République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

N° Attribué par la bibliothèque

|_|_|_|_|_|_|_|_|_|_|

Année univ.: 2020/2021

Robust nonparametric regression with stationary ergodic and censored data

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saida - Dr Moulay Tahar
Discipline : MATHEMATIQUES
Spécialité:Analyse Stoquastiques Statistique des processus et Applications.

par

Mohammed **C**hikouche **E**lias **T**aki **E**ddine [1]

Sous la direction de

Dr. R. Rouane

Soutenue le 13/07/2021 devant le jury composé de

Dr. F. Benziadi	Université Dr Tahar Moulay - Saïda	Présidente
Dr. R. Rouane	Université Dr Tahar Moulay - Saïda	Encadreur
Pr. S. Rahmani	Université Dr Tahar Moulay - Saïda	Examinatrice
Dr. M. Kadi	Université Dr Tahar Moulay - Saïda	Examinateur

[1]e-mail: eliaschikouche@gmail.com

Dedication

*All praise to **Allah**, today we fold the day's tiredness and the errand summing up between the cover of this humble work.*

I dedicate my work to:

*My great teacher and messenger, **Mohammed-peace and grace from Allah be upon him**,* who taught us the purpose of life.

***My parents**, who have been our source of inspiration and gave usstrength when we thought of giving up, who continually provide their moral, spiritual, emotional, and financial. God save them.*

My grandmother. God bless her.

*My brothers **Zaid** and **Sohaib**.*

*To my friends **Tayeb Hamlat**, **Moulay Larbi Mahdi**, **Khelifa Berkane**, and **Senouci Hachemi** and their famillies.*

Last but not least I am dedicating this to my aunts and my cousins.

All those if my pen forget them, my heart will not forgotten them.

All those who are looking glory and pride in Islam and nothing else.

Acknowledgments

In the Name of Allah, the Most Merciful, the Most Compassionate all praise be to Allah and prayers and peace be upon Mohammed His servant and messenger.

First and foremost, I thank *ALLAH* who helped me and gave me the strength, patience and courage during my years of study.

I warmly thank my supervisor **Dr. Rouane Rachida** for her incredibly careful and efficient guidance and I am greatly indebted for her understanding, patience, help, all advice and notes during my preparation of dissertation. I would like to say

"thank you for everything".

I would like also to thank the members of jury for kindly devoting their precious time to read and evaluate this research.

I owe a deep debt of gratitude to our university, laboratory of stochastic models, statistic and applications and his members for giving us an opportunity to complete this work.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my brothers for their love, understanding, prayers and continuing support to complete this research work.

Last but certainly not least, I would like to thank everybody who was important to the successful realization of this work, as well as expressing my apology that I could not mention personally one by one. To all a big thank you.

Contents

List of Abbreviations and Symbols

$\mathbb{R}^d$	**d-dimensional Euclidean space.**
$\mathbb{N}$	**Set of positive natural numbers.**
$\mathbb{R}$	**Set of real numbers.**
$\mathbb{Z}$	**Set of integer numbers.**
$\mathbb{R}^3$	**Three-dimensional Euclidean space.**
$\mathbb{P}(A)$	**Probability of event A.**
$(\Omega, \mathcal{F}, \mathbb{P})$	**Probability space.**
$i.i.d.$	**Independent and identically distributed.**
$\theta_n^\star$	**Is a point between** $\widehat{\theta}_x$ **and** θ_x.
$d(\cdot, \cdot)$	**Semi-metric.**
$\mathbb{E}(X)$	**Expected value of random variable X.**
$\mathbb{E}(X\|Y)$	**Expected value of random variable X given Y.**
$\{x_1, x_2, \ldots\}$	**The set with elements x1, x2,....**
$\mathbb{1}_A$	**Indicator function of A.**
$\widehat{\theta}$	**Estimator of parameter** θ.
$\widehat{\Psi}$	**Estimator of parameter** Ψ.
$O_{a.co.}$	**Rate of Almost complete convergence.**
$O_{a.s.}$	**Rate of Almost sure convergence.**
O_p	**Rate of convergence in probability.**
$\mathcal{N}(\mu, \sigma^2)$	**Normal distribution (mean** μ **, variance** σ^2).
$\overline{G}_n(\cdot)$	**Kaplan and Meier estimator.**
$\xrightarrow{a.s}$	**Almost sure convergence .**
$\xrightarrow{P}$	**Convergence in probability.**
$\xrightarrow{\mathcal{D}}$	**Convergence in distribution.**
$\xrightarrow{a.co.}$	**Almost complete convergence.**
$r.v$	**Random variable.**
$w.r.t.$	**With respect to.**
$B(x, r)$	**The ball with center x and radius r.**
$SLLN$	**Strong Law of Large Numbers.**
LIL	**Law of the Iterated Logarithm.**
NW	**Nadaraya Watson.**
CR	**Censoring Rate.**
MSE	**Mean Square Error .**
$SPSS$	**Statistical Package for the Social Sciences.**

General Introduction

Regression function estimation is the most important tool for addressing nonparametric prediction problems. However, it is well known that the nonparametric estimation of the regression function is highly sensitive to the presence of even a small proportion of outliers in the data. Outliers are understood to be observations that have been corrupted, incorrectly measured, mis-recorded, drawn under different conditions than those intended, or so atypical as to require separate modeling. Robust regression was introduced to solve this kind of problem.

The literature on the nonparametric robust estimation method is quite important in the statistical literature when the variable of interest is completely observed. One can refer, among others, to Huber (1964)[33] and Härdle (1984)[30] for the independent and identically distributed (i.i.d.) case, Collomb and Härdle (1986)[17] and Boente and Fraiman (1989)[8] for mixing processes. Azzedine et al. (2008)[5] established the almost complete convergence rate of the kernel-type estimator in the i.i.d. case while Cai and Roussas (1992)[11] studied its asymptotic properties under an α-mixing assumption. Crambes et al. (2008)[19] studied the same problem for a functional covariate. They established the exact asymptotic expression of the convergence rate in $\mathbb{L}^p$ norm. However, in many practical situation such as in medical follow-up or in engineering life-test study, the variable of interest may not be completely observable. This case occur when dealing with censored data. For example, in the clinical trials domain, it frequently happens that patients from the same hospital have correlated survival times due to unmeasured variables such as the quality of hospital equipment.

The regression model in presence of censored data has been studied by several authors. Cox (1972)[18] considered the linear regression model and estimated the slope via the proportional hazard model. In the general linear case, many approaches have been used, see for instance Koul et al. (1981)[40] and Koul and Stute (1998)[41]. Ren and Gu (1997)[50] proposed a type of M-estimators for the linear regression model with random design when the response observations are doubly censored. For the nonlinear regression model, Beran (1981)[6] introduced a class of nonparametric estimators for the conditional survival function in the presence of right-censoring. He proved some consistency results of these estimates, his work has extended by Dabrowska, D.M. (1987, 1989)[20][21]. Jin (2007)[36] constructed M-estimators for the regression parameters in semi-parametric linear models for censored data and established the asymptotic normality.

Notice that, the estimator of the conditional survival function could be used in order to get a consistent estimate of the regression function $m(\cdot)$ in the presence of censored data. However, the computations may be difficult practically. To overcome this drawback, Carbonez el al. (1995)[13] introduced a general nonparametric partitioning estimate of $m(\cdot)$ and proved its strong consistency. Kohler et al. (2002)[39] gave a simpler proof for kernel, nearest neighbor, least squares and penalized least squares estimates.

In all papers mentioned above when dealing with dependent data the condition of α-mixing is assumed to be fulfilled. A large class of processes satisfy this condition. Therefore, many processes fail to satisfy these conditions. However, there are still a great number of models where such assumption does not hold. For example, in some cases the first order linear autoregressive process in discrete time is not strongly mixing. In particular, the stationary process $(X_t)_{t\in\mathbb{Z}}$ checking the model AR (1) defined by $X_t = \theta X_{t-1} + \epsilon_t$, where the $(\epsilon_t)_{t\in\mathbb{Z}}$ is an innovation sequence of Bernoulli i.i.d, is not strongly mixing. However, ergodicity is conserved by taking measurable functions from an ergodic process. As the autoregressive process, above, can be represented as a linear function of the ϵ_t, it follows then that it is also ergodic. It is then necessary to consider a general larger dependency framework as is the ergodicity.

Recently Laïb and Louani (2011)[44] studied (in the case of complete data) the asymptotic properties of the regression function using functional stationary ergodic data. In the case of right censored response, Chaouch and Khardani (2015)[16] considered the conditional quantile estimation based on functional stationary ergodic data.

In this work, we present a study on the kernel smoothing estimation of the Robust regression function for right censored and stationary ergodic data, by developing the article of M.Chaouch et al (2016)[14].

This master memory falls into four chapters.

In chapter 1, we give some background and some concepts, nonparametric regression, robust nonparametric analysis, ergodic theory, we recall the preliminaries on survival models. We introduce the main functions in survival analysis: survival function, survival rate and the different forms of the risk rate etc. we also give the different models and types of censoring.

In chapter 2, we focus on the robust regression in the case where the explanatory variable is functional and the observations are completely observed. The main objective is to prove the almost complete convergence (with rate) and the asymptotic normality for the proposed estimator. This result is obtained under an ergodic stationary process assumption, without the aid of mixing conditions traditional.

In chapter 3, we consider a robust regression estimator when the interest random variable is subject to random right-censoring and assumed to be sampled from a stationary and ergodic process. The strong consistency (with rate) and the asymptotic distribution of the estimator are established under mild assumptions. Moreover, a usable confidence interval is provided which does not depend on any unknown quantity. Our results hold without assuming any type of mixing conditions and the existence of marginal and conditional densities. To prove our results we use only the martingale differences idea combined with the ergodic theory, then we avoid any additional condition on the structure of the process under study as is in the case of α-mixing one.

In chapter 4, we present numerical simulations illustrating the performance of robust estimator.

Chapter 1

Introduction and preliminaries

In this chapter we introduce some basic notions, nonparametric regression, robust nonparametric analysis, ergodic and incomplete data, which will be useful later.

1.1 Nonparametric regression

Nonparametric regression functions have been widely used in recent decades, not just in statistics, but in different fields such as medicine, signal processing, economics and biology ... The regression function is a general function that characterizes the relationship between two variables. For example, we want to know if reducing speed reduces the number of accidents on the road, does increasing study hours allow to improve the student's average, etc. The latter represents one of the first quantities that a practitioner can study when interested in explaining a variable through another. We can see this problem as follows: We have two real random variables (v.a.) Y (variable of interest / response) and X (explanatory variable / co-variable) linked by the following relation:

$$Y = m(X) + \epsilon.$$

Such that ϵ is a random variable independent of X.

When we want to describe the influence of a quantitative variable on an event, or the link between an explanatory variable X and a variable called response variable Y, Having observed X, the mean value of Y is given by a regression function:it is this function that us inform about the type of dependence there is between these two variables. This regression function is defined for all $X \in \mathbb{R}$ by:

$$m(x) = \mathbb{E}(Y \mid X = x). \tag{1.1}$$

Which is the mean of the conditional distribution of Y given $X = x$.

Note that, the main advantage of nonparametric regression is that it does not assume any specific shape for the estimator, which gives it more flexibility in practice. However, one of the main drawback of classical regression is that the estimation of the regression function is sensitive to outliers, and may be insufficient in some cases, such as when the distribution is asymmetric or multimodal. Thus we must seek alternative approaches that are sufficiently insensitive to the effects of outliers.

1.2 Robust nonparametric analysis

The robustness of a usual statistical procedure (estimation, test) is of the utmost significance in statistics. It allows to control the stability of this procedure with respect to the deviation of the model and/or observations. It should be noted that this problem was the subject of a long debate at the end of the 19th century, several scientists already had a relatively clear idea of this notion of robustness. In fact, the first mathematical work on robust estimation seems to date back to 1818 with Laplace's work in his second supplement to the analytical theory of probability. More exactly, the term "robust" was introduced in 1954 by G. E. P. Box[10]. But this concept was not recognized as a field of research until the mid-1960s. It is especially with the work of Huber P.J. (1964)[33], Hampel F.R. (1971)[28] that a coherent theory of robust statistics was developed based on min-max criteria and uses essentially convexity arguments. From another point of view other authors Huber (1973 and 1981)[34][35], Andrews (1974)[4], Krasker and Welsh (1982)[42], have developed robust automatic adjustment methods, which have the advantage of being as effective as the least squares method when there are no outliers, but more effective in the presence of atypical observations or when the error distribution in the model follows a long-tailed distribution.

The robust estimation of the regression function is a topic of great interest in nonparametric statistics. This is an area in which the first consequent results were established in the early sixties by Huber (1964)[33], of which he obtained the consistency and asymptotic normality of a class of estimators for this function. Robinson (1984)[51], Hardle (1984)[30] and Hardle and Tsybakov (1989)[31] established under mixing conditions the asymptotic normality of a weighted family of estimators derived from the kernel method for the regression function. At the same time, Boente and Fraiman (1989, 1990)[8][9] used the Robinson estimator (1984)[51] to simultaneously study the two position and scale parameters. The consistency of the constructed estimators is obtained under general conditions and in the two independent and strongly mixing cases. The uniform convergence of the robust regression estimator was obtained by Collomb and Hardle (1986)[17] by considering mixed ϕ observations. An alternative method of robust estimation of the regression function was proposed by Fan and al.(1994)[22]. This method makes it possible to encompass several nonparametric models and robustify classical regression. Laïb and Ould-Saïd (2000)[45] adapted the estimator of Collomb and Hardle (1986)[17] for the model of autoregression of a stationary ergodic process. They obtained the uniform convergence of this estimator even when the objective function is unbounded. Cai and Ould-Saïd (2003)[12] used a robust version of the local polynomial method estimation for the regression function. They demonstrated under standard conditions and when the observations are alpha mixing, the asymptotic normality and the almost sure convergence of these estimators. This paper also provides root mean square convergence result as well as the smoothing parameter optimizing this error. Recently, Ghement and al. (2008)[26] introduce a family of robust M-estimators for the dispersion parameter of which they have shown the consistency and asymptotic normality. A criterion for quantifying robustness is also proposed in this work.

1.3 Ergodic

The term "ergodic" comes from the Greek words (ergon, odos) which mean (work, path), Ludwing Boltzmann choose it while he was working on a problem in statistical mechanics. The branch of mathematics that studies ergodic systems is known as ergodic theory. The later is a fundamental hypothesis of statistical physics. more precisely, it has experienced a development in the use of dynamic systems, chaos theory as well in signal processing. Researchers as Birkhoff and Von Neumann (1931) were interested in this theory and developed two of the main theorems on this subject. We refer to Krengel's book (1985)[43] for a series of results on ergodic theory. And since then, ergodic theory has occupied a place in different branches of mathematics such as functional analysis and group theory, calculates probabilities and more precisely Markovian processes, estimation theory... .

