi(h_n) \to \infty$ as $n \to \infty$.

5.2.2 Comments on the assumptions

The condition **(H1)** plays an important role in the ergodic and functional context of this paper condition used here share some similarities with that used in Laib and Louani (2010). Conditions **(H2)**(i) are classical in the nonparametric regression estimation. **(H2)**(ii) stand as regularity conditions that are of usual nature. **(H3)** is necessary to establish consistency. The condition **(H4)** on the density $f(\cdot)$ is a classical Lipschitz-type nonparametric functional model. **(H5)** The conditions on the kernel are not very restrictive. **(H7)** rules out too large or too small bandwidths without the consistency that could not be obtained.

5.3 Main results

Below, we write $Z \overset{\mathcal{D}}{=} \mathcal{N}(\mu, \sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2, $\overset{\mathcal{D}}{\to}$ denotes the convergence in distribution and $\overset{\mathbb{P}}{\to}$ the convergence in probability.

Theorem 19. *[Uniform Consistency] Suppose that the hypotheses* ***(H1)-(H7)*** *hold. Let $\mathscr{C}$ be a class of measurable sets for which*

$$\mathcal{N}(\varepsilon, \mathscr{C}, \mathbb{G}(\cdot \mid x)) < \infty,$$

for any $\varepsilon > 0$. Suppose further that $\forall C \in \mathscr{C}$

$$|\mathbb{G}(C, y) f(y) - \mathbb{G}(C, x) f(x)| \longrightarrow 0, \quad as \quad y \to x.$$

If $n\phi(h_n) \to \infty$ and $h_n \to 0$ as $n \to \infty$, then

$$\sup_{C\in\mathscr{C}} |\mathbb{G}_n(C,x) - \mathbb{E}(\mathbb{G}_n(C,x))| \xrightarrow{\mathbb{P}} 0.$$

Remark that, the proof of Theorem 19 is a direct consequence of the decomposition

$$\begin{aligned}\mathbb{G}_n(C,x) - \mathbb{E}(\mathbb{G}_n(C,x)) &= \frac{1}{\mathbb{E}(\widehat{f_n}(x))}\left[\widehat{F_n}(C,x) - \mathbb{E}\left(\widehat{F_n}(C,x)\right)\right] \\ &\quad - \frac{\mathbb{G}_n(C,x)}{\mathbb{E}(\widehat{f_n}(x))}\left[\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n}(x))\right],\end{aligned}$$

where

$$\begin{aligned}\widehat{F_n}(C,x) &= \frac{1}{n\phi(h_n)}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i\in C\}} K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right), \\ \widehat{f_n}(x) &= \frac{1}{n\phi(h_n)}\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right).\end{aligned}$$

Puting $\Delta_i(x) = K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)$. We have

$$\begin{aligned}\widehat{F_n}(C,x) &= \frac{1}{n\phi(h_n)}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i\in C\}}\Delta_i(x), \\ \widehat{f_n}(x) &= \frac{1}{n\phi(h_n)}\sum_{i=1}^{n} \Delta_i(x).\end{aligned}$$

From now for $x \in \mathcal{E}$, set

$$\mathbb{E}(\widehat{F_n}(C,x)) = \frac{1}{n\mathbb{E}(\Delta_1(x))}\sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{Y_i\in C\}}\Delta_i(x) \mid \mathcal{F}_{i-1}\right),$$

and

$$\mathbb{E}(\widehat{f_n}(x)) = \frac{1}{n\mathbb{E}(\Delta_1(x))}\sum_{i=1}^{n} \mathbb{E}(\Delta_i(x) \mid \mathcal{F}_{i-1}),$$

where $\mathbb{E}(X \mid \mathcal{F})$ is the conditional expectation of the random variables X given the σ-field $\mathcal{F}$. Lemmas 3 and 4 are important steps towards Theorem 19, for which the proofs are given in the Appendix.

Lemma 3. *Suppose that the hypotheses **(H1)-(H7)** hold and for every fixed $C \in \mathscr{C}$ as $n \to \infty$ we have :*

$$\sup_{C\in\mathscr{C}} \left|\widehat{F_n}(C,x) - \mathbb{E}\left(\widehat{F_n}(C,x)\right)\right| = o_{\mathbb{P}}(1).$$

Lemma 4. *Suppose that the hypotheses **(H1)-(H7)** hold and for every fixed $N_{\mathcal{E}}$ neighborhood of x in the functional space $\mathcal{E}$ as $n \to \infty$, we have*

$$\sup_{x\in N_{\mathcal{E}}} \left|\widehat{f_n}(x) - \mathbb{E}\left(\widehat{f_n}(x)\right)\right| = o_{\mathbb{P}}(1).$$

To establish the asymptotic normality define the "bias" term by

$$B_n(x) = \frac{\mathbb{E}\left(\widehat{f_n}(x)\right) - \mathbb{G}_n(C,x)\mathbb{E}\left(\widehat{F_n}(C,x)\right)}{\mathbb{E}\left(\widehat{F_n}(C,x)\right)} = M_n(x) - \mathbb{G}_n(C,x), \tag{5.5}$$

where

$$M_n(x) = \frac{\mathbb{E}\left(\widehat{f_n}(x)\right)}{\mathbb{E}\left(\widehat{F_n}(C,x)\right)}.$$

By stationarity of order one of the (X_i)'s, we have

$$\mathbb{E}(\widehat{f_n}(x)) = 1. \tag{5.6}$$

The following result gives the weak convergence. Keep in mind that $f_1(x)$ is given in **(H1)**.

Theorem 20 (Asymptotic normality)**.** *Let **(H1)-(H7)** hold. Then as $n \to \infty$, for $m \geq 1$ and $C_1, \ldots, C_m \in \mathscr{C}$,*

$$\{\widetilde{\nu}_n(C_i, x)_{i=1,\ldots,m}\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where $\Sigma = \sigma_{ij}(x), i, j = 1, \ldots, m$ and

$$\sigma_{ij}(x) = \frac{\mathfrak{C}_2}{\mathfrak{C}_1^2 f_1(x)} W_2(x),$$

whenever $f_1(x) > 0$ and

$$\mathfrak{C}_1 = k(1) - \int_0^1 K'(u)\tau_0(u)d(u), \;\; \mathfrak{C}_2 = K^2(1) - \int_0^1 (K^2)'(u)\tau_0(u)du.$$

To establish the density of the process, we need to introduce the following function which provides the information on the asymptotic behaviour of the modulus of continuity

$$\Lambda_\gamma(\sigma^2, n) = \begin{cases} \sqrt{\sigma^2 \log \dfrac{1}{\sigma^2}}, & \text{if} \quad \gamma = 0; \\ \max\left((\sigma^2)^{(1-\gamma)/2}, n\phi(h_n)^{(3\gamma-1)/(2(3\gamma+1))}\right), & \text{if} \quad \gamma > 0. \end{cases}$$

Theorem 21. *Suppose that **(H1)-(H7)** hold. For each $\sigma^2 > 0$, let $\mathscr{C}_\sigma \subset \mathscr{C}$ be a class of measurable sets with*

$$\sup_{C \in \mathscr{C}_\sigma} \mathbb{G}(C,x) \leq \sigma^2 \leq 1,$$

and suppose that $\mathscr{C}$ fulfils (5.3) with $\gamma \geq 0$. Further, we assume that $\phi(h_n) \to 0$ and $n\phi(h_n) \to +\infty$ as $n \to +\infty$, such that

$$n\phi(h_n) \leq \left(\Lambda_\gamma(\sigma^2, n)\right)^2,$$

and as $n \to +\infty$, we have

$$\frac{n\phi\left(\sigma^2 \log\left(\frac{1}{\sigma^2}\right)\right)^{1+\gamma}}{\log(n)} \to \infty.$$

Further we assume that $\sigma^2 \geq h^2$. For $\gamma > 0$ and $d = 1, 2$, the later has to be replaced by $\sigma^2 \geq \phi(h_n) \log\left(\frac{1}{\phi(h_n)}\right)$, then under conditions of Theorem 20 we have the process:

$$\{\widetilde{\nu}_n(C, x) : C \in \mathscr{C}\},$$

converges in law to a Gaussian process $\{\widetilde{\nu}(C, x) : C \in \mathscr{C}\}$, that admits a version with uniformly bounded and uniformly continuous paths with respect to $\|\cdot\|_2-$norm with covariance $\sigma_{ij}(x)$ given in Theorem 20.

Remark 11. *Central limit theorems are usually used to establish confidence intervals for the target to be estimated. In the context of non-parametric estimation the asymptotic variance $\Sigma(x) := \sigma_{i,j}(x)$ in the central limit depends on certain functions only approximate confidence intervals can be obtained in practice, even when $\Sigma(x)$ functionally specified. observing now in (20) that the limiting variance contains the unknown function f_1 and that the normalization depends on the function $\phi(\cdot)$ which is not identifiable explicitly. Moreover, we have to estimate the quantities W_2 and τ_0 the corollary below is a slight modification of (20) allows to have usable form of our results in practice as usually the conditional variance $W_2(x)$ is estimated by*

$$\begin{aligned} W_{2,n} &= \frac{\sum_{i=1}^{n}(1\!\!1_{\{Y_i \in C\}} - \mathbb{G}_n(x))^2 K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)} \\ &= \frac{\sum_{i=1}^{n}(1\!\!1_{\{Y_i \in C\}} - \mathbb{G}_n(x))^2 K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h}\right)} - (\mathbb{G}_n(x))^2 \\ &= \hat{g_n}(x) - (\mathbb{G}_n(x))^2. \end{aligned}$$

Let us introduce the following estimate

$$\mathcal{F}_{x,n}(t) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{\{d(x,X_i) \leq t\}}.$$

Making use the decomposition of $\tau_0(\cdot)$ in (H1)(i) one may estimate $\tau_0(\cdot)$ by

$$\tau_n(t) = \frac{\mathcal{F}_{x,n}(th)}{\mathcal{F}_{x,n}(h)}.$$

Subsequently, for a given kernel $K(\cdot)$ and the quantities $\mathfrak{C}_1$ and $\mathfrak{C}_2$ can be estimated as *follows*

$$\mathfrak{C}_{1,n} = K(1) - \int_0^1 K^{'}(s)\tau_n(s)ds, \ \mathfrak{C}_{2,n} = K^2(1) - \int_0^1 (K^2)^{'}(s)\tau_n(s)ds.$$

Introduce now some further conditions needed to state

(H8) **(i)** *The conditional mean of* $\mathbb{1}_{\{Y_i^2 \in C\}}$ *given the* σ*-field* $\mathfrak{G}_{i-1}$ *depends only on* X_i*, i.e., there exist a function* g *such that for any* $i \geq 1$, $\mathbb{E}(\mathbb{1}_{\{Y_i^2 \in C\}} \mid \mathfrak{G}_{i-1}) = g(X_i)$ *almost surely,*