For the sake of clarity, 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 $\mathcal{B}_n = \sigma(X_k; k \leq n)$ and the forward field $\mathcal{H}_m = \sigma(X_k; k \geq m)$. The sequence $\{X_i, i = 1, 2, \ldots\}$ is strongly mixing if

$$\sup_{A \in \mathcal{B}_0, B \in \mathcal{H}_n} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \alpha(n) \to 0 \quad \text{as} \quad n \to \infty.$$

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 a more stringent than what is ordinarily referred 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. Hence, strong mixing implies ergodicity. However, the converse is not true: there exist ergodic sequences which are not strong mixing. The ergodicity condition is then a condition which is lower than any type of mixing for which usual nonparametric estimators (density, regression, ...) are convergent. It seems to be a condition of obtaining law of large numbers, we are interested in the ergodic theorem, for a stationary ergodic process X.

Theorem 1.3.1. *[43]**(Ergodic theorem)** If $X = (X_t)_{t\in\mathbb{Z}}$ is a stationary ergodic process and if X_1 is integrable, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}(X_1), \text{ almost surely (a.s.)}$$

1.4 Incomplete data (Censored data)

One of the characteristics of survival data is the existence of incomplete observations. For instance, in epidemiological surveys, data is often collected incompletely. Censoring is part of a process that generates this type of data. It must be taken into account when writing the likelihood. We will talk about censored data when the survival time is only known when it is limited by a limited observation time.

1.4.1 Survival Analysis

Survival analysis involves the modeling of time to event data. The time to event data shows the time span from well defined time origin till the well defined end point of interest (event). For instance, in study of a particular type of cancer, the time point of diagnosis of that type of cancer is chosen to be time origin and the death due to that particular cancer would be the end point. Or a study might follow people from birth (time origin) until the occurrence of a disease(end point). Survival analysis is a very active research field for several decades. An important contribution that stimulated the entire field was the counting process formulation given by Aalen (1975)[1]. The flexibility of a counting process is that it allows modeling multiple (or recurrent) events. Since then a large number of fine text books have been written on survival analysis and counting processes, with some key references being Andersen et al.

(1993)[3], Fleming and Harrington (1991)[25], Kalbfleisch and Prentice (2002)[38], and Lawless (2003)[48].

In survival analysis, a data set can be exact or censored, and it may also be truncated. Exact data, also known as uncensored data, occurs when the precise time until the event of interest is known. Censored data arises when a subject's time until the event of interest is known only to occur in a certain period of time. For example, if an individual drops out of a clinical trial before the event of interest has occurred, then that individual's time-to-event is right censored at the time point at which the individual left the trial. The time until an event of interest is truncated if the event time occurs within a period of time that is outside of the observed time period.

1.4.2 Functions of Survival Times

For purposes of survival analysis, the functions of time are usually defined. They are survival function, density function, hazard function and cumulative hazard function from which survival and hazard functions are of particular interest. In traditionally established statistical models, density and cumulative distributions are used but due to the incomplete observations in survival data(censored and truncated data) these standard functions are not appropriate. So survival and hazard functions are considered more suitable.

The distribution of the random variable T can be described in a number of equivalent ways. There is of course the usual (cumulative) distribution function

$$F(t) = \mathbb{P}[T \leq t], \quad t \geq 0,$$

which is right continuous, i.e., $\lim_{u \to t+} F(u) = F(t)$. When T is a survival time. If T is a continuous random variable, then it has a density function $f(t)$, which is related to $F(t)$ through following equations

$$f(t) = \frac{dF(t)}{dt}, F(t) = \int_0^\infty f(u)du, \quad t > 0,$$

if the distribution function has a derivative at point t then

$$f(t) = \lim_{dt \to 0} \frac{\mathbb{P}(t \leq T < t + dt)}{dt} = F'(t) = -S'(t).$$

In biomedical applications, it is often common to use the survival function

$$S(t) = \mathbb{P}[T > t] = 1 - F(t).$$

The survival function $S(t)$ is a non-increasing function overtime taking on the value 1 at $t = 0$, i.e., $S(0) = 1$. For a proper random variable T, $S(\infty) = 0$.
The hazard rate is a useful way of describing the distribution of "time to event" because it has a natural interpretation that relates to the aging of a population. We motivate the definition of hazard rate by first defining the mortality rate, which is a discrete version of the hazard rate. The hazard rate $\lambda(t)$ is the limit of the mortality rate if the interval of time is taken to be small (rather than one unit). The hazard rate is the instantaneous rate of failure at time t given that an individual is alive at time t. Specifically, $\lambda(t)$ is defined by the following equation

$$\lambda(t) = \lim_{h \to 0} \frac{\mathbb{P}|t \leq T < t + h|T \geq t]}{h}.$$

This can be expressed as

$$\lambda(t) = \frac{\lim_{h\to 0} \frac{\mathbb{P}[t < T < t+h]}{h}}{\mathbb{P}[T \geq t]} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d\log(S(t))}{dt}.$$

From this, we can integrate both sides to get

$$H(t) = \int_0^t \lambda(u)du = -\log(S(t)),$$

where $H(t)$ is referred to as the cumulative hazard function. Here we used the fact that $S(0) = 1$. Hence,

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t \lambda(u)du\right).$$

The expected life time and its variance are given by:

$$\mathbb{E}[T] = \int_0^\infty S(t)dt,$$
$$\mathrm{Var}(T) = 2\int_0^\infty tS(t)dt - \{\mathbb{E}(T)\}^2.$$

Example1: Exponential distribution
Suppose that $T \sim \mathcal{E}(\gamma)$ with $\gamma > 0$

$$F(t) = 1 - \exp(-\gamma t)$$
$$S(t) = \exp(-\gamma t)$$
$$f(t) = \gamma \exp(-\gamma t)$$
$$h(t) = \gamma$$

Example2: the Weibull distribution
Suppose that $T \sim \mathcal{W}(\gamma, \alpha)$:

$$F(t) = 1 - \exp\left(-\gamma t^\alpha\right)$$
$$S(t) = \exp\left(-\gamma t^\alpha\right)$$
$$f(t) = \gamma\alpha t^{\alpha-1} \exp\left(-\gamma t^\alpha\right)$$
$$h(t) = \gamma\alpha t^{\alpha-1}$$

1.4.3 Censoring

Apart from survival analysis censoring may arise in other applications, whereby not all survival data hold censored observations. However, this is one such topic that unites a lot of applications to survival analysis because censored survival data are so common and censoring needs special treatment. Censoring has many forms the most important is right censoring, Left Censoring and Interval Censoring, and there are different causes of occurrence of censoring. The most common form of censoring is right censoring.

►Right Censoring
In survival data T is the time from start of observation until an event happens and some cases become right censored as observation breaks off before the event arise. Accordingly, if T is said to be the event as person's age at death(in years), the event is right censored at age 50 if you may only know that $T > 50$. This concept is also not confined to event times only.

The income is right censored at 75,000\$, if the only thing you know is that a person's income is more then 75,000\$ per year.

►The types of right censoring
For individual i, consider that
- his survival time X_i
- his censoring time C_i
- the actual duration observed T_i

1. Type I censoring

Let C be a fixed value, instead of observing the variables $X_1, \cdots, X_n$ which interests us, we observe X_i only when $X_i \leq C$, otherwise we only know that $X_i > C$, we use the following notation:

$$T_i = X_i \wedge C = \min(X_i, C).$$

This censoring mechanism is frequently encountered in industrial applications.

2. Type II censoring

It is present when it is decided to observe the survival times of n patients until k of them have died and to stop the study at that time. Let X_i and T_i the order statistics of the variables X_i and T_i The date of censorship is therefore X_k and we observe the following variables

$$\begin{aligned} T_1 &= X_1 \\ &\vdots \\ T_k &= X_k \\ T_{k+1} &= X_k \\ &\vdots \\ T_n &= X_k. \end{aligned}$$

3. Type III censoring(or type I random censoring)

Let $C_1, \cdots, C_n$ be random variables i.i.d. We observe the variables $T_i = X_i \wedge C_i$. The available information can be summarized by:
- the actual duration observed T_i
- let δ_i denote the indicator.

$$\delta_i = \begin{cases} 1 \text{ if the event is observed } (T_i = X_i), \\ 0 \text{ if the individual is censored } (T_i = C_i). \end{cases}$$

For example, in a follow-up study, the censoring occurs due to the end of the study, loss to follow-up, or early withdrawals.
Reasons for censoring :
- patients decide to move to another hospital
- patients quit treatment because of side-effects of a drug
- failures occur after the end of study
- etc.

Example. Figure (1.1) shows data from a study in which all the persons go through heart surgery at time 0 and followed up to 3 years. The horizontal axis shows time in years after surgery and horizontal lines tagged A to E represents different person. The vertical line at 3 is the point at which we stop following the patients. An X specify that death occurred at that point in time. Deaths occurred at point 3 or before time 3 is observed and hence are uncensored but on the other hand, deaths occurring after time point 3 are not observed thus are censored at time 3. Consequently, A,C and D are uncensored, while B and E are right-censored.

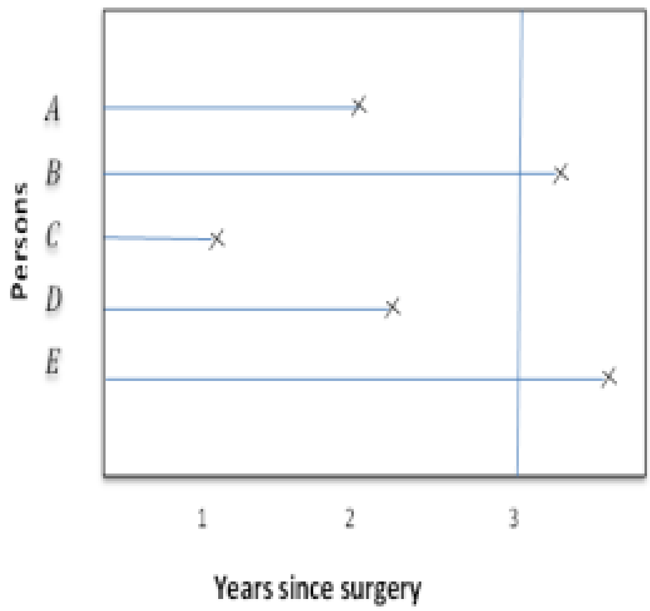

Figure 1.1: Image showing right censoring

►Left Censoring

Left censoring occur when we only know that T is less than some value. This concept is not only applicable for event time but any kind of variables. For survival data left censoring most probably occur when some of the individuals may have already experienced the event when observing a sample at a time is just started.

►Interval Censoring

Interval censoring is more common then left censoring. Both left censoring and right censoring together makes interval censoring. When you only know about variable T is $a < T < b$ for some values of a and b then T is interval censored. Interval censoring arise in survival data when the observations are made at specific time points and retroactive information on the exact timing of event cannot be achieved.

1.4.4 Kaplan-Meier Estimator

In the absence of censoring, the distribution function F is estimated very simply using the usual empirical distribution function. For example $F(t) = \mathbb{P}(T \leq t)$ is estimated by:

$$\widehat{F}_{emp}(t) = \frac{1}{n} \sum_{1 \leq i \leq n} \mathbb{1}_{\{T_i \leq t\}}.$$

Unfortunately in case the data is censored, it is impossible to use the empirical function since it involves unobserved quantities, because all the censored Y_i are not observed. It is then generally estimated that F using the Kaplan and Meier (1958) estimator. The latter is the bottom tool in statistics for non parametric estimating the distribution of a right-censored v.a. T.

The Kaplan-Meier (K-M) estimator is the most widely used duration model in practice. It intervenes in all applications that require time modeling.

Note by: $\overline{F}(t) = \mathbb{P}(T > t), \overline{G}(t) = \mathbb{P}(C > t)$ et $\overline{H}(t) = \mathbb{P}(Y > t) = \overline{F}(t)\overline{G}(t)$. The idea of this model is as follows: to survive after time t is to be alive just before t and not to die at time t, i.e. if $t_2 < t_1 < t$, using compound probabilities, we have:

$$\begin{aligned}
\overline{F}(t) &= \mathbb{P}(T > t) \\
&= \mathbb{P}\left(T > t_1, T > t\right) \\
&= \mathbb{P}\left(T > t \mid T > t_1\right) \times \mathbb{P}\left(T > t_1\right) \\
&= \mathbb{P}\left(T > t \mid T > t_1\right) \times \mathbb{P}\left(T > t_1 \mid T > t_2\right) \times \mathbb{P}\left(T > t_2\right),
\end{aligned}$$

and so on. Considering for $i = 1, \cdots, n$ only the dates when the event of interest occurs (death or censoring), we estimate quantities of the type:

$$p_i = \mathbb{P}\left(T > Y_{(i)} \mid T > Y_{(i-1)}\right),$$

where p_i is the probability of surviving in the interval $]Y_{(i-1)}, Y_{(i)}]$ knowing that we were alive in $Y_{(i-1)}$. Consider the following notations: r_i the number of individuals at risk to undergo the event just before time $Y_{(i)}$ and d_i the number of death in $Y_{(i)}$. We denote by $q_i = 1 - p_i$ the probability of dying during the interval $]Y_{(i-1)}, Y_{(i)}]$ knowing that we were alive at the beginning of this interval. so q_i can be estimated by:

$$\hat{q}_i = \frac{d_i}{r_i}.$$

As event times are assumed to be distinct, we have: $d_i = 0$ in case of censoring in $Y_{(i)}$, i.e. when $\delta_i = 0$ and $d_i = 1$ in the event of death in $Y_{(i)}$, i.e. when $\delta_i = 1$. It is clear that $r_i = n - i + 1$, we then obtain:

$$\hat{p}_i = \begin{cases} 1 - \frac{1}{n-i+1} & \text{si} \quad \delta_i = 1, \\ 1 & \text{si} \quad \delta_i = 0. \end{cases}$$

Hence, we finally arrive at the estimator of K-M of the survival function of our duration of interest T given by:

$$\widehat{S}(t) = \overline{F}_n(t) = \begin{cases} \prod_{Y_{(i)} \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{\delta_{(i)}} & \text{si } t < Y_{(n)}, \\ 0 & \text{si } t \geq Y_{(n)}. \end{cases}$$

See that the censoring situation of T by C is symmetrical to the censoring of C by T, we can define the estimator of K-M of the survival function of the censoring variable $\overline{G}(\cdot)$ by replacing $\delta_{(i)}$ by $1 - \delta_{(i)}$ which gives:

$$\overline{G}_n(t) = \begin{cases} \prod_{Y_{(i)} \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{1-\delta_{(i)}} & \text{si } t < Y_{(n)}, \\ 0 & \text{si } t \geq Y_{(n)}. \end{cases}$$

Where $Y_{(1)}, \cdots, Y_{(n)}$ are the ordered values of Y_i and $\delta_{(i)}$ is the concomitant uncensored indicator to the $Y_{(i)}$.