(ii) *The conditional variance of* $\mathbb{1}_{\{Y_i^2 \in C\}}$ *given* $\mathfrak{G}_{i-1}$ *depends only on* X_i *i.e., for any* $i \geq 1$ $\mathbb{E}\left((\mathbb{1}_{\{Y_i^2 \in C\}})^2 \mid \mathfrak{G}_{i-1}\right) = U(X_i)$ *almost surely, for some function* U*. Moreover, the function* U *is continuous in a neighborhood of* x*, that is*

$$\sup_{u:d(x,u)\leq h} |U(u) - U(x)| = o(1).$$

Corollary 5.3.1. *Assume that conditions* **(H1)**-**(H8)** *hold true* K' *and* $(K^2)'$ *are integrable functions and* $n\mathcal{F}_x(h) \longrightarrow \infty$ *and* $h^{\beta}(n\mathcal{F}_x(h))^{1/2} \longrightarrow 0$ *as* $n \to \infty$*. Then, for any* $x \in \mathcal{E}$ *such that* $f_1(x) > 0$*, we have*

$$\frac{\mathfrak{C}_{1,n}}{\sqrt{\mathfrak{C}_{2,n}}}\sqrt{\frac{n\mathcal{F}_{x,n}(h_n)}{W_{2,n}(x)}}\left(\mathbb{G}_n(C,x) - \mathbb{G}(C,x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Using Corollary (5.3.1) the asymptotic $100(1-\alpha)$confidence band given by

$$\left[\mathbb{G}_n(C,x) - c_\alpha \frac{\mathfrak{C}_{1,n}}{\sqrt{\mathfrak{C}_{2,n}}}\sqrt{\frac{W_{2,n}(x)}{n\mathcal{F}_{x,n}(h)}}, \mathbb{G}_n(C,x) + c_\alpha \frac{\mathfrak{C}_{1,n}}{\sqrt{\mathfrak{C}_{2,n}}}\sqrt{\frac{W_{2,n}(x)}{n\mathcal{F}_{x,n}(h)}}\right].$$

5.3.1 The bandwidth selection criterion

Many methods have been established and developed to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators especially for Nadaraya-Watson regression estimator we quote among them Hall (1984), Härdle (1985), Rachdi and Vieu (2007), Dony and Mason (2008), Bouzebda and El-hadjali (2020) and Bouzebda (2020). This parameter has to be selected suitably, either in the standard finite dimensional case, or in the infinite dimensional framework for insuring good practical performances. Let us define the leave-out-(X_i, Y_i) estimator for regression function

$$\mathbb{G}_{n,j}(C,x) = \frac{\sum_{i=1, i\neq j}^{n} \mathbb{1}_{\{Y_i \in C\}} K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}. \tag{5.7}$$

In order to minimize the quadratic loss function, we introduce the following criterion, we have for some (known) non-negative weight function $\mathcal{W}(\cdot)$:

$$CV\left(C,h\right):=\frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{1}_{\{Y_j\in C\}}-\mathbb{G}_{n,j}(C,X_j)\right)^2\mathcal{W}\left(X_j\right). \tag{5.8}$$

Following the ideas developed by Rachdi and Vieu (2007), a natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose $\widehat{h}_n\in[a_n,b_n]$ minimizing among $h\in[a_n,b_n]$:

$$\sup_{C\in\mathscr{C}} CV\left(\Psi,h\right).$$

The main interest of our results is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable, like in the last equation. One can replace (5.8) by

$$CV\left(C,h_n\right):=\frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{1}_{\{Y_j\in C\}}-\mathbb{G}_{n,j}(C,X_j)\right)^2\widehat{\mathcal{W}}\left(X_j,x\right). \tag{5.9}$$

In practice, one takes, for $j=1,\ldots,n$, the uniform global weights $\mathcal{W}\left(X_j\right)=1$, and the local weights

$$\widehat{W}(X_j,x)=\begin{cases}1 & \text{if} \quad d(X_j,x)\leq h_n,\\ 0 & \text{otherwise.}\end{cases}$$

For sake of brevity, we have just considered the most popular method, that is, the cross-validated selected bandwidth. This may be extended to any other bandwidth selector such the bandwidth based on Bayesian ideas Shang (2014).

5.4 Testing the independence

we consider a sample of random elements $(X_1,Y_{1,1},Y_{1,2}),\ldots,(X_n,Y_{n,1},Y_{n,2})$ copies of (X,Y_1,Y_2) that takes its value in a space $\mathcal{E}\times\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$ and define, for $(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2$,

$$\mathbb{G}_n(C_1\times C_2,x)=\frac{\sum_{i=1}^{n}\mathbb{1}_{\{Y_{i,1}\in C_1\}}\mathbb{1}_{\{Y_{i,2}\in C_2\}}K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}{\sum_{i=1}^{n}K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}, \tag{5.10}$$

$$\mathbb{G}_{n,1}(C_1,x)=\frac{\sum_{i=1}^{n}\mathbb{1}_{\{Y_{i,1}\in C_1\}}K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}{\sum_{i=1}^{n}K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}, \tag{5.11}$$

$$\mathbb{G}_{n,2}(C_2,x)=\frac{\displaystyle\sum_{i=1}^{n}\mathbb{1}_{\{Y_{i,2}\in C_2\}}K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}{\displaystyle\sum_{i=1}^{n}K\left(\frac{d_{\mathcal{E}}(x,X_i)}{h_n}\right)}. \tag{5.12}$$

We will investigate the following processes, for $(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2$,

$$\widehat{\nu}_n(C_1,C_2,x)=\sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1\times C_2,x)-\mathbb{E}(\mathbb{G}_n(C_1,x))\mathbb{E}(\mathbb{G}_n(C_2,x))\right), \tag{5.13}$$

$$\breve{\nu}_n(C_1,C_2,x)=\sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1\times C_2,x)-\mathbb{G}_{n,1}(C_1,x)\mathbb{G}_{n,2}(C_2,x)\right). \tag{5.14}$$

Notice that we have

$$\begin{aligned}\breve{\nu}_n(C_1,C_2,x) &= \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1\times C_2,x)-\mathbb{E}(\mathbb{G}_n(C_1,x))\mathbb{E}(\mathbb{G}_n(C_2,x))\right)\\ &\quad+\sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_2,x))\left(\mathbb{G}_n(C_1,x)-\mathbb{E}(\mathbb{G}_n(C_1,x))\right)\\ &\quad-\sqrt{n\phi(h_n)}(\mathbb{G}_n(C_1,x))\left(\mathbb{G}_n(C_2,x)-\mathbb{E}(\mathbb{G}_n(C_2,x))\right).\end{aligned}$$

Hence we have

$$\begin{aligned}\breve{\nu}_n(C_1,C_2,x) &\overset{d}{=} \sqrt{n\phi(h_n)}\left(\mathbb{G}_n(C_1\times C_2,x)-\mathbb{E}(\mathbb{G}_n(C_1,x))\mathbb{E}(\mathbb{G}_n(C_2,x))\right)\\ &\quad+\sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_2,x))\left(\mathbb{G}_n(C_1,x)-\mathbb{E}(\mathbb{G}_n(C_1,x))\right)\\ &\quad-\sqrt{n\phi(h_n)}\mathbb{E}(\mathbb{G}_n(C_1,x))\left(\mathbb{G}_n(C_2,x)-\mathbb{E}(\mathbb{G}_n(C_2,x))\right)\\ &= \widehat{\nu}_n(C_1,C_2,x)+\mathbb{E}(\mathbb{G}_n(C_2,x))\widetilde{\nu}_n(C_1,x)-\mathbb{E}(\mathbb{G}_n(C_1,x)) \qquad (5.15)\\ &\quad\times\widetilde{\nu}_n(C_2,x).\end{aligned}$$

Let $\{\widehat{\nu}(C_1,C_2,x):(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2\}$ be a Gaussian process. Let us introduce the following limiting process, for $(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2$,

$$\breve{\nu}(C_1,C_2,x)=\widehat{\nu}(C_1,C_2,x)+\mathbb{G}(C_2,x)\widetilde{\nu}(C_1,x)-\mathbb{G}(C_1,x)\widetilde{\nu}(C_2,x).$$

We would test the following null hypothesis

$$\mathcal{H}_0: Y_1 \ \text{ and } \ Y_2 \text{ are conditionally independent given } X=x.$$

Against the alternative

$$\mathcal{H}_1: Y_1 \ \text{ and } \ Y_2 \text{ are conditionally dependent.}$$

Statistics of independence those can be used are

$$S_{1,n} = \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2}|\widehat{\nu}_n(C_1,C_2,x)|, \tag{5.16}$$

$$S_{2,n} = \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2}|\breve{\nu}_n(C_1,C_2,x)|. \tag{5.17}$$

A combination of Theorem 21 with continuous mapping theorem we obtain the following result.

Theorem 22. *We have under condition of Theorem 21, as $n \to \infty$,*

$$S_{1,n} \to \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2} |\widehat{\nu}(C_1, C_2, x)|, \tag{5.18}$$

$$S_{2,n} \to \sup_{(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2} |\breve{\nu}(C_1, C_2, x)|. \tag{5.19}$$

5.5 Concluding remarks

In the present work, we have established the invariance principle for the conditional set-indexed empirical process formed by ergodic functional data . Our results are obtained under assumptions on the richness of the index class $\mathscr{C}$ of sets in terms of metric entropy with bracketing in the framework of ergodic variables. This paper extends the dependence setting to the cases not covered by the usual mixing structures because ideas to general dependence structure is a rather underdeveloped field. Note that the ergodic framework avoids the widely used variants to measure the dependency and our work would go well beyond the scope of the empirical process literature, recall that the theory of empirical process is useful in many applications and an application Bahadur presentation we will derive.

5.6 Appendix

This section is devoted to the proof of our results. The aforementioned notation is also used in what follows.

Proof of Lemma 3

Use finite metric entropy with inclusion, fix $\epsilon > 0$ for $C \in \mathscr{C}$. Let C_*, C^* be a bracket for C, i.e., $C_* \subset C \subset C^*$, such that

$$\mathbb{G}\left(C_* \triangle C^* \mid x\right) < \epsilon.$$

Since for $A \subset B$ we have $\mathbb{G}_n(A,x) \leq \mathbb{G}_n(B,x)$ and $\mathbb{G}(A \mid x) \leq \mathbb{G}(B \mid x)$, it follows:

$$\begin{aligned}
&\sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C,x) - \mathbb{E}\left(\mathbb{G}_n(C,x)\right)\right] \\
&\quad \leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C_*,x)\right)\right] \\
&\quad \leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C^*,x)\right)\right] \\
&\qquad + \sup_{C\in\mathscr{C}} \left[\mathbb{E}\left(\mathbb{G}_n(C^*,x)\right) - \mathbb{E}\left(\mathbb{G}_n(C_*,x)\right)\right] \\
&\quad \leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C^*,x)\right)\right] \\
&\qquad + \sup_{C\in\mathscr{C}} \mathbb{G}\left(C_* \triangle C^* \mid x\right) \\
&\quad \leq \sup_{C\in\mathscr{C}} \left[\mathbb{G}_n(C^*,x) - \mathbb{E}\left(\mathbb{G}_n(C^*,x)\right)\right] + \epsilon. \qquad (5.20)
\end{aligned}$$

An analogous lower bound holds with C^* replaced by C_*. Since the first term in the last line is a supremum over finitely sets (for fixed $\epsilon > 0$) it follows pointwise consistency of $\mathbb{G}_n(\cdot,\cdot)$ that the term is $o_{\mathbb{P}}(1)$ and hence we obtain the desired result. □

Lemma 5.6.1. *Assume that condition (**H1**(i))- (**H1**(ii))-(**H1**(iv))-(**H5**) hold true for any real numbers* $1 \leq j \leq 2+\delta$ *and* $1 \leq k \leq 2+\delta$ *with* $\delta > 0$ *as* $n \to \infty$ *we have :*

(i) $\frac{1}{\phi(h)}\mathbb{E}(\Delta_i^j(x) \mid \mathfrak{F}_{i-1}) = M_j f_{i,1}(x) + \mathcal{O}_{a.s}\left(\frac{g_{i,x}(h)}{\phi(h)}\right)$

(ii) $\frac{1}{\phi(h)}\mathbb{E}(\Delta_1^j(x)) = M_j f_1(x) + o(1)$

(iii) $\frac{1}{\phi^k(h)}(\mathbb{E}(\Delta_1(x)))^k = M_1^k f_1^k(x) + o(1)$

The reader is referred to Laib and Louani (2010) for the proof of Lemma 5.6.1. □

Proof of Lemmas 4

We shall proof that

$$\mathbb{P}\left(\left|\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n})\right| > \epsilon\right) \to 0.$$

Observe the condition in (5.6) and we use the same proof in Laib and Louani (2010) look that $\widehat{f_n}(x) - 1 = R_{1,n}(x) + R_{2,n}(x)$ where

$$R_{1,n}(x) = \frac{1}{n\mathbb{E}(\Delta_1(x))}\sum_{i=1}^{n}(\Delta_i(x) - \mathbb{E}(\Delta_i(x) \mid \mathfrak{F}_{i-1}))$$

$$\begin{aligned}
R_{2,n}(x) &= \frac{1}{n\mathbb{E}(\Delta_1(x))}\sum_{i=1}^{n}(\mathbb{E}[\Delta_i(x) \mid \mathfrak{F}_{i-1}] - \mathbb{E}(\Delta_1(x))) \\
&= \frac{1}{n\mathbb{E}(\Delta_1(x))}\sum_{i=1}^{n}\mathbb{E}[\Delta_i(x) \mid \mathfrak{F}_{i-1}] - 1
\end{aligned}$$

Combining 5.6.1 with hypothesis (H1)-(ii) and (H1)-(iii) it easy seen that $R_{2,n}(x) = o_{a.s}(1)$ as $n \to \infty$. For the first term observe that $R_{1,n}(x) = \sum_{i=1}^{n} L_{ni}(x)$, where $L_{ni}(x)$ is a triangular array of martingale differences with respect to the σ-field $\mathfrak{F}_{i-1}$ combining Burkholder Hall and Heyde (1980) and Jensen inequalities, we obtain for any $\epsilon > 0$ that exists a constant C_0 such that

$$\mathbb{P}(|R_{1,n}(x)| > \epsilon) \leq C_0 \frac{\mathbb{E}(\Delta_1^2(x))}{\epsilon^2 n(\mathbb{E}(\Delta_1(x)))^2} = \mathcal{O}\left(\frac{1}{\epsilon^2 n\phi(h)} + o(1)\right),$$

where the last equality results form (5.6.1). Since $n\phi(h) \to \infty$ as $n \to \infty$ we conclude then that $R_{1,n}(x) = o_{\mathbb{P}}(1)$. Thus the proof is complete. □

Proof of Theorem 20

We will use similar arguments to those used in the paper by Laib and Louani (2010) to prove the asymptotic normality of the process we shall use the following notation, recall the decomposition:

$$\begin{aligned}\mathbb{G}_n(C,x) - \mathbb{E}\left(\mathbb{G}_n(C,x)\right) &= \frac{1}{\mathbb{E}(\widehat{f_n}(x))}\left[\widehat{F_n}(C,x) - \mathbb{E}\left(\widehat{F_n}(C,x)\right)\right] \\ &\quad - \frac{\mathbb{G}_n(C,x)}{\mathbb{E}(\widehat{f_n}(x))}\left[\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n}(x))\right] \\ &= \frac{Q_n(x)}{\mathbb{E}(\widehat{f_n}(x))},\end{aligned}$$

where

$$Q_n(x) = \left[\widehat{F_n}(C,x) - \mathbb{E}\left(\widehat{F_n}(C,x)\right)\right] - \mathbb{G}_n(C,x)\left[\widehat{f_n}(x) - \mathbb{E}(\widehat{f_n}(x))\right].$$

Lemma 5.6.2. *Assume that the hypotheses **(H1)-(H7)** are satisfied, then we have for any $x \in \mathcal{E}$ such that $f_1(x) > 0$, we have :*

$$\sqrt{n\phi(h_n)}Q_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x)), \qquad as \quad n \to \infty.$$

Proof of Lemma 5.6.2

Let us introduce some notation. Set

$$\eta_{ni} = \left(\frac{\phi(h)}{n}\right)^{1/2} (\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(x)) \frac{\Delta_i(x)}{\mathbb{E}(\Delta_1(x))}, \tag{5.21}$$

and define $\xi_{ni} = \eta_{ni} - \mathbb{E}\left(\eta_{ni} \mid \mathfrak{F}_{i-1}\right)$. It is easily seen that

$$(n\phi(h))^{1/2} Q_n(x) = \sum_{i=1}^{n} \xi_{ni}, \tag{5.22}$$

where for any fixed $x \in \mathcal{E}$ the summands (5.22) form a triangular array of stationary martingale differences with respect to the σ-field $\mathfrak{F}_{i-1}$. This allows us to apply the central limit theorem for discrete-time arrays of real-valued martingales (see, Györfi *et al.* (1998) page 23) to establish the asymptotic normality of $Q_n(x)$. This can be done if we establish the following statements:

$$(a)\sum_{i=1}^{n}\mathbb{E}\left(\xi_{ni}^2 \mid \mathfrak{F}_{i-1}\right) \longrightarrow \sigma^2(x),$$

and

$$(b) n\mathbb{E}\left(\xi_{ni}^2 \mathbb{1}_{|\eta_{ni}|>\epsilon}\right) = o(1),$$

holds for any $\epsilon > 0$ (Lindeberg condition).

Proof of Part (a). Observe first that

$$\left|\sum_{i=1}^{n}\mathbb{E}\left(\eta_{ni}^2 \mid \mathfrak{F}_{i-1}\right) - \sum_{i=1}^{n}\mathbb{E}\left(\xi_{ni}^2 \mid \mathfrak{F}_{i-1}\right)\right| \le \sum_{i=1}^{n}\left(\mathbb{E}\left(\eta_{ni} \mid \mathfrak{F}_{i-1}\right)\right)^2.$$

Making use of the condition (H2) and Lemma 5.6.1, one has

$$\begin{aligned}
\mathbb{E}\left(\eta_{ni} \mid \mathfrak{F}_{i-1}\right) &= \frac{1}{\mathbb{E}(\Delta_i)}\left(\frac{\phi(h)}{n}\right)^{1/2} \left|\mathbb{E}\left((\mathbb{G}(X_i)-\mathbb{G}(x))\Delta_i(x) \mid \mathfrak{F}_{i-1}\right)\right| \\
&\le \frac{1}{\mathbb{E}(\Delta_i)}\left(\frac{\phi(h)}{n}\right)^{1/2} \sup_{u\in B(x,h)} |\mathbb{G}(X_i)-\mathbb{G}(x)|\mathbb{E}\left(\Delta_i(x) \mid \mathfrak{F}_{i-1}\right) \\
&= \mathcal{O}(h^\beta)\left(\frac{\phi(h)}{n}\right)^{1/2}\left(\frac{f_{i,1}(x)}{f_1(x)} + \mathcal{O}_{a.s}\left(\frac{g_{i,x}(h)}{\phi(h)}\right)\right). \qquad (5.23)
\end{aligned}$$

Thus, by (H1)(ii)–(iii), we have

$$\begin{aligned}
\sum_{i=1}^{n}\left(\mathbb{E}\left(\eta_{ni} \mid \mathfrak{F}_{i-1}\right)\right)^2 &= \mathcal{O}(h^{2\beta})\left(\frac{\phi(h)}{n}\right)\sum_{i=1}^{n}\left(\frac{f_{i,1}(x)}{f_1(x)} + \mathcal{O}_{a.s}\left(\frac{g_{i,x}(h)}{\phi(h)}\right)\right)^2 \\
&= \mathcal{O}(h^{2\beta}\phi(h))\left(\frac{1}{f_1^2(x)}\frac{1}{n}\sum_{i=1}^{n} f_{i,1}^2(x) + o_{a.s}(1)\right) \\
&= \mathcal{O}_{a.s}(\phi(h)h^{2\beta}). \qquad (5.24)
\end{aligned}$$

The statement (a) follows then if we show that

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left(\eta_{ni}^2 \mid \mathfrak{F}_{i-1}\right) = \sigma^2. \qquad (5.25)$$

To prove (5.25), observe that

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left(\eta_{ni}^2 \mid \mathfrak{F}_{i-1}\right) = \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2}\sum_{i=1}^{n}\mathbb{E}\left[(\mathbb{1}_{\{Y_i\in C\}} - \mathbb{G}(x))^2\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right] = J_{1n}+J_{2n},$$

where

$$\begin{aligned}
J_{1n} &= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2}\sum_{i=1}^{n}\mathbb{E}\left[\Delta_i^2(x)\mathbb{E}\left(\mathbb{1}_{\{Y_i\in C\}} - \mathbb{G}(x)\right)^2 \mid \mathfrak{G}_{i-1} \mid \mathfrak{F}_{i-1}\right] \\
&= \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2}\sum_{i=1}^{n}\mathbb{E}\left[W_2(X_i)\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right], \qquad (5.26)
\end{aligned}$$

than, we have

$$J_{2n} = \frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2} \sum_{i=1}^{n} \mathbb{E}\left[(\mathbb{G}(X_i) - \mathbb{G}(X))^2 \Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right].$$

we give now an upper bound for

$$\mathbb{E}\left[W_2(X_i)\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right].$$

Towards this end, we split it up into

$$I_{n1} + I_{n2},$$

with

$$I_{n1} = W_2(x)\mathbb{E}(\Delta_i^2(x) \mid \mathfrak{F}_{i-1})$$

and

$$I_{n2} = \mathbb{E}\left[(W_2(X_i) - W_2(x))\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right]$$

Making use of (H2)-(ii), one can write

$$|I_{n2}| \leq \sup_{u:d(x;u)\leq h} |W_2(u) - W_2(x))|\mathbb{E}\left[\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right]| = \mathbb{E}\left[\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right] \times o(1)$$

Thus, in view of 5.6.1 part (i), we have

$$\begin{aligned} \mathbb{E}\left[W_2(X_i)\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right] &= (o(1) + W_2(x))\mathbb{E}\left(\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right) \\ &= (o(1) + W_2(x))(M_2\phi(h)f_{i,1}(x) + O_{a.s}(g_{i.x}(h))). \quad (5.27) \end{aligned}$$

Combining again 5.6.1 and conditions (H1)(ii)–(iii), it is easily seen that $\lim_{n\to\infty} J_{1n} = \frac{M_2}{M_1^2}\frac{W_2(x)}{f_1(x)}$ almost surely, whenever $f_1(x) > 0$. Consider now the term J_{2n}. Making use of conditions (H1)(ii)–(iii) and (H2)-(i) and Lemma 5.6.1, one can write

$$\begin{aligned} |J_{n2}| &= \mathcal{O}(h^{2\beta})\frac{\phi(h)}{n(\mathbb{E}(\Delta_1(x)))^2}\sum_{i=1}^{n}\mathbb{E}\left(\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right) \\ &= \mathcal{O}(h^{2\beta})\left(\frac{M_2}{M_1^2}\frac{1}{f_1(x)} + o_{a.s}(1)\right) \to 0 \quad \textit{almost} \quad \textit{surely} \quad \textit{as} n \to \infty. \quad (5.28) \end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left(\Delta_i^2(x) \mid \mathfrak{F}_{i-1}\right) = \lim_{n\to\infty}(J_{n1} + J_{n2}) = \frac{M_2}{M_1^2}\frac{W_2(x)}{f_1(x)} =: \sigma^2(x) \quad \textit{almost} \quad \textit{surely}$$

whenever $f_1(x) > 0$, this completes the proof of Part (a).