Remark 1.4.1. *The Kaplan-Meier (1958) estimator (KM), also called the product limit estimator, is defined as*

$$\overline{G}_n(t) = \begin{cases} \prod_{1 \le i \le n} \left(1 - \frac{1 - \delta_{(i)}}{n - i + 1}\right)^{\mathbb{1}\{Y_{(i)} \le t\}} & si \quad t < Y_{(n)}, \\ 0 & si \quad t \ge Y_{(n)}. \end{cases} \tag{1.2}$$

Example. Figure (1.2) illustrates the survival function drawn by taking a hypothetical data of group of patients entered in clinical trial receiving anti-retroviral therapy for HIV infection. The data shows the time of event i.e death, occured among the patients that is: $6, 12, 21, 27, 32, 39, 43, 43, 46^+, 89, 115^+, 139^+, 181^+, 211^+, 217^+, 261, 263, 270, 295^+, 311, 335^+, 346^+, 365^+$(+ means right censored observation).

t_i	d_i	r_i	d_i/r_i	$1 - d_i/r_i$	$\widehat{S}_{KM}$
6	1	23	0.043	0.957	0.957
12	1	23	0.45	0.955	0.913
21	1	21	0.048	0.952	0.869
27	1	20	0.05	0.95	0.825
32	1	19	0.052	0.948	0.782
39	1	18	0.056	0.944	0.738
43	2	17	0.118	0.882	0.650
89	1	14	0.714	0.928	0.603
261	1	8	0.125	0.875	0.527
263	1	7	0.143	0.857	0.451
270	1	6	0.167	0.833	0.375
311	1	4	0.25	0.75	0.281

Table 1.1: Construction of the Kaplan-Meier estimator.

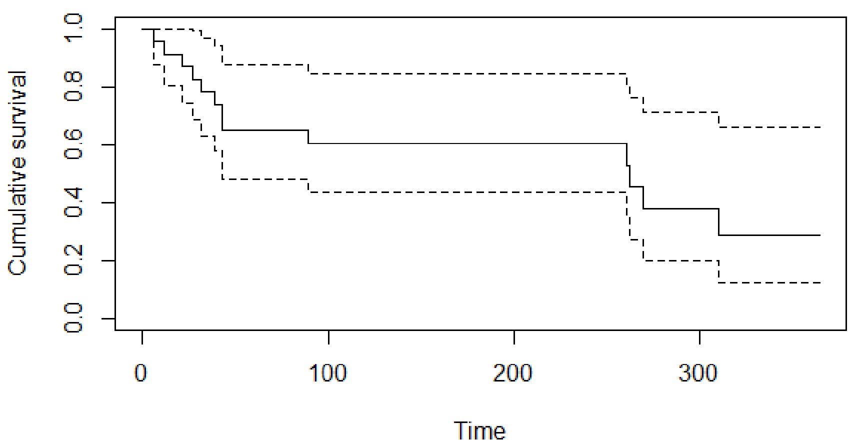

Figure 1.2: Plot of Kaplan-Meier estimates group of patients receiving ARV therapy

From Figure (1.2) we can see the estimated probability is the step function that remain unchanged even if there is a censored observation in between. The X-axis (horizontal lines) show the time past after entry into studies and the Y-axis (vertical lines) shows the estimated survival probabilities. The time t when the cumulative probability is 0.5 i.e $S(t) = 0.5$ is called median survival time which according to this example is $t = 263$. We can use different statistical programs to plot Kaplan-Meier curve such as SPSS, R, Sigma plot etc. Here in our example we have used R to plot the curve.

Chapter 2

Nonparametric robust regression for ergodic complete data

The main goal of this chapter is to treat some asymptotic results for the estimation of the robust regression function when the observations are complete and the covariates take values in an infinite dimensional space. This chapter is divided into three sections: In the first section, we will present the robust model and its estimate. In the second one, we will study the almost complete convergence. In the last section we will deal with asymptotic normality.

2.1 Robust model and its estimate

Let us consider $(X_i, Y_i)_{i=1,\dots,n}$ a couple of random variables be an $\mathcal{F} \times \mathbb{R}$-valued measurable strictly stationary ergodic process, where $\mathcal{F}$ is a semi-metric space. We denoted d the semi-metric. For $x \in \mathcal{F}$, we consider a real measurable function denoted ψ_x and we model the co-variation between X_i and Y_i through the nonparametric robust regression, denoted by θ_x, implicitly defined as a zero with respect to (w.r.t.) t of the following equation:

$$\Psi(x,t) = \mathbb{E}[\psi_x(Y_i,t)|X_i = x] = 0, \tag{2.1}$$

where ψ_x is a real-valued Borel function satisfying some regularity conditions to be stated below. We suppose that, for all $x \in \mathcal{F}$, θ_x exists and is the unique zero w.r.t t of equation (2.1). We point out that this robustification method covers and includes many important nonparametric models introduced by Huber (1964)[33], for example, $\psi_x(y,t) = (y - t)$ yields the classical regression. For more example of the function ψ_x consulted Stone (2005)[52].
For all $(x,t) \in \mathcal{F} \times \mathbb{R}$, we propose a nonparametric estimator of $\Psi(x,t)$ given by:

$$\widehat{\Psi}(x,t) = \frac{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi_x(Y_i,t)}{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))},$$

where K is a kernel and $h = h_n$ is a sequence of positive real numbers which goes to zero as n goes to infinity. A natural estimator $\widehat{\theta}_x$ of θ_x is a zero w.r.t. t of the equation

$$\widehat{\Psi}(x,t) = 0.$$

Clearly, when $\psi_x(Y,t) = Y - t$, then $\widehat{\theta}_n$ is the estimator given in Ferraty and Vieu (2006)[23]for the functional nonparametric regression, while for $\psi_x(y,t) = \mathbb{1}_{y>t} - (1-\alpha)$, we obtain the αth conditional quantile estimate studied by Laksaci et al.(2009)[47].

2.2 Almost complete convergence

The concept of almost complete convergence was introduced by Hsu and Robbins (1947)[32]. It implies almost sure convergence and lends itself well to calculations involving sums of random variables. Despite this, it did not begin to become popular in the statistical community in the 1980s after the work of Collomb. It is used mainly in nonparametric statistics.

In this section, we will list some assumptions and we will provide the main results.

2.2.1 Notations, hypotheses and comments

From now on, we will denote by C and C' some strictly positive generic constants, x is a fixed point in $\mathcal{F}$, $\mathcal{N}_x$ denotes a fixed neighborhood of x. For $r > 0$, let $B(x,r) = \{x' \in \mathcal{F}/d(x',x) < r\}$. Furthermore, for $i = 1,\dots,n$, we define $\mathcal{F}_k$ as the σ-field generated by $((X_1,Y_1),\dots,(X_k,Y_k))$ and $\mathcal{G}_k$ as the σ-field generated by $((X_1,Y_1),\dots,(X_k,Y_k),X_{k+1})$. We need the following hypotheses to establish our asymptotic results:

(H1) The process $(X_i,Y_i)_{i\in\mathbb{N}}$ satisfies:
(i) The function $\phi(x,r) = \mathbb{P}(X \in B(x,r)) > 0, \forall r > 0$.
(ii)For all $i = 1,...,n$ there exists a deterministic function $\phi_i(x,\cdot)$ such that $0 < \mathbb{P}$ $(X_i \in B(x,r)|\mathcal{F}_{i-1}) \le \phi_i(x,r), \forall r > 0$.
(iii) For all $r > 0$, $\frac{1}{n\phi(x,r)} \sum_{i=1}^{n} \mathbb{P}(X_i \in B(x,r)|\mathcal{F}_{i-1}) \longrightarrow 1$ a.co.

(H2) The function Ψ is such that :
(i) The function $\Psi(x,.)$ is of class $\mathcal{C}^1$ on $[\theta_x - \delta, \theta_x + \delta], \delta > 0$.
(ii) For each fixed $t \in [\theta_x - \delta, \theta_x + \delta]$ the function $\Psi(\cdot,t)$ is continuous at the point x.
(iii) $\forall (t_1,t_2) \in [\theta_x - \delta, \theta_x + \delta] \times [\theta_x - \delta, \theta_x + \delta], \forall (x_1,x_2) \in \mathcal{N}_x \times \mathcal{N}_x$
$|\Psi(x_1,t_1) - \Psi(x_2,t_2)| \le C d^{b_1}(x_1,x_2) + |t_1 - t_2|^{b_2}, b_1 > 0, b_2 > 0$.

(H3) For each fixed t $\in [\theta_x - \delta, \theta_x + \delta]$, $\forall j \ge 1, \mathbb{E}[\psi_x^j(Y,t)|\mathcal{G}_{i-1}] = \mathbb{E}[\psi_x^j(Y,t)|X_i] < Cj! < \infty$, a.s.

(H4) The function ψ_x is monotone w.r.t. the second component.

(H5) K is a function with support (0,1) such that
$$0 < C\mathbb{1}_{(0,1)} < K(t) < C'\mathbb{1}_{(0,1)} < \infty.$$
The derivative K' exists on [0,1] and satisfied the condition $K'(t) < 0, \forall t \in [0,1]$ and $\left|\int_0^1 (K^j)'(u)du\right| < \infty$ for $j \ge 1$.

(H6) $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} \dfrac{\varphi(x,h)\ log\ n}{n^2\phi^2(x,h)} = 0$ where $\varphi(x,h) = \sum_{i=1}^{n} \phi_i(x,h)$.

Comments on the hypotheses

Our assumptions are very standard in this context. Indeed, the ergodicity of functional data: The latter is exploited together with condition (H1) which is less restrictive to the conditions imposed by Laïb and Louani (2011)[44] since, it is not necessary to write (approximately) the concentration function $\mathbb{P}(X_i \in B(x,r)) > 0$ and the conditional concentration function $\mathbb{P}(X_i \in B(x,r)|\mathcal{F}_{i-1})$ as products of two independent nonnegative functions of the center and radius. In both the functional and finite dimensional cases, this generalization on condition (H1) is very important. Indeed, in the functional case, In addition, such a form of writing in the multivariate case needs the differentiability of the marginal (resp. the conditional) distribution function as well as the strict positivity of its densities; Thus, we can also proceed without the existence of these densities and even if these densities are vanishing at a center. Furthermore, in the multivariate case, when the marginal density (resp. the conditional density) of X given $\mathcal{F}_{i-1}$ exists and is continuous and strictly positive, the hypothesis (H1)(iii) is a direct consequence of Beck's theorem (Györfi et al., 1989, p. 49)[27]. the nonparametric model: the functional space of the model is characterized by some regularity conditions allowing the bias term to be evaluated. Thus there are two kinds of conditions: the first one ((H2)(ii)) is the continuity of the model, while the second one is based on Lipschitz-type condition (H2)(iii). The first one is necessary to get the convergence, while the second consideration is used to make precise the convergence rate of the estimate. Condition (H3) is a standard assumption over j^{th} moments of the conditional expectation of the function ψ_x. The robustness property is controlled by Condition (H4) where only the convexity (which is fundamentals constraints of the robustness properties of the M-estimators) of the score function is needed. In order to cover the classical regression studied in this ergodic functional context by Laib and Louani (2011)[44] we establish our asymptotic normality without the boundedness condition for the score function. Furthermore, they are needed to evaluate the bias term in the asymptotic properties. Assumption (H5) concerns the kernel $K(\cdot)$ which is technical and imposed for sake of simplicity whereas (H6) is classical for consistency results.

2.2.2 Results

Now we are in a position to give our main result. Our first main result is given in the following theorem which deals with pointwise almost complete convergence.

Theorem 2.2.1. *[2] Assume that* $(H1)$, $(H2)((i)-(ii))$ *and* $(H3)$-$(H6)$ *are satisfied, then* $\widehat{\theta}_x$ *exists for all sufficiently large* n. *Furthermore, we have*

$$\widehat{\theta}_x - \theta_x \to 0 \quad a.co.$$

In order to give a more accurate asymptotic result, we replace (H2) (ii) by (H2)(iii) and we obtain the following result:

Theorem 2.2.2. *[2] Assume that* $(H1)$, $(H2)((i)-(iii))$ *and* $(H3)$-$(H6)$ *are satisfied, then* $\widehat{\theta}_x$ *exists a.s. for all sufficiently large* n. *Furthermore, if* $\Psi'(x,\theta_x) \neq 0$ *we have*

$$\widehat{\theta}_x - \theta_x = \mathrm{O}(h^{b_1}) + \mathrm{O}\left(\sqrt{\frac{\varphi(x,h)\ log\ n}{n^2\phi^2(x,h)}}\right) \ a.co.$$

Proof. For the proofs of Theorems (2.2.1) and (2.2.2) we use the fact that ψ_x is monotone w.r.t. the second component. We give the proof for the case of an increasing $\psi_x(Y,\cdot)$, the decreasing case being obtained in the same manner. For this case we write, $\epsilon > 0$, we set $K_i = K\left(\frac{d(x,X_i)}{h}\right)$.

$$\sum_n \mathbb{P}[|\widehat{\theta}_x - \theta_x| \geq \epsilon] \leq \sum_n \mathbb{P}\left[\left(|\widehat{\theta}_x - \theta_x| \mathbb{1}_{\{|\widehat{\theta}_x - \theta_x| \leq \delta\}}\right) \geq \epsilon\right] + \sum_n \mathbb{P}\left[\left(|\widehat{\theta}_x - \theta_x| \mathbb{1}_{\{|\widehat{\theta}_x - \theta_x| > \delta\}}\right) \geq \epsilon\right].$$

Since $\psi_x(Y, \cdot)$ is increasing, it follows that,

$$\begin{aligned} \mathbb{P}\left(\left(|\widehat{\theta}_x - \theta_x| \mathbb{1}_{\{|\widehat{\theta}_x - \theta_x| > \delta\}}\right) \geq \epsilon\right) &\leq \mathbb{P}(|\widehat{\theta}_x - \theta_x| > \delta_0) \\ &\leq \mathbb{P}(|\widehat{\Psi}(x, \theta_x + \delta) - \Psi(x, \theta_x + \delta)| \geq \Psi(x, \theta_x + \delta)) \\ &\quad + \mathbb{P}(|\widehat{\Psi}(x, \theta_x - \delta) - \Psi(x, \theta_x - \delta)| \geq -\Psi(x, \theta_x - \delta)). \end{aligned}$$

Moreover, we can write under (H2)(i)

$$(\widehat{\theta}_x - \theta_x) \mathbb{1}_{\{|\widehat{\theta}_x - \theta_x| \leq \delta\}} = \frac{\Psi(x, \widehat{\theta}_x) - \widehat{\Psi}(x, \widehat{\theta}_x)}{\Psi'(x, \theta_n^\star)},$$

where $\theta_n^\star$ is a point between $\widehat{\theta}_x$ and θ_x.
Therefore, the remaining task is to study the convergence rate of

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} |\Psi(x, t) - \widehat{\Psi}(x, t)|$$

and to show that

$$\exists \tau > 0, \qquad \sum_{n=1}^{\infty} \mathbb{P}(\Psi'(x, \theta_n^\star) < \tau) < \infty. \tag{2.2}$$

To reach this end, we write

$$\widehat{\Psi}(x, t) = B_n(x, t) + \frac{R_n(x, t)}{\widehat{\Psi}_D(x)} + \frac{Q_n(x, t)}{\widehat{\Psi}_D(x)},$$

where

$$\begin{aligned} Q_n(x, t) &= (\widehat{\Psi}_N(x, t) - \bar{\Psi}_N(x, t)) - \Psi(x, t)(\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)) \\ B_n(x, t) &= \frac{\bar{\Psi}_N(x, t)}{\bar{\Psi}_D(x)} - \Psi(x, t), \quad \textit{and} \quad R_n(x, t) = -B_n(t)(\widehat{\Psi}_N(x, t) - \bar{\Psi}_N(x, t)), \end{aligned}$$

with

$$\begin{aligned} \widehat{\Psi}_N(x, t) &= \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^{n} K_i \psi_x(Y_i, t), \\ \bar{\Psi}_N(x, t) &= \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^{n} \mathbb{E}[K_i \psi_x(Y_i, t) | \mathcal{F}_{i-1}], \\ \widehat{\Psi}_D(x) &= \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^{n} K_i, \\ \bar{\Psi}_D(x) &= \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^{n} \mathbb{E}[K_i | \mathcal{F}_{i-1}]. \end{aligned}$$

Therefore, both theorems are a consequence of the following intermediate results.