Proof of Part (b). the Lindeberg condition results from Corollary 9.5.2 in (Chow and

Teicher (1997)) which implies that

$$n\mathbb{E}(\xi_{ni}^2 \mathbb{1}(|\xi_{ni}| > \varepsilon)) \leq 4n\mathbb{E}\left(\eta_{ni}^2 \mathbb{1}(|\eta_{ni}| > \varepsilon/2)\right).$$

Let $a > 1$ and $b > 1$ such that $\frac{1}{a} + \frac{1}{b} = 1$. Making use of Hölder and Markov inequalities one can write, for all $\varepsilon > 0$.

$$\mathbb{E}\left(\eta_{ni}^2 \mathbb{1}(|\eta_{ni}| > \varepsilon/2)\right) \leq \frac{\mathbb{E}|\eta_{ni}|^{2a}}{(\varepsilon/2)^{2a/b}},$$

taking C_0 a positive constant and $2a = 2+\delta$ (with δ as in (H1)) we obtain

$$\begin{aligned}
&4n\mathbb{E}\left(\eta_{ni}^2 \mathbb{1}(|\eta_{ni}| > \varepsilon/2)\right)\\
\leq\ & C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \mathbb{E}([|\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(x)|\Delta_i(x)]^{2+\delta})\\
\leq\ & C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \mathbb{E}\left(\mathbb{E}\left(|\mathbb{1}_{\{Y_i \in C\}} - \mathbb{G}(x)|^{2+\delta}(\Delta_i(x))^{2+\delta} \mid X_i\right)\right)\\
\leq\ & C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \mathbb{E}\left((\Delta_i(x))^{2+\delta}\overline{W}_{2+\delta}(X_i)\right)\\
\leq\ & C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n}{(\mathbb{E}(\Delta_1(x)))^{2+\delta}} \left(\mathbb{E}\left((\Delta_i(x))^{2+\delta}|\overline{W}_{2+\delta}(X_i) - \overline{W}_{2+\delta}(x)\right)\right.\\
& \left. +|\overline{W}_{2+\delta}(x)|\mathbb{E}\left[(\Delta_i(x))^{2+\delta}\right]\right)\\
\leq\ & C_0 \left(\frac{\phi(h)}{n}\right)^{(2+\delta)/2} \frac{n\mathbb{E}\left[(\Delta_1(x))^{2+\delta}\right]}{\mathbb{E}\left(\Delta_1(x)\right)^{2+\delta}} \left(|\overline{W}_{2+\delta}(x)| + o(1)\right)\\
\leq\ & C_0 (n\phi(h))^{-\delta/2} \frac{(M_{2+\delta} f_1(x) + o(1))}{(M_1^{2+\delta} f_1^{2+\delta}(x) + o(1))} \left(|\overline{W}_{2+\delta}(x)| + o(1)\right)\\
=\ & \mathcal{O}((n\phi(h))^{-\delta/2}),
\end{aligned} \tag{5.29}$$

where the last equality follows from Lemma 5.6.1. This completes the Proof of part (b), since $n\phi(h) \to \infty$ as $n \to \infty$. Thus the proof is complete. □

Proof of Theorem 21

Let us recall some facts. Let $f(\cdot) = \mathbb{1}\{\cdot \in C_1\}$ and $g(\cdot) = \mathbb{1}\{\cdot \in C_2\}$. Given random measures μ_n on $(\mathbf{X}, \mathcal{X})$, we define

$$d_{\mu_n}^{(2)}(f, g) := \left[\mu_n(f-g)^2\right]^{1/2}.$$

Say that a class of functions $\mathcal{F}$ has uniformly integrable entropy with respect to $\mathcal{L}_2$-norm if

$$\int_0^\infty \sup_{\gamma \in M(\mathbf{X}, F)} \left[\ln N\left(\epsilon\left[\gamma\left(F^2\right)\right]^{1/2}, \mathcal{F}, d_\gamma^{(2)}\right)\right]^{1/2} d\epsilon < \infty,$$

where $d_\gamma^{(2)}(f,g) := [\int_{\mathbf{X}}(f-g)^2 d\gamma]^{1/2}$. When the class $\mathcal{F}$ has uniformly integrable entropy, $\left(\mathcal{F}, d_\gamma^{(2)}\right)$ is totally bounded for any measure γ. Let κ be an envelope of $\mathscr{F}$. That is, κ a measurable function mapping $\mathscr{F}$ to $[0,\infty)$ such that

$$\sup_{f\in\mathscr{F}}|f(t)|\leq\kappa(t),\quad\text{for all}\quad t\in\mathbb{R}.$$

Let $M(\mathbb{R},\kappa)$ be the set of all measures γ on $(\mathbb{R},\mathscr{F})$ with

$$\gamma(\kappa):=\int_{\mathbb{R}}\kappa^2 d\gamma<\infty, \tag{5.30}$$

and

$$d_\gamma^{(r)}(f,g):=\left[\int_{\mathbb{R}}(f-g)^r d\gamma\right]^{1/r}.$$

Given random measures μ_n on $(\mathbb{R},\mathscr{F})$, we define

$$d_{\mu_n}^{(2)}(f,g):=[\mu_n(f-g)^2]^{1/2}.$$

Let us introduce the uniform entropy integral

$$J(\delta,\mathscr{F},d_\gamma^{(2)})=\int_0^\delta\sup_{\gamma\in(\mathbb{R},\mathscr{F})}\left[\log\left(\mathcal{N}\left(\epsilon[\gamma(\kappa^2)]^{1/2},\mathscr{F},d_\gamma^{(2)}\right)\right)\right]^{1/2}d\epsilon.$$

We say that $\mathscr{F}$ has uniformly integrable entropy with respect to L_2-norm if

$$J\left(\infty,\mathscr{F},d_\gamma^{(2)}\right)<\infty. \tag{5.31}$$

When the class $\mathscr{F}$ has uniformly integrable entropy, $\left(\mathscr{F},d_\gamma^{(2)}\right)$ is totally bounded for any measure γ. Let $\{\mathbb{B}(\varphi):\varphi\in\mathscr{F}\}$ be a Gaussian process whose sample paths are contained in

$$U_b(\mathscr{F},d_\gamma^{(2)}):=\left\{f\in\ell^\infty(\mathscr{F}):f\ \text{ is uniformly continuous with respect to }\ d_\gamma^{(2)}\right\}.$$

Let $\mathcal{L}(\bullet)$ denote the law of $\bullet$. Notice that obtaining a uniform CLT essentially means that we show the following convergence

$$\left\{\mathcal{L}(A_{n,\varphi}):\varphi\in\mathscr{F}\right\}\to\left\{\mathcal{L}(\mathbb{B}(\varphi)):f\in\mathscr{F}\right\},$$

where the processes are indexed by $\mathscr{F}$ and considered as random elements of the bounded real-valued functions on $\mathscr{F}$ defined by

$$\ell^\infty(\mathscr{F}):=\left\{f:\mathscr{F}\to\mathbb{R}:\|f\|_{\mathscr{F}}:=\sup_{\varphi\in\mathscr{F}}|f(\varphi)|<\infty\right\}, \tag{5.32}$$

which is a Banach space equipped with the sup norm. In the sequel, we use the weak convergence in the sense of Hoffmann-Jø rgensen (1991) that we recall in the following definition. Throughout the paper, $\mathbb{E}^*$ denotes the upper expectation with respect to the outer probability $\mathbb{P}^*$, we refer to (Van der Vaart and Wellner, 1996, p.6) and (Kosorok, 2008, §6.2, p.88) for further details and discussion.

Definition 5.6.3. *A sequence of $\ell^\infty(\mathscr{F})$-valued random functions $\{T_n : n \geq 1\}$ converges in law to a $\ell^\infty(\mathscr{F})$-valued Borel measurable random function T whose law concentrates on a separable subset of $\ell^\infty(\mathscr{F})$, denoted $T_n \rightsquigarrow T$, if,*

$$\mathbb{E}g(T) = \lim_{n\to\infty} \mathbb{E}^* g(T_n), \quad \forall g \in C(\ell^\infty(\mathscr{F}), \|\cdot\|_{\mathscr{F}}),$$

where $C(\ell^\infty(\mathscr{F}), \|\cdot\|_{\mathscr{F}})$ is the set of all bounded $\|\cdot\|_{\mathscr{F}}$-continuous functions from $(\ell^\infty(\mathscr{F}), \|\cdot\|_{\mathscr{F}})$ into $\mathbb{R}$.