The asymptotic behavior of the term $\widehat{\Psi}_D(x)$ is described by the following lemma

Lemma 2.2.1. *[2] Assume that hypotheses (H1)-(H5) and (H6), we have*

$$\widehat{\Psi}_D(x) - \bar{\Psi}_D(x) = \mathrm{O}\left(\sqrt{\frac{\varphi(x,h)\ \log(n)}{n^2\phi^2(x,h)}}\right) \quad a.co.$$

Proof. We have

$$\widehat{\Psi}_D(x) - 1 = R_{n,1}(x) + R_{n,2}(x),$$

where

$$R_{n,1}(x) = \frac{1}{n\mathbb{E}\left[K_1(x)\right]}\sum_{i=1}^{n}\left(K_i(x) - \mathbb{E}\left[K_i(x) \mid \mathcal{F}_{i-1}\right]\right) = \widehat{\Psi}(x) - \bar{\Psi}(x),$$

$$R_{n,2}(x) = \frac{1}{n\mathbb{E}\left[K_1(x)\right]}\sum_{i=1}^{n}\left(\mathbb{E}\left[K_i(x) \mid \mathcal{F}_{i-1}\right] - \mathbb{E}K_1(x)\right).$$

With condition (H6)(iii), it is easily seen that $R_{n,2}(x) = \mathrm{o}_{\text{a.s.}}(1)$ as $n \to \infty$.
For the the first term, since the kernel K is bounded, observe that $R_{n,1}(x) = 1/n\mathbb{E}\left[K_1(x)\right]\sum_{i=1}^{n} L_{n,i}(x)$, where $\{L_{n,i}(x)\}$ is a triangular array of bounded martingale differences with respect to the sequence of σ-fields $(\mathcal{F}_{i-1})_{i\geq 1}$. Using the Jensen inequality (4.6.1), we can deduce that

$$\mathbb{E}\left[L_{n,i}^2(x) \mid \mathcal{F}_{i-1}\right] \leq 2\mathbb{E}\left[K_i^2(x) \mid \mathcal{F}_{i-1}\right].$$

By using a similar argument stated in Lemma 2 in Laib and Louani, (2010) [46], we obtained,

$$R_{n,1}(x) = \mathrm{O}_{a.co.}\left(\sqrt{\tfrac{\varphi(x,h)\ log\ n}{n^2\phi^2(x,h)}}\right). \qquad \square \tag{2.3}$$

Corollary 2.2.1. *[2] Under Hypotheses of Lemma (2.2.1), we have,*

$$\exists C > 0 \quad \sum_{n=1}^{\infty}\mathbb{P}(\widehat{\Psi}_D(x) < C) < \infty.$$

Proof. It is clear that, under (H5), there exists $0 < C < C' < \infty$ where

$$0 < C\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathcal{F}_{i-1}) < \bar{\Psi}_D(x) < |\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)| + \widehat{\Psi}_D(x).$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\widehat{\Psi}_D(x) \leq \tfrac{C}{2}\right) &\leq \mathbb{P}\left(\tfrac{C}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathcal{F}_{i-1}) < \frac{C}{2} + |\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)|\right)\\
&\leq \mathbb{P}\left(\left|\tfrac{C}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathcal{F}_{i-1}) - |\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)| - C\right| > \tfrac{C}{2}\right).
\end{aligned}$$

It is obvious that the previous lemma and (H1)(iii) enable us to arrive at this conclusion.

$$\sum_n \mathbb{P}\left(\left|\tfrac{C}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathcal{F}_{i-1}) - |\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)| - C\right| > \tfrac{C}{2}\right),$$

which gives the result. $\square$

Lemma 2.2.2. *[2] Under hypotheses $(H1)$, $(H2)((i),(ii))$, $(H5)$ and $(H6)$, we have*

$$\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |B_n(x,t)| = \mathrm{o}(1),$$

If we replace $(H2)(ii)$ by $(H2)(iii)$, we have

$$\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |B_n(x,t)| = \mathrm{O}(h^{b_1}).$$

Proof. We begin by assessing the conditional bias term. Observe that:

$$B_n(x,t) = \frac{\bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)}.$$

Similarly as in Lemma (2.2.1), it is easily seen that $\bar{\Psi}_D(x) = \mathrm{O}_{a.s.}(1)$. Therefore, we have to establish that

$$\widetilde{B}_n(x,t) = \bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x) = \mathrm{O}_{a.co.}(h^{b_1}).$$

Making use of conditions (H2)(iii) and (H3) one can easily see that

$$\begin{aligned}
\left|\tilde{B}_n(x)\right| &= \left|\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^n \mathbb{E}\left[\left(\psi_x(Y_i,t)-\Psi(x,t)\right)K_i(x)\mid \mathcal{F}_{i-1}\right]\right| \\
&= \left|\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\left(\psi_x(Y_i,t)-\Psi(x,t)\right)K_i(x)\mid \mathcal{G}_{i-1}\right]\mid \mathcal{F}_{i-1}\right]\right| \\
&= \left|\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\left(\psi_x(Y_i,t)-\Psi(x,t)\right)K_i(x)\mid X_i\right]\mid \mathcal{F}_{i-1}\right]\right| \\
&= \left|\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^n \mathbb{E}\left[\left(\Psi\left(X_i,t\right)-\Psi(x,t)\right)K_i(x)\mid \mathcal{F}_{i-1}\right]\right| \\
&\leq \sup_{u\in B(x,h)}|\Psi(u,t)-\Psi(x,t)|\left|\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^n \mathbb{E}\left[K_i(x)\mid \mathcal{F}_{i-1}\right]\right| = \mathrm{O}_{a.co.}\left(h^{b_1}\right),
\end{aligned}$$

since the support of the kernel K is the interval [0,1]. □

Lemma 2.2.3. *[2] Under hypotheses $(H1)$ and $(H3)$-$(H6)$, we have*

$$\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)| = \mathrm{O}\left(\sqrt{\frac{\varphi(x,h)\ log\ n}{n^2\phi^2(x,h)}}\right) \quad a.co.$$

Proof. Using the compactness of $[\theta_x-\delta,\theta_x+\delta]$, we cover the compact by $[\theta_x-\delta,\theta_x+\delta] \subset \bigcup_{j=1}^{d_n}(t_j-l_n, t_j+l_n)$ with $l_n = n^{-1/2b_2}$ and $d_n = \mathrm{O}(n^{1/2b_2})$. To do that, we denote by $\mathcal{G}_n$ the subset of the intervals extremities grid

$$\mathcal{G}_n = \{t_j - l_n, t_j + l_n, 1 \leq j \leq d_n\}. \tag{2.4}$$

We use the monotony of ψ_x to show that

$$\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)| \leq \max_{1\leq j\leq d_n}\max_{z\in\{t_j-l_n,t_j+l_n\}} |\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z)| + 2^{b_2}C_2 l_n^{b_2}\bar{\Psi}_D(x).$$

From Lemma (2.2.1), we deduce that

$$l_n^{b_2}\bar{\Psi}_D(x) = \mathrm{O}_{a.co.}\left(\sqrt{\tfrac{1}{n}}\right) = \mathrm{O}_{a.co.}\left(\sqrt{\tfrac{\varphi(x,h)\ log\ n}{n^2\phi^2(x,h)}}\right) \quad .$$

Therefore, all it remains to show that

$$\max_{1\le j\le d_n}\max_{z\in\{t_j-l_n,t_j+l_n\}}|\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)| = \mathrm{O}_{a.co.}\left(\sqrt{\tfrac{\varphi(x,h)\ log\ n}{n^2\phi^2(x,h)}}\right) \quad . \tag{2.5}$$

The proof of (2.5) is based on the application of the exponential inequality (4.6.1), on the triangular array of martingale differences (according to the σ-fields $(\mathcal{F}_{i-1})_i$):

$$L_i = K_i(x)\psi_x(Y_i,z) - \mathbb{E}\left[K_i(x)\psi_x(Y_i,z)\mid\mathcal{F}_{i-1}\right], \quad z\in\mathcal{G}_n$$

the application of the mentioned inequality is based on evaluation of $\mathbb{E}\left[L_i^p(x,z)\mid\mathcal{F}_{i-1}\right]$. The latter can be evaluated by using the same arguments as were invoked for proving Lemma 5 in Laïb and Louani (2011)[44], allowing us to write, under (H3), for any $p\in\mathbb{N}-\{0\}$, observe that

$$L_i^p(x,z) = \sum_{k=0}^{p} C_p^k\left(K_i(x)\psi_x(Y_i,z)\right)^k(-1)^{p-k}\left[\mathbb{E}\left(K_i(x)\psi_x(Y_i,z)\mid\mathcal{F}_{i-1}\right)\right]^{p-k}.$$

Thus,

$$\mathbb{E}\left(L_i^p(x,z)\mid\mathcal{F}_{i-1}\right) = \sum_{k=0}^{p} C_p^k\mathbb{E}\left[\left(K_i(x)\psi_x(Y_i,z)\right)^k\mid\mathcal{F}_{i-1}\right](-1)^{p-k}\left[\mathbb{E}\left(K_i(x)\psi_x(Y_i,z)\mid\mathcal{F}_{i-1}\right)\right]^{p-k}.$$

Therefore,

$$\mathbb{E}\left[L_i^p(x,z)\mid\mathcal{F}_{i-1}\right]\le C\phi_i(x,h).$$

Thus, we can presently apply the aforementioned exponential inequality and obtain: for all $\eta>0$ and $d_n = l_n^{-1}$,
we have

$$l_n^{-1}\max_{z\in\mathcal{G}_n}\mathbb{P}\left(\left|\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)\right| > \eta\sqrt{\frac{\varphi(x,h)\log n}{n^2\phi^2(h)}}\right)\le C'n^{-C\eta^2+1/2b_2}.$$

As a result, a suitable choice of η completes the proof of this lemma. □

Lemma 2.2.4. *[2] Under Assumptions (H1),(H2)((i),(ii)) and (H3)-(H6), $\widehat{\theta}_x$ exists a.s.for all sufficiently large n and there exists $\zeta_1>0$ such that*

$$\sum_{n\ge1}\mathbb{P}\{\Psi'(x,\theta_n^\star)<\zeta_1\}<\infty.$$

Proof. It is clear that, the monotony of ψ_x, for all $\epsilon>0$

$$\Psi(x,\theta_x-\epsilon)\le\Psi(x,\theta_x)\le\Psi(x,\theta_x+\epsilon).$$

By using a similar argument as those used in the previous Lemmas, we show that

$$\widehat{\Psi}(x,\theta_x)-\Psi(x,\theta_x) = \mathrm{O}\left(h^{b_1}+\sqrt{\frac{\log n}{n\phi(x,h)}}\right), \qquad \text{a.co.}$$

for all real fixed t. So, for sufficiently large n and for all ϵ small enough

$$\widehat{\Psi}\left(x, \theta_x - \epsilon\right) \leq 0 \leq \widehat{\Psi}\left(x, \theta_x + \epsilon\right). \quad \text{a.co.}$$

Since ψ_x and K are continuous functions, it follows that $\widehat{\Psi}(x,t)$ is a continuous function of t and, there exists a $\widehat{\theta}_x \in [\theta_x - \epsilon, \theta_x + \epsilon]$ such that $\widehat{\Psi}(x, \widehat{\theta}_x) = 0$. Concerning, the uniqueness of $\widehat{\theta}_x$, we point out that the latter is a direct consequence of the strict monotonicity of ψ_x and the positivity of K, and its second part is a direct consequence of the regularity assumption (H2)(i) on $\Psi(x, \cdot)$. □

2.3 Asymptotic normality

The asymptotic property discussed in this section is asymptotic normality, this is of a great significance in statistics, this section deals, under the hypothesis of a stationary ergodic process, the asymptotic normality of the same estimator.

2.3.1 Notations, hypotheses and comments

Now, we study the asymptotic normality of $\widehat{\theta}_x$. We replace (H1),(H2),(H3),(H4),(H5) and (H6) by the following hypotheses, respectively.

(H'1) The concentration property (H1) holds.

(H'2) Condition (H2(i)) holds. Moreover, the condition (H2 (ii)) is replaced by the function $\Psi(\cdot, t)$ and $\lambda_2(\cdot, t) = \mathbb{E}[\psi_x^2(Y,t)|X = \cdot]$ are continuous at the point x. And the condition (H2 (iii)) is replaced by the derivative function
$\Phi(s, z) = \mathbb{E}[\Psi(X_1, z) - \Psi(x, z)|d(x, X_1) = s]$ exists at $s = 0$ and is continuous w.r.t the second component at $\mathcal{N}_x$.

(H'3) The condition (H3) stays the same.

(H'4) The condition (H4) for the function ψ_x stays the same.

(H'5) The kernel K satisfies (H5) and is a positive function supported on]0,1[.

(H'6) There exists a function $\tau_x(\cdot)$ such that

$$\forall t \in [0,1] \ \lim_{h \to 0} \frac{\phi(x, th)}{\phi(x, h)} = \tau_x(t).$$

$$K^2(1) - \int_0^1 (K^2(u))' \tau_x(u) du > 0 \qquad \textit{and} \qquad K(1) - \int_0^1 K'(u) \tau_x(u) du \neq 0.$$

Comments on the hypotheses

Our assumptions are quite mild. In this work, the functional space of our model is characterized by the regularity condition (H'2(iii)). This condition replace the Lipschitz condition usually assumed in nonparametric functional data analysis. This change is useful in order to explicit asymptotically the bias term. However, the Lipschitz condition gives inexact/inaccurate asymptotic bias term which is not interesting for the asymptotic normality. The conditions (H'5) concerns the kernel $K(\cdot)$ which is technical and imposed for sake of simplicity whereas. Moreover, the function $\tau_x(\cdot)$ defined in (H'6) plays a fundamental role in the asymptotic normality result. It permits to give the variance term explicitly.