We set

$$\eta_{n;i}(f,x) := \eta_{n;i}(C_1,x) := \left(\frac{\phi(h)}{n}\right)^{1/2} \left(\mathbb{1}_{\{(Y_i \in C_1\}} - \mathbb{G}(C,x)\right) \frac{\Delta_i(x)}{\mathbb{E}(\Delta_i(x))}.$$

with $\Delta_i(x) = K(h^{-1}d(x,X_i))$, and define $\eta_{n;i}(g,x)$ in a similar way. Let

$$\xi_{n;i}(f,x) := \eta_{n;i}(f,x) - \mathbb{E}(\eta_{n;i}(f,x) \mid \mathfrak{F}_{i-1}).$$

Let us define

$$\sigma_n^2(f,g) = \sum_{i=1}^n \left(\xi_{n;i}(f,x) - \xi_{n;i}(g,x)\right)^2.$$

To prove Theorem 21, using Theorem 2 of Bae *et al.* (2010), it suffices to show that, for all constant $L > 0$, as n tends to infinity, that

$$\mathbb{P}^* \left\{ \sup_{f,g\in\mathscr{F}} \frac{\sigma_n^2(f,g)}{(d_{\mu_n}^{(2)}(f,g))^2} > L \right\} \to 0, \tag{5.33}$$

which is implied by the following,

$$\mathbb{E}^* \sup_{d^{(2)}(f,g)\leq\delta_n} \sum_{i=1}^n \frac{\mathbb{E}((\xi_{n;i}(f,x) - \xi_{n;i}(g,x))^2 \mid \mathfrak{F}_{i-1})}{(d^{(2)}(f,g))^2} \to 0, \quad \text{as} \quad \delta_n \to 0,$$

where we recall

$$d^{(2)}(f,g) := \left[\int_{\mathbb{R}} (f-g)^2 d\mathbb{P}\right]^{1/2}.$$

In the rest of the proof, denote by $\beta_n(x) = \frac{\sqrt{\phi(h)}}{\mathbb{E}[\Delta_1(x)]}$, and

$$\zeta(f,x) = \zeta(C_1,x) := \left(\mathbb{1}_{\{(Y_i \in C_1\}} - \mathbb{G}(C,x)\right) \Delta_i(x).$$

Therefore, we have the following

$$\begin{aligned}
&\sum_{i=1}^{n}\frac{\mathbb{E}((\xi_{n;i}(f,x)-\xi_{n;i}(g,x))^2\mid\mathfrak{F}_{i-1})}{d^{(2)}(f,g)}\\
&= \frac{\beta_n^2(x)}{nd^{(2)}(f,g)}\sum_{i=1}^{n}\mathbb{E}\left[\Big((\zeta(f,x)-\zeta(g,x))-\mathbb{E}\left[\zeta(f,x)-\zeta(g,x)\mid\mathcal{F}_{i-1}\right]\Big)^2\mid\mathfrak{F}_{i-1}\right]^2\\
&\leq \frac{\beta_n^2(x)}{nd^{(2)}(f,g)}\sum_{i=1}^{n}2\mathbb{E}\left[\Big(\zeta(f,x)-\zeta(g,x)\Big)^2\mid\mathfrak{F}_{i-1}\right]-2\mathbb{E}\left\{\left[\mathbb{E}\left[\Big(\zeta(f,x)-\zeta(g,x)\Big)\mid\mathfrak{F}_{i-1}\right]\right]^2\right\}\\
&:= T_{1,n}+T_{2,n}.
\end{aligned}$$

We first evaluate $T_{1,n}$. We have

$$\begin{aligned}
T_{1,n} &\leq \frac{2\beta_n^2(x)}{nd^{(2)}(f,g)}\sum_{i=1}^{n}2\mathbb{E}\left[\Delta_i^2(x)\left(f(Y_i)-g(Y_i)\right)^2\mid\mathfrak{F}_{i-1}\right]\\
&\quad +2\mathbb{E}\left[\Delta_i^2(x)\left(\mathbb{G}(C_1,x)-\mathbb{G}(C_2,x)\right)^2\mid\mathfrak{F}_{i-1}\right]\\
&:= T_{1,n,1}+T_{1,n,2}.
\end{aligned}$$

Using the fact that $\mathbb{E}(\Delta_1^2(x))=\mathcal{O}(\phi(h))$ (in view of Lemma 3), the class of functions $\mathscr{F}$ admits a constant envelope and $K(\cdot)$ is bounded and bounded away from zero, one may get the following upper bound of the last equation, for some positive constant,

$$\begin{aligned}
T_{1,n,1} &\leq \frac{C\sqrt{\phi(h)}}{d^{(2)}(f,g)}\mathbb{E}\left[\Delta_1(x)\left(f(Y_1)-g(Y_1)\right)\right]\\
&\leq \frac{C\sqrt{\phi(h)}}{d^{(2)}(f,g)}\mathbb{E}\left[\Delta_1(x)^2\right]^{1/2}\mathbb{E}\left[\left(f(Y_1)-g(Y_1)\right)^2\right]^{1/2}\\
&= \frac{C\sqrt{\phi(h)}}{\overline{G}^2(\zeta)}\mathbb{E}\left[\Delta_1(x)^2\right]^{1/2}\\
&= \mathcal{O}(\phi(h))=o(1).
\end{aligned}$$

Making use of similar arguments, we infer that

$$T_{1,n,2} = \frac{C\phi(h)^{3/2}}{d^{(2)}(f,g)}\left(\mathbb{E}\left[(f(Y)-g(Y))|X=x\right]\right)^2=\mathcal{O}(\phi(h)^{3/2}=o(1).$$

We readily obtain that, $T_{1,n}=o(1)$. We have, by similar arguments to those used in the proof of the preceding statement, $T_{2,n}=o(1)$. Making use of Lindeberg conditions of the preceding proof and (5.33) combined with Theorem 1 of Bae *et al.* (2010), we obtain, for given $\varepsilon>0$ and $\gamma>0$, there exists $\eta>0$, such that

$$\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{d(C_1,C_2))\leq\eta}|\widetilde{\nu}_n(C_1,x)-\widetilde{\nu}_n(C_2,x)|\geq 5\gamma\right\}\leq 3\varepsilon. \tag{5.34}$$

Now the proof theorem is completed by combining this last equation with Theorem 21. ■

5.7 Examples

Example 6. *(Laib and Louani (2010)) On the Hilbert space $\mathcal{E}$ equipped with the norm $\|\cdot\|$ associated to the inner product $\langle\cdot,\cdot\rangle$, consider the Hilbert autoregressive model of order one defined, for $n \geq 1$, by*

$$X_n = \rho(X_{n-1}) + \epsilon_n,$$

where $(\epsilon_n)_{n\geq 1}$ is an i.i.d. sequence of Hilbert random variables such that ϵ_n is independent of X_{n-1} and $\mathbb{E}(\|\epsilon_n\|^2) < \infty$ and ρ is a functional operator on $\mathcal{E}$. For $k \in \mathbb{N}^*$, consider the semi-metric d_k defined for for any $(x, y) \in \mathcal{E}^2$, by

$$d_k(x, y) = \left(\sum_{j=1}^{k}\langle x - y, e_j\rangle^2\right)^{\frac{1}{2}}. \tag{5.35}$$

Taking the semi-metric defined in the statement (5.35), observe that

$$F_x^{\mathcal{F}_{i-1}}(u) = \mathbb{P}(d_k(x, X_i) \leq u \mid \mathcal{F}_{i-1}) = \mathbb{P}(d_k(x, \rho(X_i) + \epsilon_i) \leq u \mid \mathcal{F}_{i-1}).$$

Since we can write $\epsilon_i = \sum_{j=1}^{\infty} \epsilon_i^j e_j$ and for any $s \in \mathcal{E}$ $\rho(s) = \sum_{j=1}^{\infty}(\rho(s))_j e_j$, it follows that

$$\begin{aligned} F_{X_i|X_{i-1}=s}(u) &= \mathbb{P}(d_k(x, \rho(X_i) + \epsilon_i) \leq u \mid X_{i-1} = s) \\ &= \mathbb{P}\left(\sum_{j=1}^{k}\langle x_j - (\rho(s))_j + \epsilon_i^j, e_j\rangle^2 \leq u^2\right) \\ &= \mathbb{P}(\|\bar{\epsilon_i} - (\bar{\rho}(s)) - x\|_{Ecld} \leq u) = \mathbb{P}(\bar{\epsilon_i} \in B_k((\bar{\rho}(s)) - x, u)), \end{aligned}$$

where $\bar{\epsilon_i} = (\epsilon_i^1, \ldots, \epsilon_i^k)$, $\bar{\rho}(s) = ((\rho(s))_1, \ldots, (\rho(s))_k)$ and $B_k(\bar{\rho}(s) - x, u)$ is the ball in $\mathbb{R}^k$ of center $\bar{\rho}(s) - x$ and radius u. Denote by g the density function of $\bar{\epsilon_i}$ Clearly, we have

$$\begin{aligned} F_{X_i|X_{i-1}=s}(u) &= \int \cdots \int_{B_k(\bar{\rho}(s)-x,u)} g(t_1, \ldots, t_k)dt_1 \ldots dt_k \\ &= \int \cdots \int_{B_k(\bar{\rho}(s)-x,u)} |g(t_1, \ldots, t_k) - g(\bar{\rho}(s) - x)|dt_1 \ldots dt_k + Cu^k g(\bar{\rho}(s) - x). \end{aligned}$$

When g is assumed to be a Lipschitz function of order 1 with a constant $C > 0$, we obtain

$$F_{X_i|X_{i-1}=s}(u) = Cu^k g(\bar{\rho}(s) - x) + o(u^k).$$

Therefore,

$$F_x^{\mathcal{F}_{i-1}}(u) = F_{X_i|X_{i-1}=s}(u) = Cu^k g(\bar{\rho}(X_{i-1}) - x) + o(u^k).$$

Example 7. *(Laib and Louani (2010)) Let $\mathcal{C}$ be a separate abstract space equipped with a semi-distance. Consider the autoregressive model of order one defined, for any $i \geq 1$, by $X_i = \rho(X_{i-1}) + \epsilon_i$ where $\epsilon_i = \eta_i h$ with a real random variable η_i independent of X_{i-1} and $h \in \mathcal{C}$ and ρ is a functional operator on $\mathcal{C}$. For $(x, y) \in \mathcal{C}$ consider the semi-distance* *between x and y given by*

$$d(x,y) = \left|\int (x(t) - y(t)dt\right|.$$

Observe, for any $u > 0$, that we have

$$F_x^{\mathcal{F}_{i-1}}(u) = \mathbb{P}(d_k(x, X_i) \leq u \mid \mathcal{F}_{i-1}) = \mathbb{P}(d(x, X_i) \leq u \mid X_{i-1}).$$

Consequently, whenever $0 \neq \int h(t)dt < \infty$, we have

$$\begin{aligned}
F_{X_i|X_{i-1}=s}(u) &= \mathbb{P}(d(x, X_i) \leq u \mid X_{i-1} = s) \\
&= \mathbb{P}\left(\left|\int x(t) - X_i(t)dt\right| \leq u \mid X_{i-1} = s\right) \\
&= \mathbb{P}\left(\left|\int x(t) - \rho(X_{i-1})(t) - \eta_i h(t)dt\right| \leq u \mid X_{i-1} = s\right) \\
&= \mathbb{P}\left(\left|\int x(t) - \rho(s)(t) - \eta_i h(t)dt\right| \leq u\right) \\
&= \mathbb{P}\left(\frac{-u + \int x(t)dt - \int \rho(s)(t)dt}{\int h(t)dt} \leq \eta_i \leq \frac{u + \int x(t)dt - \int \rho(s)(t)dt}{\int h(t)dt}\right) \\
&= \Phi\left(\frac{u + \int x(t)dt - \int \rho(s)(t)dt}{\int h(t)dt}\right) - \Phi\left(\frac{-u + \int x(t)dt - \int \rho(s)(t)dt}{\int h(t)dt}\right),
\end{aligned}$$

where Φ is the cumulative distribution function of η_i. Assuming now that $0 < \int h(t)dt < \infty$, $|\int x(t)dt| < \infty$ and $|\int \rho(s)(t)dt| < \infty$ for any $s \in \mathcal{C}$ and tacking Φ as the $\mathcal{N}(0,1)$ cumulative distribution function, we obtain

$$F_{X_i|X_{i-1}=s}(u) = \frac{u}{\int h(t)dt}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}\left(\frac{\int x(t)dt - \int \rho(s)(t)dt}{\int h(t)dt}\right)^2\right)(1 + o(1)).$$

Thus

$$F_x^{\mathcal{F}_{i-1}}(u) = \frac{u}{\int h(t)dt}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}\left(\frac{\int x(t)dt - \int \rho(s)(t)dt}{\int h(t)dt}\right)^2\right)(1 + o(1)),$$

and the condition **(H1)***(ii) is satisfied with*

$$\phi(u) = \frac{u}{\int h(t)dt}\sqrt{\frac{2}{\pi}}.$$

CHAPTER 6

CONCLUSION AND PROSPECTS

6.1 Conclusion

In this thesis we were interested specifically to Nadaraya-Watson conditional empirical processes model when the covariates are functional, the model studied, concerns the set-indexed conditional empirical process and we established the asymptotic properties of the built estimator :

For our model we studied the invariance principle in which we established the weak convergence uniform and asymptotic normality with giving variance formula where we used same results obtained by Masry (2005) under some assumptions also the density where we applied the chaining method developed by Doukhan (1995) , and for the problem of choosing the optimal smoothing parameter we proposed the bandwidth selection criterion rules, we have just considered the most popular method, that is the cross-validated selected bandwidth, In addition we applied our main results for testing the conditional independence we point out that our results are the first in empirical process with functional framework, the present work extended the results of Polonik and Yao (2002) given functional mixing approach on note the main difficulty when dealing with functional variables on relies on the difficulty for choosing some appropriate measures of reference in infinite dimensional space because for the theory of empirical process many technique have been studies in the literature but functional framework no results exist, than we extend our work to the ergodic data where we use paper Laib and Louani (2010) for our contribution.

The theoretical results in this thesis will become principal reference for many further developments in functional data analysis because this contribution open lot of search work in the following section we give some prospects.

6.2 Prospects

Our work in this thesis is the first results given for conditional set-indexed empirical process when the covarite are functional so this work offers many perspectives to short and long terms to improve and extend our results in functional framework:

About the model :

One of the next work we can replace the set by an function in the model studied and we give the same results in this thesis see the model given by Poryvaĭ (2005) in the multivariate framework then we propose the following functional one:

We consider a sample of random elements $(X_1, Y_1), \ldots, (X_n, Y_n)$ copies of (X, Y) that takes its value in a space $\mathcal{E} \times \mathcal{F}$. The functional space $\mathcal{E}$ and $\mathcal{F}$ are equipped with a semi-metric $d_{\mathcal{E}}(\cdot, \cdot)$ and $d_{\mathcal{F}}(\cdot, \cdot)$ respectively, for some measurable function f in a class of functions $\mathscr{F}$, We will consider the following Nadaraya Watson-type conditional empirical distribution:

$$\mathbb{G}_n(f, x) = \frac{\sum_{i=1}^{n} f(Y_i) K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d_{\mathcal{E}}(x, X_i)}{h_n}\right)},$$

where $K(\cdot)$ is a real-valued kernel function from $[0, \infty)$ into $[0, \infty)$ and h_n is a smoothing parameter f is a measurable function, and $x \in \mathcal{E}$.Concerning the semi-metric topology defined on $\mathcal{E}$ and $\mathcal{F}$ we will use the notation

$$B(x, t) = \{x_1 \in \mathcal{E} : d_{\mathcal{E}}(x_1, x) \leq t\},$$

for the ball in $\mathcal{E}$ with center x and radius t. We denote

$$\phi_x(t) = \mathbb{P}(d_{\mathcal{E}}(x, X) \leq t) = \mathbb{P}(X \in B(x, t)),$$

which is the small ball probability function. We defined the conditional empirical process indexed by class of functions :

$$\nu_n(f, x) = \sqrt{n\phi(h_n)} \left(\mathbb{G}_n(f(y), x) - \mathbb{E}(\mathbb{G}_n(f(y), x))\right), \text{ for } f \in \mathscr{F}.$$

About the methods :

The k-nearest neighbour algorithm is among the most popular methods used in statistical pattern recognition we can apply this method to obtain same results in this thesis.

About the nature of variables:

We can extend our results for the conditional set-indexed empirical process to the case of the functional ergodic data with missing at random , also weak dependency process Doukhan and Louhichi (1999)recall also in the case of independents variables no results be studied.

About the applications :

We can also study the Bahadur representation for the conditional quantile that will extend the work given by Polonik and Yao (2002).

BIBLIOGRAPHY

Aenssler(1985). Bootstrapping empirical measures indexed by Vapnikçhervonenkis classes of sets. *Probab. Theory and Math. Statistics.* (Prokhorov et al.,Eds.) pp. 467-481. VNU Press, Utrecht.

Andrews, D. W. K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review / Revue Internationale de Statistique*, **62**(1), 119–132.

Akritas,M. and Politis,D. (2003(ed)). Recent advances and tends in nonparametric statistics. Elsever,Amsterdam.

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*, **12**(4), 1041–1067.

Andrews, D. W. K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review / Revue Internationale de Statistique*, **62**(1), 119–132.

Aneiros, G., Bongiorno, E. G., Cao, R., and Vieu, P., editors (2017). *Functional statistics and related fields. Selected papers based on the presentations at the 4th international workshop on functional and operational statistics, IWFOS, Corunna, Spain, June 15–17, 2017.* Cham: Springer.

Arcones, M. A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.*, **22**(4), 2242–2274.

Bae, J., Jun, D., and Levental, S. (2010). The uniform CLT for martingale difference arrays under the uniformly integrable entropy. *Bull. Korean Math. Soc.*, **47**(1), 39–51.

Bass, R. F. and Pyke, R. (1984). A strong law of large numbers for partial-sum processes indexed by sets. *Ann. Probab.*, **12**(1), 268–271.

Berlinet, A. Gannoun, A., Matzner-Lober,E. (1998a). Normalité asymptotique d'estimateur convergents du mode conditionnel . *La Rev.Canad. de Statist.* , **26** ,365–380.

Berlinet, A. Gannoun, A., Matzner-Lober,E. (1998a). Propriétés asymptotiques d'estimateur convergents des quantiles conditionnels. *C R Acad. Sci.,Paris, Sér.I, Math.* , **326** ,611–614.

Besse, P Ramsay, J.O.(1986). Principal components analysis of sampled functions. *Psychometrika.*,**51**(2), 285–311.

Blondin D. (2004) Lois limites uniformes et estimation non-paramétrique de la régression. *PhD Thesis.*.

Bradley R.C. (2007). Introduction to Strong Mixing Conditions. *Vols. 1, 2, and 3. Kendrick Press, Heber City (Utah)*,

Brunk, H.D. (1970). Estimation by isotonic regression. *In Nonparametric Techniques in Statistical Inference (M.L. Puri ed.)*, Cambridge Univ. Press.177–197.

Billingsly P. (1968). *Convergence of Probability Measures . Wiley, New York*.

Bongiorno, E. G., Goia, A., Salinelli, E., and Vieu, P. (2014). An overview of IWFOS'2014. In *Contributions in infinite-dimensional statistics and related topics*, pages 1–5. Esculapio, Bologna.

Bosq, D., Lecoutre, J. P.(1987). *Théorie de l'estimation fonctionnelle.,.* ECONOMICA. Paris.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition. Estimation and prediction.

Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications.

Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester; Dunod, Paris.

Bouzebda, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, **21**(3), 153–188.

Bouzebda, S. and Chaouch, M. (2022). Uniform limit theorems for a class of conditional z-estimators when covariates are functions. *J. Multivariate Anal.*, **187**, Paper No. XX, 24.

Bouzebda, S., Madani, F., and Souddi, Y. (2021). Some asymptotic properties of the conditional set-indexed empirical process based on dependent functional data. *Int. J. Math. Stat.*, **22**(3), 1–29.

Bouzebda, S. and Cherfi, M. (2012). General bootstrap for dual ϕ-divergence estimates. *J. Probab. Stat.*, pages Art. ID 834107, 33.

Bouzebda, S., Madani, F., and Souddi, Y. (2021). Some asymptotic properties of the conditional set-indexed empirical process based on dependent functional data. *Int. J. Math. Stat.*, **xx**(x), x–xx.

Bouzebda,S., Nemouchi,B.(2020). Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U-statistics involving functional data. *J. Nonparametr. Stat.*, **32**,.

Bouzebda, S. and El-hadjali, T. (2020). Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.*, **32**(4), 864–914.

Bouzebda, S. and Didi, S. (2017a). Additive regression model for stationary and ergodic continuous time processes. *Comm. Statist. Theory Methods*, **46**(5), 2454–2493.

Bouzebda, S., Didi, S., and El Hajj, L. (2015). Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: asymptotic results. *Math. Methods Statist.*, **24**(3), 163–199.

Bouzebda,S., Nemouchi,B.(2020). Weak-convergence of empirical conditional processes and conditional U-processes involving functional mixing data. Submitted.

Bouzebda, S. and Nezzal, A. (2021). Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data. *Submitted.*

Bouzebda, S., Chaouch, M., and Laïb, N. (2016). Limiting law results for a class of conditional mode estimates for functional stationary ergodic data. *Math. Methods Statist.*, **25**(3), 168–195.

Bouzebda, S., Papamichail, C., and Limnios, N. (2018). On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications. *J. Nonparametr. Stat.*, **30**(1), 49–86.

Bouzebda, S. and Didi, S. (2017b). Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: asymptotic results. *Comm. Statist. Theory Methods*, **46**(3), 1367–1406.

Bouzebda, S. and Didi, S. (2021). Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes. *Rev. Mat. Complut.*, **34**(3), 811–852.

Cantelli F. P.(1933). Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist. Ital. Attuari*,**4**,221–424.

Carbon, M. and Tran, L. T. (1996). On histograms for linear processes. *J. Statist.Plann. Inference* .,**53 (3)**,403-419..

Chow, Y. S. and Teicher, H. (1997). *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, third edition. Independence, interchangeability, martingales.

Chu, C.-K. et Marron, J. S. (1991). Choosing a kernel regression estimator. *Statistical Science*,**6.4**,404–436.

Clarkson,D.B.,Fraley,C and Ramsay,J.S (2005). S+ functional data analysis user's guide. Springer Series in Statistics. Springer-Verlag, New York.

Csörgo, M. and Révész, P. (1978) Strong approximations of the quantile process. *Annals of Statistics*, **6**, 882–894.

Collomb, G. (1981). Estimation non-paramétrique de la régression. *: Revue bibliographique Inter. Statist. Review.*, **49**, 75–93.

Collomb, G. (1983). Méthodes non paramétriques en régression, analyse de séries temporelles, prédiction et discrimination. *Doctorat d'état, Toulouse.*

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. W. Gebiete.*, **66**, 441–460.

Collomb, G. (1985) Nonparametric regression *: an up to date bibliography Statistics*,**2**,297–307.