2.3.2 Results

Our main result is detailed in the following theorem

Theorem 2.3.1. *[7] Assume that (H'1)-(H'6) hold, then $\widehat{\theta}_x$ exists and is unique with great probability and for any $x \in \mathcal{A}$, we have*

$$\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta}_x - \theta_x - B_n(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n \to \infty$$

where

$$B_n(x) = h\Phi'(0,\theta_x)\frac{\beta_0}{\beta_1} + o(h) \text{ and } \sigma^2(x,\theta_x) = \frac{\beta_2\lambda_2(x,\theta_x)}{\beta_1^2(\Gamma_1(x,\theta_x))^2}$$

with

$$\beta_0 = -\int_0^1 (sK(s))'\beta_x(s)ds, \quad (\beta_j = -\int_0^1 \left(K^j\right)'(s)\beta_x(s)ds, \text{ for } j = 1,2),$$

$$\Gamma_1(x,\theta_x) = \frac{\partial}{\partial t}\Psi(x,\theta_x) \text{ and } \quad \mathcal{A} = \{x \in \mathcal{F}, \lambda_2(x,\theta_x)\Gamma_1(x,\theta_x) \neq 0\}$$

and $\xrightarrow{\mathcal{D}}$ means the convergence in distribution.

Corollary 2.3.1. *[7] Under the hypotheses of Theorem (2.3.1) and if the bandwidth parameter h satisfies $nh^2\phi(x,h) \to 0$ as $n \to \infty$, then*

$$\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta}_x - \theta_x\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n \to \infty.$$

Proof.(Proof of Theorem (2.3.1) and Corollary (2.3.1)) We give the proof for the case of an increasing ψ_x, decreasing case being obtained by considering $-\psi_x$. In this case, we define, for all $u \in \mathbb{R}, z = \theta_x - B_n(x) + u[n\phi(x,h)]^{-1/2}\sigma(x,\theta_x)$. Let us notice that,

$$\mathbb{P}\left\{\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta}_x - \theta_x + B_n(x)\right) < u\right\} = \mathbb{P}\left\{\widehat{\theta}_x < \theta_x - B_n(x) + u[n\phi(x,h)]^{-1/2}\sigma(x,\theta_x)\right\}$$
$$= \mathbb{P}\{0 < \widehat{\Psi}(x,z)\}.$$

It is clear that we can write

$$\widehat{\Psi}(x,t) = B_n(x,t) + \frac{R_n(x,t)}{\widehat{\Psi}_D(x)} + \frac{Q_n(x,t)}{\widehat{\Psi}_D(x)}.$$

It follows that

$$\mathbb{P}\left\{\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta}_x - \theta_x + B_n(x)\right) < u\right\} = \mathbb{P}\left\{-\widehat{\Psi}_D(x)B_n(x,z) - R_n(x,z) < Q_n(x,z)\right\}.$$

Our main result is a consequence of the following intermediates results,

Lemma 2.3.1. *[7] Under the hypotheses of Theorem (2.3.1), we have for any $x \in \mathcal{A}$*

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} Q_n(x,z) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as\ n \to \infty.$$

Proof. First, we introduce some notations. We note $K_i(x) = K(h^{-1}d(x, X_i))$ and

$$\eta_{ni} = \left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)} \right)^{1/2} \left(\psi_x\left(Y_i,z\right) - \Psi(x,z)\right) \frac{K_i(x)}{\mathbb{E}K_1(x)}, \tag{2.6}$$

and defined $\zeta_{ni} = \eta_{ni} - \mathbb{E}[\eta_{ni}|\mathcal{F}_{i-1}]$. It is easily seen that

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)} \right)^{1/2} Q_n(x,z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_{ni}.$$

Since ζ_{ni} is a triangular array of martingale differences according to the σ-fields $(\mathcal{F}_{i-1})_i$, we can apply the central limit theorem based on the unconditional Lindeberg condition. More precisely, we must verify the following conditions

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\zeta_{ni}^2 \mid \mathcal{F}_{i-1}\right] \to 1 \text{ in probability },\tag{2.7}$$

and

$$\forall \epsilon > 0 \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\zeta_{ni}^2 \mathbb{1}_{\zeta_{ni}^2 > \epsilon n}\right] \to 0. \tag{2.8}$$

We begin by proving (2.7). To do that, we write

$$\mathbb{E}\left[\zeta_{ni}^2 \mid \mathcal{F}_{i-1}\right] = \mathbb{E}\left[\eta_{ni}^2 \mid \mathcal{F}_{i-1}\right] - \mathbb{E}^2\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right].$$

Therefore, it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}^2\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right] \xrightarrow{\mathbb{P}} \quad 0, \tag{2.9}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\eta_{ni}^2 \mid \mathcal{F}_{i-1}\right] \xrightarrow{\mathbb{P}} \quad 1. \tag{2.10}$$

For the first convergence, we have

$$\begin{aligned}
\left|\mathbb{E}\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right]\right| &= \frac{1}{\mathbb{E}K_1(x)} \left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)} \right)^{1/2} \left|\mathbb{E}\left[\left(\Psi\left(X_i,t\right) - \Psi(x,t)\right) K_i(x) \mid \mathcal{F}_{i-1}\right]\right| \\
&\le \frac{1}{\mathbb{E}K_1(x)} \left(\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)} \right) \right)^{1/2} \sup_{u \in B(x,h)} |\Psi(u,t) - \Psi(x,t)| \mathbb{E}\left[K_i(x) \mid \mathcal{F}_{i-1}\right].
\end{aligned}$$

Obviously, under (H'1) and (H'5) we have

$$C\phi_i(x,h) \le \mathbb{E}\left[K_i \mid \mathcal{F}_{i-1}\right] \le C'\phi_i(x,h),$$

and

$$C\phi(x,h) \le \mathbb{E}\left[\Delta_i(x)\right] \le C'\phi(x,h).$$

On other hand condition (H'2(ii)) implies that

$$\sup_{u \in B(x,h)} |\Psi(u,t) - \Psi(x,t)| = o(1).$$

Combining these three lasts results, we obtain

$$\begin{aligned}(|\ \mathbb{E}\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right]\ |)^2 &\leq \sup_{u\in B(x,h)} \left|\Psi(u,t)-\Psi(x,t)\left(\frac{\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)}\right)\right| \frac{1}{\phi(x,h)}\phi_i^2(x,h)\\ &\leq \sup_{u\in B(x,h)} \left|\Psi(u,t)-\Psi(x,t)\right| \left(\frac{\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)}\right) \frac{1}{\phi(x,h)}\phi_i(x,h).\end{aligned}$$

Finally, providing the fact that (see, (H'1(iii)))

$$\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\phi_i(x,h) \xrightarrow{\mathbb{P}} \quad 1,$$

we obtain

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right]\right)^2 &= \sup_{u\in B(x,h)} \left|\Psi(u,t)-\Psi(x,t)\right| \left(\frac{\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)}\right)\left(\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\phi_i(x,h)\right)\\ &= o_p(1).\end{aligned}$$

Now, we move to the convergence in (2.10). We write

$$\begin{aligned}\tfrac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\eta_{ni}^2 \mid \mathcal{F}_{i-1}\right] &= \frac{1}{n\left(\mathbb{E}K_1(x)\right)^2}\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)}\right)\sum_{i=1}^{n}\mathbb{E}\left[\left(\psi_x\left(Y_i,z\right)-\Psi(x,z)\right)^2 K_i^2(x) \mid \mathcal{F}_{i-1}\right]\\ &= \tfrac{1}{n(\mathbb{E}K_1(x))^2}\left(\tfrac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)\left(\sum_{i=1}^{n}\mathbb{E}\left[\psi_x^2\left(Y_i,z\right)K_i^2(x) \mid \mathcal{F}_{i-1}\right]\right.\\ &\quad -2\Psi(x,z)\sum_{i=1}^{n}\mathbb{E}\left[\psi_x\left(Y_i,z\right)K_i^2(x) \mid \mathcal{F}_{i-1}\right]\\ &\quad \left.+\Psi^2(x,z)\sum_{i=1}^{n}\mathbb{E}\left[K_i^2(x) \mid \mathcal{F}_{i-1}\right]\right).\end{aligned}$$

Denote

$$D_1=\sum_{i=1}^{n}\mathbb{E}\left[\psi_x^2\left(Y_i,z\right)K_i^2(x) \mid \mathcal{F}_{i-1}\right], \quad D_2=\sum_{i=1}^{n}\mathbb{E}\left[\psi_x\left(Y_i,z\right)K_i^2(x) \mid \mathcal{F}_{i-1}\right],$$

and

$$D_3=\sum_{i=1}^{n}\mathbb{E}\left[K_i^2(x) \mid \mathcal{F}_{i-1}\right].$$

It is clear that

$$\begin{aligned}D_1 &= \lambda_2(x,z)\sum_{i=1}^{n}\mathbb{E}[K_i^2(x) \mid \mathcal{F}_{i-1}]+\sum_{i=1}^{n}[\mathbb{E}[\psi_x^2\left(Y_i,z\right)K_i^2(x) \mid \mathcal{F}_{i-1}]\\ &\quad -\lambda_2(x,z)\mathbb{E}[K_i^2 \mid \mathcal{F}_{i-1}]]\\ D_1 &= \lambda_2(x,z)\sum_{i=1}^{n}\mathbb{E}[K_i^2(x) \mid \mathcal{F}_{i-1}]+\sum_{i=1}^{n}[\mathbb{E}[K_i^2(x)\mathbb{E}[\psi_x^2\left(Y_i,z\right) \mid \mathcal{G}_{i-1}] \mid \mathcal{F}_{i-1}]\\ &\quad -\lambda_2(x,z)\mathbb{E}[K_i^2(x) \mid \mathcal{F}_{i-1}]]\\ D_1 &= \lambda_2(x,z)\sum_{i=1}^{n}\mathbb{E}[K_i^2(x) \mid \mathcal{F}_{i-1}]+\sum_{i=1}^{n}[\mathbb{E}[K_i^2(x)\mathbb{E}[\psi_x^2\left(Y_i,z\right) \mid X_i] \mid \mathcal{F}_{i-1}]\\ &\quad -\lambda_2(x,z)\mathbb{E}[K_i^2(x) \mid \mathcal{F}_{i-1}]].\end{aligned}$$

Using the same arguments as those used in (2.9), to evaluate the second term.Then, we have,

$$\frac{1}{n\mathbb{E}[K_1(x)]}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i^2(x)\mathbb{E}\left[\psi^2\left(Y_i,z\right)\mid X_i\right]\mid\mathcal{F}_{i-1}\right]-\lambda_2(x,z)\mathbb{E}\left[K_i^2(x)\mid\mathcal{F}_{i-1}\right]\right]$$
$$\leq\sup_{u\in B(x,h)}\left|\lambda_2(x,z)-\lambda_2(x,z)\right|\left(\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}\left(X_i\in B(x,h)\mid\mathcal{F}_{i-1}\right)\right).$$

Moreover, we use the continuity of $\lambda_2(x,\cdot)$to write

$$\lambda_2(x,z)=\lambda_2(x,\theta_x)+o(1).$$

Thus,

$$\frac{1}{n\mathbb{E}\left[K_1(x)\right]}D_1=\lambda_2(x,\theta(x))\frac{1}{n\mathbb{E}\left[K_1(x)\right]}\sum_{i=1}^{n}\mathbb{E}\left[K_i^2(x)\mid\mathcal{F}_{i-1}\right]+o(1),$$

and similarly we can get

$$\frac{1}{n\mathbb{E}\left[K_1(x)\right]}D_2=\Psi_x(x,\theta(x))\frac{1}{n\mathbb{E}\left[K_1(x)\right]}\sum_{i=1}^{n}\mathbb{E}\left[K_i^2(x)\mid\mathcal{F}_{i-1}\right]+o(1)=o(1).$$

Finally, we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\eta_{ni}^2\mid\mathcal{F}_{i-1}\right]=\frac{1}{n\left(\mathbb{E}K_1(x)\right)^2}\left(\frac{\phi(x,h)\beta_1^2}{\beta_2}\right)\sum_{i=1}^{n}\mathbb{E}\left[K_i^2(x)\mid\mathcal{F}_{i-1}\right]+o(1).$$

Next, we apply the same concepts used in Ferraty et al. (2009)[24] to get

$$\mathbb{E}\left[K_i^2(x)\mid\mathcal{F}_{i-1}\right]=K^2(1)\phi_i(x,h)-\int_0^1\left(K^2(u)\right)'\phi_i(x,uh)du,$$

and

$$\mathbb{E}\left[K_1(x)\right]=K(1)\phi(x,h)-\int_0^1(K(u))'\phi(x,uh)du.$$

It follows that

$$\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{E}\left[K_i^2(x)\mid\mathcal{F}_{i-1}\right]=\frac{K^2(1)}{n\phi(x,h)}\sum_{i=1}^{n}\phi_i(x,h)-\int_0^1\left(K^2(u)\right)'\frac{\phi(x,uh)}{n\phi(x,h)\phi(x,uh)}\sum_{i=1}^{n}\phi_i(x,uh)du$$
$$=K^2(1)-\int_0^1\left(K^2(u)\right)'\tau_x(u)du+o_p(1)=\beta_2+o_p(1),$$

and

$$\frac{1}{n\phi(x,h)}\mathbb{E}\left[K_1(x)\right]=\beta_1+o(1).$$

We deduce that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\eta_{ni}^2\mid\mathcal{F}_{i-1}\right]=1,$$

which completes the proof of (2.7)
Concerning (2.8), we write

$$\zeta_{ni}^2\mathbb{1}_{\zeta_{ni}^2>\epsilon n}\leq\frac{\left|\zeta_{ni}\right|^{2+\delta}}{\sqrt{(\epsilon n)^\delta}}\text{ for every }\delta>0.$$

Notice that

$$\mathbb{E}\left[\zeta_{ni}^{2+\delta}\right] = \mathbb{E}\left[|\eta_{ni}(x) - \mathbb{E}\left[\eta_{ni}(x) \mid \mathcal{F}_{i-1}\right]|^{2+\delta}\right]$$
$$\leq 2^{1+\delta}\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right] + 2^{1+\delta}|\mathbb{E}\left[\mathbb{E}\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right]^{2+\delta}\right]|.$$

Using Jensen's inequality (4.6.1) we obtain

$$\mathbb{E}\left[\zeta_{ni}^{2+\delta}\right] \leq C\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right].$$

Accordingly, it remains to evaluate $\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right]$. For this, once again we use the C_r-inequality (4.6.4). We obtain

$$\mathbb{E}\left[|\eta_{ni}(x)|^{2+K}\right] \leq C\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)\mathbb{E}^2\left[K_1\right]}\right)^{1+\delta/2}\mathbb{E}\left[K_i^{2+\delta}(x)\psi^{2+\delta}\left(Y_i,t\right)\right] + \Psi^{2+\delta}(x,z)\mathbb{E}\left[K_i^{2+\delta}\right].$$