Collomb, G. et Härdle, W. and Hassani,S(1987) A note on prediction via conditional mode estimation. *J.Statist. Plann. and Inf*,**15** ,227–236.

Darling, D.-A. (1957) The Kolmogorov-Smirnov, Crame'r-von Mises tests. *Ann. Math. Statist.*, **28**,823–838.

Dawid, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B*, **41**(1), 1–31.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.*, **8**(3), 598–617.

Deheuvels, P. et Mason, D. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Prob.*,**20**,1248–1287.

Dehling, H., Denker, M., and Phillip , W. (1987). The almost sure invariance principle for the empirical process of U-statistic structure. *Ann. Inst. H. Poincure Probab. Statist.*, **23**, 121–134.

Demongeot J.,Naceri A., Laksaci A, et al.(2017). Local linear regression modelization when all variables are curves. *Statist Probab Lett.* , **121**,37–44.

Dereich, S. (2003) High resolution coding of stochastic processes and small ball probabilities. *PhD Thesis.*.

Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist..*, **6 (2) No.2** , 179–191.

Dony, J. and Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, **14**(4), 1108–1133.

Donsker, M.D.(1952) Justifcation and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.*,**23**,277–281. newblock*Stoch. Proc. Appl.*, **84**,313—342.

Doukhan, P. (1994) *Mixing : Properties and Examples. Springer-Verlag. Lect. Notes Stat.*, **85**

Doukhan, P., Massart, P., and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **31**(2), 393–427.

Doukhan, P., Louhichi, S.,(1999) A new weak dependence condition and application to moment inequalities. *Stoch. Proc. Appl.*, **84**, 313–342.

Doukhan, P., Fermanian, J. D., Lang, G.,(2009) An empirical central limit theorem with applications to copulas under weak dependence. *Stat. Infer. Stoch. Process.*,**12**,65–87.

Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.*, **6**(6), 899–929 (1979).

Dudley, R. M. (1984). A course on empirical processes. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 1–142. Springer, Berlin.

Dudley, R. M. (1999). *Uniform central limit theorems*, volume 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimal character of the sample distribution function and of the multinomial estimator. *Ann. Math. Statist.*,**27**, 642–669.

Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.*, **2**(5), 291–293.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**(3), 1380–1403.

Fan, J. (1992). Design-adaptative nonparametric regression. *J. Amer. Statist.Assoc.*, **87**, 998–1004.

Fan, F. et Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability, 66. Chapman Hall, London.

Ferraty,F., Vieu,P.(2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés., *Elsevier,*. Amsterdam,330 , No.2, 139–142.

Ferraty,F., Vieu,P.(2002). The functional nonparametric model and application to spectrometric data., *Comput. Statist. and Data Anal.*,bf330, No.2, 139–142.

Ferraty,F., Vieu,P.(2004). Nonparametric models for functional data, with application in regression times series prediction and curves discrimination., *J.* Nonparametric Statist.,**16**,111–125.

Ferraty F., Laksaci A., Vieu P.(2006). Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci. Paris.* ,**340**,389–392.

Ferraty,F. Rabhi,A. Vieu,P (2005). Conditional quantiles for dependent functional data with application to the climatic ElNiño phenomenon . *Sankhayā* , **67** (2),378 –398.

Ferraty F, Van Keilegom I, Vieu P. (2012). Regression when both response and predictor are functions. *J Multivariate Anal.*,**109**,10–28.

Ferraty F., Laksaci A., Vieu P.(2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat Inference Stoch Process.* ,**9**(1),47–76.

Ferraty,F., Vieu,P.(2003). *Functional nonparametric statistics: a double infinite dimensional framework. In: M. Akritas and D. Politis (eds.) Recent advances and trends in nonparametric statistics., Elsevier,.* Amsterdam, 61–78

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis.* Springer Series in Statistics. Springer, New York. Theory and practice.

Ferraty F. and Vieu P. (2006a) *Nonparametric modelling for functional data . Springer-Verlag,New York.*

Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.*, **49**(3), 267–286.

Gänssler, P., Stute, W.(1979). Empirical Processes: A survey of results for independent and identically distributed random variables. *Ann. Probab.*, **7** , 193–243.

Galileo Galilei (1632) *Dialogue concerning the two chief world systems.* . Modern library, 2001. Translated by Stillman Drake

Gao, F. et Li, W.V. (2007) Small ball probabilities for the Slepian Gaussian fields. *Trans. Amer. Math. Soc.* , **359** , 1339–1350 (electronic).

Gasser, T. et Müller, H. G. (1979). Kernel estimation of regression functions.In Smoothing Techniques for Curve Estimation Lecture Notes in Mathematics, *Springer Verlag, Berlin.* , **757** , 23-68.

Gasser, T., Hall, P., and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60**(4), 681–691.

Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.*, **12**(4), 929–998. With discussion.

Giné, E., Zinn, J., (1986) Lectures on the central limit theorem for empirical processes, in: *Probability and Banach spaces (Zaragoza, 1985).* vol. 1221 of Lecture Notes in Math.,Springer, Berlin,50–113.

Glivenko, V. (1933) Sulla determinazione empirica della legge di probabilita. *Giorn. Ist. Ital.Attuari* ,**4**, 92–99.

Györfi, L., Härdle, W., Sarda, P. et Vieu, P. (1989). Nonparametric curve estimation for time series. *Lecture Notes in Statistics.*, 60, Springer-Verlag.

Györfi, L., Morvai, G., and Yakowitz, S. J. (1998). Limits to consistent on-line forecasting for ergodic time series. *IEEE Trans. Inform. Theory*, **44**(2), 886–892.

Hall P. (1984). Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function *Z. Wahrsch. Verw. Gebiete*, **69** , 175–196.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London.

Härdle W. and Marron, J(1985). Optimal bandwidth selection in nonparametric regression function estimation *Z. Wahrsch. Verw. Gebiete*, **13** , 1465–1481.

Härdle, W.(1990) Applied nonparametric regression. *Cambridge Univ. Press*,Cambridge,UK.

Harel M. and Puri M. L. (1987). Weak convergence of weighted multivariate empirical processes under mixing conditions. *in* "Mathematical Statistics and Probability Theory, Vol. A.Theoretical Aspects" (M. L. Puri, P. Révész, and W. Wertz, Eds.), pp. 121–141, Reidel, Dordrecht.

Hastie, T. et Loader, C. (1993). Local regression : automatic kernel carpentry. *Statistical Science,* **8.2**, 120–143

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications.* Springer Series in Statistics. Springer, New York.

Horváth, L. and Yandell, B. S. (1988). Asymptotics of conditional empirical processes. *J. Multivariate Anal.*, **26**(2), 184–206.

Hong-zhy, A. and Bing, C. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *Internat. Statist. Rev.* , **59**, 287–307.

Hoffmann-Jø rgensen, J. (1991). *Stochastic processes on Polish spaces*, volume 39 of *Various Publications Series (Aarhus).* Aarhus Universitet, Matematisk Institut, Aarhus.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Ioannides, D. A Matzner-Lober,E. (2002). Nonparamertic estimation of the conditional mode with errors-in-variables:strong consistency for mixing processes . *J Nonparameter. Stat.*, **14** ,341 –352.

Ioannides, D. A Matzner-Lober,E. (2004). A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. *J Nonparameter. Stat.*, **16** ,515–524.

Jennen-Steinmetz, C. and Gasser, T. (1988). A unifying approach to nonparametric regression estimation. *J. Amer. Statist. Assoc., 83, No.* **404** ,1084-1088.

Johnston, G. (1979). *Smooth nonparametric regression analysis.* . Inst. of Stat. Mimeo Series 1253, Ph.D. dissertation, Univ. of No. Carolina et Chapell Hill.

Johnston, G. (1982). Probabilities of maximal deviations for nonparametric regression function estimation. *J. Mult. Analysis,* **12** ,402–414.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Koul, H.L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.*, **27**,204–236.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA. Principles and techniques.

Kolmogorov,A. and Tikhomirov, V.M(1959). ε-entropy and ε-capacity. *Uspekhi Mat.Nauk.***14**Pages 3–86. (Eng. Transl. Amer. Math. Soc. Transl. Ser., **2**, Pages 277–364, (1964)).

Kuelbs,J. Li,W. (1993). Metrric entropy and small ball problem for Gaussian measures. *J.Funct.Anal.* , **116** ,133 –157.

Laib, N. and Louani, D. (2010). Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J. Multivariate Anal.*, **101**(10), 2266–2281.

Laksaci, A. and Yousfate,A. (2002). Estimation fonctionnelle de densité de l'oprateur de transition d'un processus de Morkov *à* temps discret. *C.R., Math.,Acad.Sci.Paris* , **334** ,1035–1038.

Le Cam, L. (1983). A remark on empirical measures. In *A Festschrift for Erich L. Lehmann*, Wadsworth Statist./Probab. Ser., pages 305–327. Wadsworth, Belmont, CA.

Li, W.V. and Shao, Q.M. (2001) *Gaussian processes : inequalities, small ball probabilities and applications. In :C.R. Rao and D. Shanbhag (eds.)* Stochastic processes, Theory and Methods. *Handbook of Statitics* , **19** , North-Holland, Amsterdam.

Lian, H.(2011). Convergence of functional k-nearest neighbor regression estimate with functional responses . *J Multivariate Anal.*, **5** ,31–40.

Lifshits, M.A., Linde, W. et Shi, Z. (2006) Small deviations of Riemann-Liouville processes in Lq-spaces with respect to fractal measures. *Proc. London Math. Soc.*,(3) **92**(1), 224–250.

Louani, D. and Ould-Sad,E.(1999). Asymptotic normality of kernel estimator of the conditional mode under strong mixing hypothesis. *J.Nonparametric Stat.*, **11** ,413–442.

Marcin D. (2017) An empirical almost sure central limit theorem under the weak dependence assumptions and its application to copula processes. *VOL. LXXI, NO.*,**1**,11–29.

Masry E.(2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process Appl.*,**115**(1),155–177.

Massart, P. (1988). Invariance principles for empirical processes: The weakly dependent case. Ph.D. dissertation, Univ. Paris.

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* . Lecture Notes in Statistics. 46. Springer-Verlag, Berlin.

Müller, H.-G. and Song, K.-S. (1993). Identity reproducing multivariate nonparametric regression. *J. of Mult. Analysis,* **46**,237–253.

Naceri A. (2016). Functional local linear regression model with functional response. *JPSS J Probab Stat Sci.*,**,14**(1),1–11.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory Probab. Appl.*, **9**, 141–142.

Nadaraya, E. A. (1989) Nonparametric Estimation of Probability Densities and Regression Curves. *Kluwer, Dordrecht.*

Nobel, A. and Dembo, A. (1993). A note on uniform laws of averages for dependent processes. *Statist. Probab. Lett.*, **17**(3), 169–172.

Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.*, **15**(3), 897–919.

Ould-said,E (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scand.J.Stat.*,**24**,231–2399.

Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.*,**33**,1065-1076.

Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge, second edition. Models, reasoning, and inference.

Phillip, W. (1984). Invariance principles for sums of mixing random elements and the multivariate empirical process. *In Limit Theorems in Probability and Statistics. Colloq. Math.Soc. Jdnos Bolyai 36.* North-Holland, Amsterdam.

Pollard, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A*, **33**(2), 235–248.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, volume 2 of NSF-CBMS *Regional Conference Series in Probability and Statistics.* Institute of Mathematical Statistics and American Statistical Association, Hayward, California.

Polonik, W. and Yao, Q. (2002). Set-indexed conditional empirical and quantile processes based on dependent data. *J. Multivariate Anal.*, **80**(2), 234–255.

Poryvaĭ, D. V. (2005). An invariance principle for conditional empirical processes formed by dependent random variables. *Izv. Ross. Akad. Nauk Ser. Mat.*, **69**(4), 129–148.

Priestley, M.B. et Chao, M.T. (1972) Nonparametric function fitting. *J. Royal Statist. Soc. B* , **23** ,395–436.

Pyke, R. (1968). The weak convergence of the empirical process with random sample size. *Proc. Cam. Phil. Soc.*, **64**, 155–160.

Quintela del Rio, and Vieu,Ph. (1997) Non parametric conditionnal mode estimate. *Nonparametric Statistics*,**8** ,253–266.

Rachdi, M. and Vieu, P.(2007). Nonparametric regression for functional data: automatic smoothing parameter selection *J. Statist. Plann. Inference*, **137** ,2784–2801.

Radulović, D. (1996) The bootstrap for empirical processes based on stationary observations. *Stochastic Processes and Their Applications*,**65**, 259–279.

Ramsay, J. O. et Silverman, B. W. (1997). *Functional data analysis*. Springer Series in Statistics Springer, New York.

Ramsay, J. O. et Silverman, B. W. (2002). *Applied functional data analysis.* Series in Statistics. Springer-Verlag, New York. Methods and case studies.

Ramsay, J. O. and Silverman, B. W. (2005a). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

Ramsay, J. O. and Silverman, B. W. (2005b). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

Rio, E. (2000). *Théorie asymptotique des processus aléaltoires faiblement d'ependants, vol 31 Mathematics and applications Springer-Verlag,Berlin*

Révész, P.(1976) Three theorems of multivariate empirical process. *Lecture Notes in Math*,Empirical Distributions and Processes. No. 566, 106–126. Springer-Verlag, Berlin.

Rosenblatt, M. (1972). Uniform ergodicity and strong mixing. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **24**, 79–84.

Roussas G.(1969). Nonparametric estimation of the transition distribution function of Markov process. *Ann.Math.Statit.*,**40**,1386–1400.

Rosenblatt, F. (1956) Remarks on some nonparametric estimates of a density function. , *Ann. Math. Statist.*,**38**,482–49.

Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA*,**42** ,43–47.

Roussas,G. (1991). Estimation of transiotion distribution function and its quantiles in Markov process:strong consistency and asymtotic normality. *it Nonparametric functional estimation and related topics* ,443–462.

Ruppert, D. et Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.*, **22**, 1346–1370.

Samanta,M. Thavaneswaran, A (1990) Nonparametric estimation of the conditional mode. *Comm.Statis.theory Methods*,**19**(12),4515–454.

Samanta,M.(1989). Non parametric estimation of conditional quantiles. *Statist.Proba. Letters*,**7**,407–412.

Sarda, P., Vieu, P. (2000). *Computation and Application.,*. Wiley Series in Pprobability and Statistics. Wiley, New York

Schimek, M. (2000). *Smoothing and Regression : Approaches, computation, and application,*. Ed. M.G. Schimek. Wiley Series in Probability and Statistics.

Schuster, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of points. *Annals of Mathematical Statistics,* **43.1** ,84–88.

Schwartz, S.C. (1967) Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.*,**38**,1261-1265.

Scott, D. W. (1992). *Multivariate Density Estimation* . Theory, Practice and Visualization. Wiley, New York.

Scott, Tran, L. T. (1994). Density estimation for time series by histograms, *J. Statist. Plann. Inference* ,**40 (1)**,61-79.

Sen P.K. (1971). A note on weak convergence of empirical processes for sequences of ϕ-mixing random variables. *Ann. Math. Statist.*,**42** , 2131–2133.

Shang, H. L. (2014). Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density. *J. Nonparametr. Stat.*, **26**(3), 599–615.

Shorack G. R. and Wellner J. A.(1986) *Empirical Processes with Applications to Statistics. Wiley,New York.*

Shi, J. Q. and Choi, T. (2011). *Gaussian process regression analysis for functional data*. CRC Press, Boca Raton, FL.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*,**5**,595–645.

Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann Statist.*,**10**(4),1040–1053.

Stute, W. (1986a). Conditional empirical processes. *Ann. Statist.*, **14**(2), 638–647.

Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.*, **14**(3), 891–901.

Stute, W. (1997) Nonparametric model checks for regression. *Ann. Statist.*, **25**,613–641.

Stute, W., Thies, S. and Zhu, L.X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.*, **26**, 1916–1934.

Su, L. and White,H.(2007) A consistent characteristic function-based test for conditional independence.J. Econometrics , **141**, 807–834..

Su, L. and White, H. (2008). A nonparametric Hellinger metric test for conditional independence.. *Econometric Theory*, **24**, 829–864.

Sue, J.Q. and Wei, L.J. (1991). A lack of fit test for the mean function in a generalized linear model. *J Amer. Statist. Assoc.*, **86**, 420–426..

Theodoros N. Yannis G.Y. (1997) Rates of convergence of estimate Kolmogorov entropy and the dimensionality reduction principale in regression . *The Annals of Statistics*, **25**(6) ,2493 – 2511.

Tsybakov, A. B. (2003). Introduction à l'estimation non-paramétrique. Mathématiques et Applications, 41. Springer.

Tukey, J.W. (1961). Curve as parameters, and touch estimation. *Proceedings of the 4th Symposium on Mathematics, Statistics and Probability*, Berkeley, CA, USA , 681–694.

Van de Geer, S. A. (2000). *Empirical Processes in M-Estimation. Cambridge University Press, New York.*

Van de Geer, S.(2006) *Empirical Process Theory and Applications. ETH, Zürich.*

Van der Vaart A. W., Wellner . J. A. (2000) *Weak Convergence and Empirical Processes. With Applications to Statistics.*, Springer-Verlag, New York.

Van der Vaart, A. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Stat.*, **5**, 192–203.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Vapnik, V. N. and Červonenkis, A. J. (1971). The uniform convergence of frequencies of the appearance of events to their probabilities. *Teor. Verojatnost. i Primenen.*, **16**, 264–279.

Vieu, P. (1991). Quadratic errors for nonparametric estimates under dependence. *J. Multivariate Anal.*, **39 (2)** , 324–347.

Wand, M. P. et Jones, M. C. (1995) *Kernel Smoothing.* Chapman and Hall, London.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya A* , **26**, 359–372.

Wellner, J. A. (1989). Discussion of Gill's paper "Non- and semiparametric maximum likelihood estimators and the von Mises method (part I)". *Scandinavian Journal of Statistics*, **16**, 124–127.

Withers C. S.(1975). Convergence of empirical processes of mixing rv's on [0, 1]. *Ann. Statist.*,**3**,1101–1108.

Wong, W. H. and Severini, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Annals of Statistics*, **19**, 603–632.

Yang Y., Ling NX.,Wang J., et al.(2014). Convergence rate of recursive kernel estimation of nonparametric regression with functional responses. *J Hefei Univ Technol Nat Sci.* Chinese ,**37**(12),1528–1531

Yao,F. and Lee, T.C.M.(2006). Penalised spline models for functional principal component. *analysis J.R.Stat.Soc.B.*,**86**(1), 3–25.

Yoshihara, K.-i. (1990). Conditional empirical processes defined by ϕ-mixing sequences. *Comput. Math. Appl.*, **19**(1), 149–158.

Yoshihara, K.-i. (1988). Asymptotic normality of nearest neighbor regression function estimates based on some dependent observations. *Yokohama Math. J.*, **36**, 55–68.

Youndjé,E. (1993). Estimation non paramtrique de la densité conditionnelle par la methode du noyau. *thése de Doctorat, Université de Rouen.*

Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.*, **22**(1), 94–116.

Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2011). Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, UAI'11, page 804–813, Arlington, Virginia, USA. AUAI Press.

Zhang, J.-T. (2014). *Analysis of variance for functional data*, volume 127 of *Monographs on Statistics and Applied Probability.* CRC Press, Boca Raton, FL.

Zhou, Y. Liang, H. (2000). Asymptotic normality for L^1-norm kernel estimator of conditional median under α-mixing dependence . *J. Multivarite Anal.* , **73** ,136–154.

Zhou, Y., Liu, J., and Zhu, L. (2020). Test for conditional independence with application to conditional screening. *J. Multivariate Anal.*, **175**, 104557, 18.

Ziegler K. (2003). On the asymptotic normality of kernel regression estimators of the mode in the nonparametric random design model. *J. Statist. Plann. Inf*, **115** , 123–144.

على التقارب الضعيف للعمليات التجريبية المحلية

الملخص / يركز مشروع هذه الأطروحة على مبدأ الثبات للعمليات التجريبية الشرطية من خلال إدخال مقدر نوع ناداريا واتسون عندما تكون المتغيرات المشتركة وظيفية لقد اقترحنا عملية تجريبية مشروطة مفهرسة بفئة مجموعة حيث نؤسس تناسقًا ضعيفًا وحالة مقاربة طبيعية للمقدر المقترح في ظل ظروف معينة عندما تكون المتغيرات ثابتة ومختلطة بشدة. في ما يلي نستخدم نتائجنا الرئيسية لاختبار الاستقلال الشرطي ونوسع نتائجنا لتشمل بيانات ذات خلط عام

الكلمات المفتا حية / العمليات التجريبية ، المتغيرات الوظيفية، ناداريا واتسون، خلط قوي ، خلط عام

«On the weak convergence of local empirical processes»

Abstract : The project of this thesis focuses on the principle of invariance for conditional empirical processes by introducing the Nadaraya Watson type estimator when the covariates are functional. We have proposed a conditional empirical process indexed by an ensemble class where we establish weak consistency and asymptotic normality for the proposed estimator under certain conditions when the variables are stationary and strongly mixed. In the following we use our main results to test conditional independence and we extend our results to ergodic data.

Keywords: Conditional empirical process, Nadaraya Watson, functional covarites, strong mixing, ergodic.

« Sur la convergence faible des processus empiriques locaux »

Résumé : Le projet de cette thèse se focalise le principe d'invariance pour des processus empiriques conditionnels en introduisant l'estimateur du type Nadaraya Watson lorsque les covariables sont fonctionnelles. Nous avons proposé un processus empirique conditionnel indexé par une classe d'ensemble où nous établissons la consistance faible et la normalité asymptotique pour l'estimateur proposé sous certaines conditions lorsque les variables sont stationnaires et fortement mixtes. Dans la suite nous utilisons nos principaux résultats pour tester l'indépendance conditionnelle et nous étendons nos résultats aux données ergodiques.

Mots clés : Processus empirique conditionnel, Nadaraya Watson , les covariables sont fonctionnelles, Fort mixtes, ergodiques.