We condition on X_i, using the fact that

$$\mathbb{E}\left[\psi^{2+\delta}\left(Y_i,t\right) \mid X_i\right] < \infty.$$

To obtain

$$\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right] \leq C\left(\frac{1}{\phi(x,h)}\right)^{1+\delta/2}\mathbb{E}\left(\left[K_i(x)\right]^{2+\delta}\right) \leq C\left(\frac{1}{\phi(x,h)}\right)^{\delta/2}. \tag{2.11}$$

Consequently,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\zeta_{ni}^2\mathbb{1}_{\zeta_{ni}^2>\epsilon n}\right] \leq C\left(\frac{1}{n\phi(x,h)}\right)^{\delta/2} \to 0,$$

and the proof is complete. □

Lemma 2.3.2. *[46] Under Hypotheses (H'1) and (H'4)-(H'6), we have,*

$$\widehat{\Psi}_D(x) - 1 = o_P(1).$$

□

Lemma 2.3.3. *[7] Under hypotheses (H'1), (H'2), and (H'4)-(H'6) we have*

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x\right)}\right)^{1/2}B_n(x,z) = u + o(1), \quad as \quad n \to \infty.$$

Proof. From a simple manipulation we obtain

$$\frac{\bar{\Psi}_N(x,z)}{\bar{\Psi}_D(x)} = \frac{1}{\sum_{i=1}^{n}\mathbb{E}\left[K_i(x) \mid \mathcal{F}_{i-1}\right]}\sum_{i=1}^{n}\mathbb{E}\left[K_i\left[\mathbb{E}\left[\psi_x\left(Y,z\right) \mid X_1\right] - \mathbb{E}\left[\psi_x\left(Y,z\right) \mid X=x\right]\right] \mid \mathcal{F}_{i-1}\right]$$
$$+\mathbb{E}\left[\psi_x\left(Y,z\right) \mid X=x\right] - \mathbb{E}\left[\psi_x\left(Y,\theta(x)\right) \mid X=x\right] = J_1 + J_2. \tag{2.12}$$

For $J_1(x)$ the main basic idea of the proof is to take the same ideas as in Ferraty et al. (2007)[23]. Under (H'2(iii)), we obtain

$$\begin{aligned} A_i &= \mathbb{E}[K_i[\mathbb{E}[\psi_x(Y,z) \mid X_i] - \mathbb{E}[\psi_x(Y,z) \mid X = x]] \mid \mathcal{F}_{i-1}] \\ &= \mathbb{E}[K_i[\mathbb{E}[\Psi(X_i,z) - \Psi(x,z) \mid d(x,X_i)] \mid \mathcal{F}_{i-1}]] \\ &= \mathbb{E}[K_i \Phi(d(x,X_i),z) \mid \mathcal{F}_{i-1}] \\ &= \int \Phi(th,z) K(t) d\mathbb{P}^{\mathcal{F}_{i-1}}(th) \\ &= h\Phi'(0,z) \int tK(t) d\mathbb{P}^{\mathcal{F}_{i-1}}(th). \end{aligned} \tag{2.13}$$

We use the continuity of $\Phi'(0,\cdot)$ and the fact that

$$\int tK(t) d\mathbb{P}^{\mathcal{F}_{i-1}}(th) = K(1)\phi_i(x,h) - \int_0^1 (sK(s))'\phi_i(x,sh)ds,$$

to obtain

$$\frac{1}{n}\sum_{i=1}^n A_i = h\Phi'(0,\theta_x)\left(K(1) - \int_0^1 (sK(s))'\tau_x(s)ds\right) + o_p(h).$$

In similar fashion, we have

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[K_i(x) \mid \mathcal{F}_{i-1}\right] = \left(K(1) - \int_0^1 K'(s)\tau_x(s)ds\right) + o_p(1).$$

Finally,

$$J_1 = B_n(x) + o(h).$$

Concerning J_2 we use a Taylor expansion to get, under (H'2)

$$J_2 = -B_n(x) + u[n\phi(x,h)]^{-1/2}\sigma(x,\theta_x)\frac{\partial}{\partial t}\Psi(x,\theta_x) + o\left([n\phi(x,h)]^{-1/2}\right).$$

The result is then a consequence of the decomposition in (2.12).

Lemma 2.3.4. *[7] Under hypotheses (H'1), (H'2), and (H'4)-(H'6) we have,*

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} R_n(x,z) = o_P(1) \quad a.co.$$

Proof. Clearly, it suffices to show that

$$\frac{\bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} = o_p(1),$$

and

$$\left|\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)\right| = o_p(1).$$

On the one hand

$$\frac{\bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} =$$

$$=\frac{1}{n\mathbb{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i(x)\mathbb{E}\left[\psi\left(Y_i,t\right)\mid\mathcal{G}_{i-1}\right]\mid\mathcal{F}_{i-1}\right]-\Psi(x,t)\mathbb{E}\left[K_i(x)\mid\mathcal{F}_{i-1}\right]\right]$$
$$=\frac{1}{n\mathbb{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i(x)\mathbb{E}\left[\psi\left(Y_i,t\right)\mid X_i\right]\mid\mathcal{F}_{i-1}\right]-\Psi(x,t)\mathbb{E}\left[K_i(x)\mid\mathcal{F}_{i-1}\right]\right]$$
$$\leq\frac{1}{n\mathbb{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i(x)\left|\Psi\left(X_i,t\right)-\Psi(x,t)\right|\mid\mathcal{F}_{i-1}\right]\right].$$

Using (H'1(ii)),we deduce that

$$\left|\frac{\bar{\Psi}_N(x,t)-\Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)}\right|\leq\sup_{x'\in B(x,h)}\left|\Psi\left(x',t\right)-\Psi(x,t)\right|\rightarrow 0.$$

On the other hand,

$$\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)=o_p(1).$$

Our next aim is to show the following two results :

$$\mathbb{E}\left[\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)\right]\rightarrow 0,$$

and

$$\operatorname{Var}\left[\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)\right]\rightarrow 0.$$

The first one is an outcome of the definitions of $\widehat{\Psi}_N(x,z)$ and $\bar{\Psi}_N(x,z)$. Next, for the second one, we have

$$\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)=\sum_{i=1}^{n}\Delta_i(x,z),$$

where

$$\Delta_i(x,z)=\frac{1}{n\mathbb{E}\left[K_1\right]}K_i\psi\left(Y_i,z\right)-\mathbb{E}\left[K_i\psi\left(Y_i,z\right)\mid\mathcal{F}_{i-1}\right].$$

By Burkholder's inequality (4.6.1), we have

$$\mathbb{E}\left[\sum_{i=1}^{n}\Delta_i(x,z)\right]^2\leq\sum_{i=1}^{n}\mathbb{E}\left[\Delta_i(x,z)\right]^2.$$

Furthermore, through Jensen's inequality (4.6.1) we show that

$$\mathbb{E}\left[\Delta_i(x,z)\right]^2\leq\frac{1}{n^2\mathbb{E}^2\left[K_1\right]}\mathbb{E}\left[K_i^2\psi^2\left(Y_i,z\right)\right]\leq\frac{1}{n^2\mathbb{E}^2\left[K_1\right]}\mathbb{E}\left[K_i^2\right]\leq\frac{1}{n\phi^2(x,h)}\phi_i(x,h).$$

Now, (H'1) gives

$$\operatorname{Var}\left[\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)\right]\rightarrow 0.$$

□

Lemma 2.3.5. *[7] Under Hypotheses (H'1) and (H'4)-(H'6),* $\widehat{\theta}_x$ *exists a.s. for all sufficiently large* n

Proof. It is clear that the proof of lemma (2.3.5) is similar to the proof of lemma (2.2.4). □

Chapter 3

Nonparametric robust regression for right censored with stationary ergodic data

The present chapter deals with a nonparametric robust regression for right censored when the covariate takes values in $\mathbb{R}^d (d \geq 1)$ and the data are sampled from a stationary ergodic process. This chapter is divided into two sections: In the first section, we will give a definition to the model and the estimator. In the other section, we will present the assumptions under which the almost sure consistency (with rate) and the asymptotic distribution of the estimator are established, then we will deal with confidence intervals.

3.1 Definition of the model and the estimator

Consider a triple (X, C, T) of r.v's defined in $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$, where T is the variable of interest (typically a lifetime variable), C a censoring variable and $X = (X_1, ..., X_d)$ a vector of covariates. We refer to $F(\cdot)$ (resp. $G(\cdot)$) the distribution function of T (resp. C) that are supposed to be unknown and continuous. The continuity of G enables as to use the convergence results for the Kaplan and Meier (1958)[37] estimator of G. In the following, we assume that:

$$(T, X) \text{ and } (C) \text{ are independent.} \tag{3.1}$$

In the right censoring model, we don't directly observe the pair (T, C) with $Y = \min(T, C)$ and $\delta = \mathbb{1}_{\{T \leq C\}}$. Consequently, we presume that a sample $\{(X_i, \delta_i, Y_i), i = 1, \ldots, n\}$ is at our disposal. In addition, we suppose throughout the chapter that $(X_i, T_i)_{i=1,\ldots,n}$ is a strictly stationary ergodic sequence, in the sense that satisfies the theorem (1.3.1), and $(C_i)_i$ is a sequence of (i.i.d.) r.v's which is independent of $(X_i, T_i)_{i=1,\ldots,n}$. It is noticeable that, the unobserved sample $(X_i, T_i, C_i)_{i=1,\ldots,n}$ is ergodic by continuity of the identity application. Let us also consider $\mathfrak{G} : \mathbb{R}^3 \to \mathbb{R} \times \{0, 1\} \times \mathbb{R}$ be a measurable application defined by $(T, C, X) \mapsto (Y, \delta, X)$. Clearly, by the multidimensional ergodic theorem, the observed sample is stationary and ergodic one, as soon as the unobserved sample is. Now put

$$Z = \frac{\delta Y}{\overline{G}(Y)}, \quad \text{where} \quad \overline{G}(\cdot) = 1 - G(\cdot). \tag{3.2}$$

Then from (3.1) and (3.2) we get

$$
\begin{aligned}
\mathbb{E}(Z \mid X) &= \mathbb{E}\left[\frac{\delta Y}{\overline{G}(Y)} \mid X\right] \\
&= \mathbb{E}\left\{\mathbb{E}\left[\frac{\mathbb{1}_{\{T\leq C\}}T}{\overline{G}(T)} \mid X, T\right] \mid X\right\} \\
&= \mathbb{E}\left\{\frac{T}{\overline{G}(T)}\mathbb{E}\left(\mathbb{1}_{\{T\leq C\}} \mid X, T\right) \mid X\right\} \\
&= \mathbb{E}(T \mid X).
\end{aligned}
\tag{3.3}
$$

As a result, any estimator for the regression function $\mathbb{E}(Z|X=x)$, which can be derived from fully observed data(Y_i, C_i) , turns out to be an estimator for the regression function $\mathbb{E}(T|X=x)$ based on the unobserved data.
In order to define the ψ-regression function under the right censoring model, consider (as in (3.2)), the ψ^*-function

$$\psi^*(T-\theta) = \frac{\delta\psi(T-\theta)}{\overline{G}(T)}. \tag{3.4}$$

The issue being found the parameter $\theta_\psi(x)$ which is a zero w.r.t. θ of

$$\Psi(x,\theta) = \mathbb{E}\left[\frac{\delta\psi(T-\theta)}{\overline{G}(T)} \mid X = x\right] = 0. \tag{3.5}$$

Note here that the ψ^*-function, given by (3.4), inherits the monotony property from $\psi(T-\cdot)$. According to the observed sample $(X_i, \delta_i, Y_i)_{i=1,\dots,n}$, we define the following "pseudo-estimator" of $\Psi(x,\theta)$, which will be used as intermediate estimator

$$\widetilde{\Psi}_n(x,\theta) = \frac{\sum_{i=1}^{n} K\left(h^{-1}(x-X_i)\right)\delta_i\left(\overline{G}(Y_i)\right)^{-1}\psi(Y_i-\theta)}{\sum_{i=1}^{n} K\left(h^{-1}(x-X_i)\right)} = \frac{\widetilde{\Psi}_N(x,\theta)}{\widehat{\Psi}_D(x)}, \tag{3.6}$$

where

$$\widetilde{\Psi}_N(x,\theta) = \frac{1}{n\mathbb{E}(K_1(x))}\sum_{i=1}^{n} K_i(x)\frac{\delta_i\psi(Y_i-\theta)}{\overline{G}(Y_i)} \quad \text{and} \quad \widehat{\Psi}_D(x) = \frac{1}{n\mathbb{E}(K_1(x))}\sum_{i=1}^{n} K_i(x), \tag{3.7}$$

with $K_i(x) = K\left(h^{-1}(x-X_i)\right)$. Since G is unknown in practice, and in order to get a feasible estimator, we replace $\overline{G}(\cdot)$ by its Kaplan and Meier (1958) estimator $\overline{G}_n(\cdot)$ given by (1.2). Thus, an estimator of $\Psi(x,\theta)$ is given by

$$\widehat{\Psi}_n(x,\theta) = \frac{\widehat{\Psi}_N(x,\theta)}{\widehat{\Psi}_D(x)}, \tag{3.8}$$

where the denominator has the same expression as in (3.7) and the numerator is analogous to $\widetilde{\Psi}_N(x,\theta)$ by replacing $\overline{G}(\cdot)$ by $\overline{G}_n(\cdot)$.
Therefore, an estimate of the ψ-regression function $\theta_\psi(x)$ denoted by $\widehat{\theta}_{\psi,n}(x)$, may be defined as a zero w.r.t. θ of the equation $\widehat{\Psi}_n(x,\theta) = 0$, which satisfies

$$\widehat{\Psi}_n\left(x, \widehat{\theta}_{\psi,n}(x)\right) = 0. \tag{3.9}$$

3.2 Assumptions and main results

The introduction of some notation will be the way to present our findings. Let $\mathcal{F}_i$ be the σ-field generated by $((X_1,T_1),...,(X_i,T_i))$ and $\mathcal{G}_i$ the one generated by $((X_1,T_1),...,(X_i,T_i),X_{i+1})$. Set $\|\cdot\|_2$ for the Euclidean norm and $\|\cdot\|$ the supremum norm in $\mathbb{R}^d$. For any fixed x in $\mathbb{R}^d$ and for $r>0$, denoted by $S_{r,x}=\{v : \|v-x\|\leq r\}$ the sphere of radius r centered at x. For any Borel set $A\subset\mathbb{R}^d$, set $\mathbb{P}^{\mathcal{F}_{i-1}}_{X_i}(A)=\mathbb{P}\left(X_i\in A\mid \mathcal{F}_{i-1}\right)$. For some $\tau>0$, let $\mathcal{C}_{\psi,\tau}$ be a subset of $\mathbb{R}$ and $\mathcal{N}_x$ be a neighborhood of x. Denote by $\mathrm{O}_{a.s.}(u)$ a real random function g such that $g(u)/u$ is almost surely bounded as u goes to zero. In what follows, for any distribution function $L(\cdot)$, let $\zeta_L=\sup\{t,\text{ such that } L(t)<1\}$ be the support's right endpoint. We suppose that $\theta_\psi(x)\in\mathcal{C}_{\psi,\tau}\cap(-\infty,\zeta]$, where $\zeta<\zeta_G\wedge\zeta_F$.

We will make use of the following assumptions gathered here for easy reference.

(H1) There exist a constant c_0 and a nonnegative bounded random function $f_i(x,w)=f_i(x)(\text{resp.} f(x))$, $w\in\Omega$, defined on $\mathbb{R}^d$, such that, for any $i\geq 1$,

$$\mathbb{P}^{\mathcal{F}_{i-1}}_{X_i}\left(S_{r,x}\right)=c_0 f_i(x) r^d \quad \text{a.s.} \quad \text{as} \quad r\to 0, \tag{3.10}$$

$$\text{and } \mathbb{P}_{X_i}\left(S_{r,x}\right)=c_0 f(x) r^d \quad \text{as} \quad r\to 0. \tag{3.11}$$

(H2) For any $x\in\mathbb{R}^d$, and $j=1,2$, $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} f_i^j(x)=f^j(x) \quad$ a.s.

(H3) $\{h_n\}$ is a non-increasing sequence of positive constants such that
(i) $h_n\to 0 \quad$ and $\quad \log n/nh_n^d\to 0 \quad$ as $\quad n\to\infty$.

(ii) $h_n^d\ln\ln n\to 0 \quad$ and $\quad nh_n^{1+2\alpha_1 d}\to 0 \quad$ as $\quad n\to\infty \quad$ for some $\quad \alpha_1>0$.

(H4) (i) K is a spherically symmetric density function with a spherical bounded support. That is, there exists $k:\mathbb{R}^+\to\mathbb{R}$ satisfying $k(0)>0$, $k(v)=0$ for $v>1$, and $K(x)=k\left(\|x\|_2\right)$. Furthermore, for $j\geq 1$, $k^j(\cdot)$ is of class $\mathcal{C}^1$.
(ii) For any $m\geq 1$, $\int_0^1 k^m(u)u^{d-1}du\leq 1/d$.

(H5) The function Ψ given by (3.5) is such that :
(i) $\Psi(x,\cdot)$ is of class $\mathcal{C}^1$ on $\mathcal{C}_{\psi,\tau}$.
(ii) For each fixed $\theta\in\mathcal{C}_{\psi,\tau}$, $\Psi(\cdot,\theta)$ is continuous at the point x.
(iii) $\forall\left(\theta_1,\theta_2\right)\in\mathcal{C}^2_{\psi,\tau}$ and $\forall\left(x_1,x_2\right)\in\mathcal{N}_x^2$, the derivative with respect to θ of order p ($p\in\{0,1\}$) $\Psi^{(p)}(\cdot,\cdot)$ of $\Psi(\cdot,\cdot)$ satisfying

$$\left|\Psi^{(p)}\left(x_1,\theta_1\right)-\Psi^{(p)}\left(x_2,\theta_2\right)\right|\leq\gamma_1\left\|x_1-x_2\right\|_2^{\alpha_1}+\gamma_2\left|\theta_1-\theta_2\right|^{\alpha_2},$$

for $\quad \alpha_1>0$ and $\alpha_2>0$.

(H6) The function ψ is such that :
(i) For any fixed $\theta\in\mathcal{C}_{\psi,\tau}$ and any $j\geq 1$
$\mathbb{E}\left[\left(\psi^{(p)}\left(T_i-\theta\right)\right)^j\mid\mathcal{G}_{i-1}\right]=\mathbb{E}\left[\left(\psi^{(p)}\left(T_i-\theta\right)\right)^j\mid X_i\right]<\gamma_3 j!<\infty \quad$ a.s. $\quad$ with $\quad p\in\{0,1\}$ and $\gamma_3>0$.
(ii) $\psi(\cdot)$ is strictly monotone, bounded, continuously differentiable such that: $\forall\theta\in\mathbb{R}$, $|\psi'(\cdot-\theta)|>\gamma_4>0$ with γ_4 is a positive constant.

(H7) (i) For any $(x_1, x_2) \in \mathcal{N}_x^2$, the function

$$\mathbb{M}(x,\theta) = \mathbb{E}\left(\frac{\psi^2(T-\theta)}{\overline{G}(T)} \mid X = x\right), \quad x \in \mathbb{R}^d \tag{3.12}$$

satisfies

$$|\mathbb{M}(x_1,\theta) - \mathbb{M}(x_2,\theta)| \leq \gamma_5 \|x_1 - x_2\|^{\alpha_3}$$

γ_5 and α_3 are some nonnegative constants uniformly on θ.
(ii) For $m = 1, 2, \mathbb{E}\left(|\delta_1 \overline{G}^{-1}(T)\psi(T-\theta)|^m\right) < \infty$ and for any $x \in \mathbb{R}^d$, the conditional variance of $\delta_1 \overline{G}^{-1}(T)\psi(T-\theta)$ given $X = x$ exists, that is

$$\mathbb{V}(\theta \mid x) = \mathbb{E}\left[\left(\delta_1 \overline{G}^{-1}(T)\psi(T-\theta) - \Psi(x,\theta)\right)^2 \mid X = x\right] \tag{3.13}$$

for every θ.

(H8) (i) The conditional variance of $\delta_i \overline{G}^{-1}(T_i)\,\psi(T_i - \theta)$ given the σ-field $\mathcal{G}_{i-1}$ depends only on X_i, i.e., for any $i \geq 1, \mathbb{E}\left[\left(\delta_i \overline{G}^{-1}(T_i)\,\psi(T_i-\theta) - \Psi(X_i,\theta)\right)^2 \mid \mathcal{G}_{i-1}\right] = \mathbb{V}(\theta \mid X_i)$ almost surely.
(ii) For some $\varsigma > 0, \mathbb{E}\left(\left|\delta_1 \overline{G}^{-1}(T_1)\,\psi(T_1-\theta)\right|^{2+\varsigma}\right) < \infty$ and the function

$$\widetilde{\mathbb{V}}_{2+\varsigma}(\theta \mid u) = \mathbb{E}\left[\left(\delta_i \overline{G}^{-1}(T_i)\,\psi(T_i-\theta) - \Psi(X_i,\theta)\right)^{2+\varsigma} \mid X_i = u\right], \quad u \in \mathbb{R}^d \tag{3.14}$$

is continuous over $\mathcal{N}_x$.

Comments on the hypotheses

The condition (H1) significate that the conditional probability of the d-dimensional ball, given the σ-field $\mathcal{F}_{i-1}$, is asymptotically controlled by a local dimension when the radius r tends to zero. This assumption may be interpreted in terms of the fractal dimension, which ensures that our results are established without assuming the existence of marginal and conditional densities. It obvious that whenever the $\mathbb{P}_{X_i}^{\mathcal{F}_{i-1}}$ has a continuous conditional density $f_{X_i}^{\mathcal{F}_{i-1}}(x) = f_i(x)$ at any point of the set $\{x : f_{X_i}^{\mathcal{F}_{i-1}} > 0\}$. then the constant c_0 becomes $\pi^{d/2}\Gamma((d+2)/2)$ where Γ stands for the Gamma function. The assumption (H2) means that the random function $f_i(\cdot)$ completed the ergodic property. The first condition of (H3) is used to establish the pointwise consistency rate of the estimator, the second condition used to vanish the bias term when dealing with the asymptotic normality. Assumption (H4) is meant to be an alternative to the use of a product to construct a multivariate function. The function $k(\|x\|_2)$ may be displayed as

$$k(\|x\|_2) = \frac{K(\|x\|_2)}{\int_{[0,1]^d} K(\|u\|)\,du} = c_d K(\|x_2\|).$$

Where c_d is the normalizing constant.
Condition (H5) deals with some regularities of $\Psi(\cdot)$. Condition (H6)(i) is a standard assumption over j^{th} moments of the conditional expectation of $\psi(\cdot)$. The existence and uniqueness of the (3.5) solution is ensured by (ii). Assumption (H7)(i) is a classical condition to obtain the asymptotic normality. On the other hand (H7)(ii) is necessary to define the conditional variance given in (3.13). Assumption (H8)(i) is a Markov condition and (H8)(ii) ensures the existence of the quantity defined in (3.14).

3.2.1 Pointwise consistency with rate

This result states the pointwise consistency with rate of the M-estimator $\widehat{\theta}_{\psi,n}(x)$.

Theorem 3.2.1. *[14] Suppose that* $(H1)$-$(H6)$ *are satisfied, then* $\widehat{\theta}_{\psi,n}(x)$ *exists and is unique a.s., and*

$$\widehat{\theta}_{\psi,n}(x) - \theta_{\psi}(x) = \mathrm{O}_{a.s.}\left(h_n^{d\alpha_1}\right) + \mathrm{O}_{a.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$$

Proof. To prove the theorem some additional notations are needed. Let

$$\bar{\bar{\Psi}}_N(x,\theta) = \frac{1}{n\mathbb{E}\left(K_1(x)\right)} \sum_{i=1}^{n} \mathbb{E}[K_i(x)\frac{\delta_i \psi\left(Y_i - \theta\right)}{\overline{G}\left(Y_i\right)}|\mathcal{F}_{i-1}] \quad \text{and} \quad \bar{\bar{\Psi}}_D(x) = \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^{n} \mathbb{E}[K_i|\mathcal{F}_{i-1}]. \tag{3.15}$$

Where

$$B_n(x,\theta) = \frac{\bar{\bar{\Psi}}_N(x,\theta)}{\bar{\bar{\Psi}}_D(x)} - \Psi(x,\theta), \tag{3.16}$$

is the pseudo-conditional bias of $\widetilde{\Psi}_n(x,\theta)$.
And

$$R_n(x,\theta) = -B_n(x,\theta)(\widehat{\Psi}_D(x) - \bar{\bar{\Psi}}_D(x)), \tag{3.17}$$

$$Q_n(x,\theta) = (\widetilde{\Psi}_N(x,\theta) - \bar{\bar{\Psi}}_N(x,\theta)) - \Psi(x,\theta)(\widehat{\Psi}_D(x) - \bar{\bar{\Psi}}_D(x)). \tag{3.18}$$

Thus, from (3.15), (3.16), (3.17) and (3.18) we have:

$$\widetilde{\Psi}(x,\theta) - \Psi(x,\theta) = B_n(x,\theta) + \frac{R_n(x,\theta)}{\widehat{\Psi}_D(x)} + \frac{Q_n(x,\theta)}{\widehat{\Psi}_D(x)}. \tag{3.19}$$

As, $\Psi(x,\theta_{\psi}(x)) = 0$ and $\widehat{\Psi}_n(x,\widehat{\theta}_{\psi,n}(x)) = 0$, then through Taylor's expansion of $\Psi(x,\cdot)$ around $\theta_{\psi}(x)$ leads to

$$\begin{aligned} \widehat{\theta}_{\psi,n}(x) - \theta_{\psi}(x) &= \frac{\Psi(x,\widehat{\theta}_{\psi,n}(x))}{\Psi'(x,\theta_n^{\star})} \\ &= \frac{\Psi(x,\widehat{\theta}_{\psi,n}(x)) - \widehat{\Psi}_n(x,\widehat{\theta}_{\psi,n}(x))}{\Psi'(x,\theta_n^{\star})}, \end{aligned} \tag{3.20}$$

where $\theta_n^{\star}$ is between $\widehat{\theta}_{\psi,n}(x)$ and $\theta_{\psi}(x)$. It results from Assumption (H6)(ii) that $|\Psi'\left(x,\theta_n^{\star}\right)| > \gamma$ and therefore

$$\left|\widehat{\theta}_{\psi,n}(x) - \theta_{\psi}(x)\right| = \mathrm{O}_{a.s.}\left(\sup_{\theta\in\mathcal{C}_{\psi,\tau}} \left|\widehat{\Psi}_n(x,\theta) - \Psi(x,\theta)\right|\right). \tag{3.21}$$

Finally, from the following proposition which provides the almost sure consistency with rate of the estimator $\widehat{\Psi}_n(\cdot,\cdot)$ uniformly w.r.t. the second component. We proof the Theorem (3.2.1). □

Proposition 3.2.1. *Assume that* $(H1)$-$(H6)$ *hold true, we have, for some* $\alpha_1 > 0$.

$$\sup_{\theta\in\mathcal{C}_{\psi,\tau}} \left|\widehat{\Psi}_n(x,\theta) - \Psi(x,\theta)\right| = \mathrm{O}_{a.s.}\left(h_n^{d\alpha_1}\right) + \mathrm{O}_{a.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$$

Proof. Using the following decomposition

$$\begin{aligned}\widehat{\Psi}_n(x,\theta)-\Psi(x,\theta) &= \left(\widehat{\Psi}_n(x,\theta)-\widetilde{\Psi}_n(x,\theta)\right)+\left(\widetilde{\Psi}_n(x,\theta)-\Psi(x,\theta)\right)\\ &= \left(\widehat{\Psi}_n(x,\theta)-\widetilde{\Psi}_n(x,\theta)\right)+B_n(x,\theta)+\frac{R_n(x,\theta)+Q_n(x,\theta)}{\widehat{\Psi}_D(x)}.\end{aligned} \tag{3.22}$$

The proof of Proposition (3.2.1) is divided into some lemmas, the convergence of $\widehat{\Psi}_D(x)$ to 1 and the convergence of the terms $B_n(x)$, $R_n(x)$ and $Q_n(x)$. □

We consider now the following technical lemma, which plays the same role as Bochner Theorem.

Lemma 3.2.1. *[14] Assume that Assumptions (H1) and (H4) hold true. For any $j\geq 1$, we have*

(i) $\mathbb{E}\left(\Delta_i^j(x)\mid\mathcal{F}_{i-1}\right)=c_0 f_{i,d}(x)h_n^d d\int_0^1 k^j(u)u^{d-1}du \quad a.s.$

(ii) $\mathbb{E}\left(\Delta_1^j(x)\right)=c_0 f(x)h_n^d d\int_0^1 k^j(u)u^{d-1}du.$

Proof. Using Assumption (H4), we can state

$$\begin{aligned}\mathbb{E}\left(\Delta_i^j(x)\mid\mathcal{F}_{i-1}\right) &= \mathbb{E}\left(k^j\left(\|X_i-x\|_2/h_n\right)\mid\mathcal{F}_{i-1}\right)\\ &= \int_0^{h_n} k^j\left(u/h_n\right)d\mathbb{P}^{\mathcal{F}_{i-1}}\left(\|X_i-x\|_2\leq u\right)\\ &= \int_0^1 k^j(t)d\mathbb{P}^{\mathcal{F}_{i-1}}\left(\|X_i-x\|_2/h_n\leq t\right)\\ &= k^j(1)\mathbb{P}_{X_i}^{\mathcal{F}_{i-1}}\left(S_{h_n,x}\right)-\int_0^1\left(k^j(u)\right)'\mathbb{P}_{X_i}^{\mathcal{F}_{i-1}}\left(S_{uh_n,x}\right)du.\end{aligned}$$

At last, with the Assumption (H1), we can easily concluded the proof of (i). As a result that of (ii) follows from part (i) by considering $\mathcal{F}_{i-1}$ the trivial σ-field. □

The following lemma provides the convergence rate of $\widehat{\Psi}_D(x)$.

Lemma 3.2.2. *[14] Suppose that Assumptions (H1)-(H4) hold true, then we have*

(i) $\lim_{n\to\infty}\widehat{\Psi}_D(x)=\lim_{n\to\infty}\bar{\widehat{\Psi}}_D(x)=1, \quad a.s.$

(ii) $\widehat{\Psi}_D(x)-\bar{\widehat{\Psi}}_D(x)=\mathrm{O}_{a.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$

Proof. First, observe that $\widehat{\Psi}_D(x)-1=R_{n,1}(x)+R_{n,2}(x)$, where

$$R_{n,1}(x)=\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^n\left(K_i(x)-\mathbb{E}\left(K_i(x)\mid\mathcal{F}_{i-1}\right)\right),$$

and

$$R_{n,2}(x)=\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum^n\left[\mathbb{E}\left(K_i(x)\mid\mathcal{F}_{i-1}\right)-\mathbb{E}\left(K_1(x)\right)\right].$$

It can be easily seen that $\mathbb{E}\left(K_i(x)\mid\mathcal{F}_{i-1}\right)-\mathbb{E}\left(K_1(x)\right)=c_0h_n^d d\left[f_i(x)-f(x)\right]\int_0^1 k(u)u^{d-1}du$. And assumption (H2) combined with Lemma (3.2.1) allow to conclude that $R_{n,2}(x)=o(1)$ a.s.

as $n \to \infty$.
To deal with $R_{n,1}(x)$ write

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

Where $L_{n,i}(x) = \left[K_i(x) - \mathbb{E}\left(K_i(x) \mid \mathcal{F}_{i-1}\right)\right]/\mathbb{E}\left(K_1(x)\right)$ is a martingale difference.
Therefore, let us check the condition of Lemma (3.2.1).
Observe that

$$|L_{n,i}(x)| = \frac{|K_i(x) - \mathbb{E}\left(K_i(x) \mid \mathcal{F}_{i-1}\right)|}{|\mathbb{E}\left(K_1(x)\right)|} \leq \frac{2\overline{k}}{c_0 h_n^d d f(x) \int_0^1 k(u)u^{d-1}du} = M,$$

where $\overline{k} = \sup_{x\in\mathbb{R}^+} k(x)$.
On the other hand, making use of Lemma (4.6.1), one can see that

$$\mathbb{E}\left(L_{n,i}^2(x) \mid \mathcal{F}_{i-1}\right) = \frac{\int_0^1 k^2(u)u^{d-1}du}{c_0 h_n^d d f^2(x)\left(\int_0^1 k(u)u^{d-1}du\right)^2} f_i(x) = d_i^2.$$

Then using Lemma 1 in Chaouch et al. (2016)[15], with $D_n = \sum_{i=1}^{n} d_i^2$, we obtain for any $\lambda > 0$

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} L_{n,i}(x)\right| \geq \lambda\right) &\leq 2\exp\left\{-\frac{n^2\lambda^2}{4D_n + 2Mn\lambda}\right\}\\ &= 2\exp\left\{-\frac{n\lambda^2}{4\frac{D_n}{n} + 2M\lambda}\right\}.\end{aligned}$$

Moreover, using Lemma (3.2.1), we get $4\frac{D_n}{n} + 2M\lambda = \frac{C_\lambda(x)}{h_n^d}$, where

$$C_\lambda(x) = \frac{4\left(\int_0^1 k^2(u)u^{d-1}du\right)\left(\int_0^1 k(u)u^{d-1}du\right)^{-2} + 4\lambda\overline{k}\left(\int_0^1 k(u)u^{d-1}du\right)}{c_0 d f(x)}.$$

Consequently,

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} L_{n,i}(x)\right| \geq \lambda\right) \leq 2\exp\left\{-\frac{\lambda^2}{C_\lambda(x)} n h_n^d\right\}.$$

Considering the Assumption (H3), the finding is concluded by Borel-Cantelli Lemma (4.7.1). The quantity $\overline{\widehat{\Psi}}_D(x)$ and $\widehat{\Psi}_D(x)$ are processed in a similar way. The same arguments can be used concerning the item (ii) as the study of $R_{n,1}(x)$. □

The following lemma establishes the convergence almost surely (with rate) of the conditional bias term B_n and the central term as R_n defined in (3.16) and (3.17).

Lemma 3.2.3. *[14] Assume that Assumptions (H1)-(H4), (H5)(iii) and (H6)(i) hold true, then we get*

$$\sup_{\theta\in\mathcal{C}_{\psi,\tau}} |B_n(x,\theta)| = \mathrm{O}_{a.s.}\left(h_n^{d\alpha_1}\right), \tag{3.23}$$

and

$$\sup_{\theta\in\mathcal{C}_{\psi,\tau}} |R_n(x,\theta)| = \mathrm{O}_{a.s.}\left(h_n^{d\alpha_1}\sqrt{\frac{\log n}{n h_n^d}}\right). \tag{3.24}$$

Proof. Recall that

$$B_n(x,\theta) = \frac{\bar{\tilde{\Psi}}_N(x,\theta) - \bar{\tilde{\Psi}}_D(x)\Psi(x,\theta)}{\bar{\tilde{\Psi}}_D(x)}.$$

By a double conditioning w.r.t. the σ-fields $(\mathcal{G}_{i-1}, T_i)$ and $\mathcal{F}_{i-1}$, it follows from (H6)(i) that

$$\begin{aligned}\bar{\tilde{\Psi}}_N(x,\theta) &= \frac{1}{n\mathbb{E}\left(K_1(x)\right)} \sum_{i=1}^{n} \mathbb{E}\left\{K_i(x)\mathbb{E}\left[\frac{\delta_i \psi\left(Y_i-\theta\right)}{\overline{G}\left(Y_i\right)} \mid \mathcal{G}_{i-1}, T_i\right] \mid \mathcal{F}_{i-1}\right\} \\ &= \frac{1}{n\mathbb{E}\left(K_1(x)\right)} \sum_{i=1}^{n} \mathbb{E}\left\{K_i(x)\mathbb{E}\left[\frac{\delta_i \psi\left(Y_i-\theta\right)}{\overline{G}\left(Y_i\right)} \mid X_i, T_i\right] \mid \mathcal{F}_{i-1}\right\}.\end{aligned}$$

Therefore, we obtain

$$\bar{\tilde{\Psi}}_N(x,\theta) - \bar{\tilde{\Psi}}_D(x)\Psi(x,\theta) = \frac{1}{n\mathbb{E}\left(K_1(x)\right)} \sum_{i=1}^{n} \mathbb{E}\left\{K_i(x)\left[\Psi\left(X_i,\theta\right) - \Psi(x,\theta)\right] \mid \mathcal{F}_{i-1}\right\}.$$

Making use of the triangular inequality and Assumption (H5)(iii) we get

$$\bar{\tilde{\Psi}}_N(x,\theta) - \bar{\hat{\Psi}}_D(x)\Psi(x,\theta) = \mathrm{O}_{a.s.}\left(h_n^{d\alpha_1}\right) \times \bar{\hat{\Psi}}_D(x).$$

Finally the use of Lemma (3.2.2) enable us to deduce the consistency rate of $B_n(x,\theta)$ given by the equation (3.23).
Furthermore, since $R_n(x,\theta) = -B_n(x,\theta)\left[\widehat{\Psi}_D(x) - \bar{\tilde{\Psi}}_D(x)\right]$, then equation (3.23) and Lemma (3.2.2) permit to get the consistency rate of $R_n(x,\theta)$ given by the equation (3.24).

□

The following lemma deals with the convergence rate of the numerator $\widetilde{\Psi}_N(x,\theta)$ defined in (3.7) of the pseudo-estimator $\widetilde{\Psi}_n(x,\theta)$.

Lemma 3.2.4. *[14] Under Assumptions (H4) and (H5)(iii) and (3.1), we have*

$$\sup_{\theta\in\mathcal{C}_{\psi,\tau}} \left|\widetilde{\Psi}_N(x,\theta) - \bar{\tilde{\Psi}}_N(x,\theta)\right| = \mathrm{O}_{a.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right). \tag{3.25}$$

Proof. Since $\mathcal{C}_{\psi,\tau} = [\theta_\psi(x)-\tau, \theta_\psi(x)+\tau]$ is a compact set, it admits a covering by a finite number d_n of balls $\mathcal{B}_j(t_j,\ell_j)$ centered at t_j, $1 \le j \le d_n$ satisfies $\ell_n = n^{-1/2}$ and $d_n = \mathrm{O}(n^{-\alpha_2/2})$. Therefore,

$$\begin{aligned}\sup_{\theta\in\mathcal{C}_{\psi,\tau}} \left|\widetilde{\Psi}_N(x,\theta) - \bar{\tilde{\Psi}}_N(x,\theta)\right| \le & \max_{1\le j\le d_n} \sup_{\theta\in\mathcal{B}_j} \left|\widetilde{\Psi}_N(x,\theta) - \widetilde{\Psi}_N\left(x,t_j\right)\right| + \max_{1\le j\le d_n} \left|\widetilde{\Psi}_N\left(x,t_j\right) - \bar{\tilde{\Psi}}_N\left(x,t_j\right)\right| \\ &+ \max_{1\le j\le d_n} \sup_{\theta\in\mathcal{B}_j} \left|\bar{\tilde{\Psi}}_N\left(x,t_j\right) - \bar{\tilde{\Psi}}_N(x,\theta)\right| \\ =& \mathcal{I}_{n,1} + \mathcal{I}_{n,2} + \mathcal{I}_{n,3}.\end{aligned} \tag{3.26}$$

Consistency of the first term $\mathcal{I}_{n,1}$

Using Assumption (H6)(ii), we get

$$\left|\widetilde{\Psi}_N(x,\theta)-\widetilde{\Psi}_N\left(x,t_j\right)\right| \leq \frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^{n} K_i(x)\frac{\delta_i}{\overline{G}\left(Y_i\right)}\left|\psi\left(Y_i-\theta\right)-\psi\left(Y_i-t_j\right)\right|$$
$$\leq \frac{\ell_n}{\overline{G}\left(\tau_F\right)}\widehat{\Psi}_D(x).$$

Therefore, through Lemma (3.2.2) $\widehat{\Psi}_D(x)=\mathrm{O}_{a.s.}(1)$ and $\overline{G}\left(\tau_F\right)>0$, we have

$$\mathcal{I}_{n,1}=\mathrm{O}_{a.s.}\left(\sqrt{\frac{1}{n}}\right). \tag{3.27}$$

Consistency of the first term $\mathcal{I}_{n,3}$

Using a double conditioning with respect to the σ-algebra $\mathcal{G}_{i-1}$ and the definition of $\Psi(x,\theta)$ given by equation (3.5), it is unproblematic to obtain

$$\bar{\widetilde{\Psi}}_N\left(x,t_j\right)-\bar{\widetilde{\Psi}}_N(x,\theta)=\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\sum_{i=1}^{n}\mathbb{E}\left\{K_i(x)\left[\Psi\left(X_i,t_j\right)-\Psi\left(X_i,\theta\right)\right]\mid\mathcal{F}_{i-1}\right\}.$$

Then using Assumption (H5)(iii) and Lemma (3.2.2) we have

$$\mathcal{I}_{n,3}=\mathrm{O}_{\mathrm{a.s.}}\left(n^{-\alpha_2/2}\right). \tag{3.28}$$

Consistency of the first term $\mathcal{I}_{n,2}$

First, consider that $\mathcal{I}_{n,2}$ can be written as

$$\mathcal{I}_{n,2}=\max_{1\leq j\leq d_n}\left|\widetilde{\Psi}_N\left(x,t_j\right)-\bar{\widetilde{\Psi}}_N\left(x,t_j\right)\right|$$
$$=\max_{1\leq j\leq d_n}\frac{1}{n\mathbb{E}\left(K_1(x)\right)}\left|\sum_{i=1}^{n}S_{n,i}\left(x,t_j\right)\right|,$$

where $S_{n,i}\left(x,t_j\right)=K_i(x)\frac{\delta_i\psi(Y_i-t_j)}{\overline{G}(Y_i)}-\mathbb{E}\left[K_i(x)\frac{\delta_i\psi(Y_i-t_j)}{\overline{G}(Y_i)}\mid\mathcal{F}_{i-1}\right]$ is a martingale difference. Thus, we can use Lemma(4.6.1) to obtain an exponential upper bound relative to the quantity $\widetilde{\Psi}_N\left(x,t_j\right)-\bar{\widetilde{\Psi}}_N\left(x,t_j\right)$. The condition under which lemma (4.6.1) is allowed to be applied are verified in what follows. for any $p\in\mathbb{N}-\{0\}$, that

$$S_{n,i}^{p}\left(x,t_j\right)=\sum_{k=0}^{p}C_p^k\left(\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-t_j\right)K_i(x)\right)^k(-1)^{p-k}\left[\mathbb{E}\left(\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-t_j\right)K_i(x)\mid\mathcal{F}_{i-1}\right)\right]^{p-k}.$$

In view of Assumption (H6)(i), $\left[\mathbb{E}\left(\frac{\delta_i}{\overline{G}(Y_i)}\psi\left(Y_i-t_j\right)K_i(x)\mid\mathcal{F}_{i-1}\right)\right]^{p-k}$ is $\mathcal{F}_{i-1}$-measurable. As a result, using Jensen's inequality(4.6.1), one gets

$$\left|\mathbb{E}\left(S_{n,i}^{p}\left(x,t_j\right)\mid\mathcal{F}_{i-1}\right)\right|\leq\sum_{k=0}^{p}C_p^k\mathbb{E}\left\{\left|\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-t_j\right)K_i(x)\right|^k\mid\mathcal{F}_{i-1}\right\}$$
$$\times\mathbb{E}\left\{\left|\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-t_j\right)K_i(x)\right|^{p-k}\mid\mathcal{F}_{i-1}\right\}.$$

With a double conditioning w.r.t. the σ-field $\mathcal{G}_{i-1}$, Assumption (H6)(i) and (3.1), we get, for any $m \geq 1$, that

$$\begin{aligned}\mathbb{E}\left\{\left|\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-t_j\right)K_i(x)\right|^m \mid \mathcal{F}_{i-1}\right\} &\leq \frac{1}{\left(\overline{G}\left(\tau_F\right)\right)^{m-1}}\mathbb{E}\left\{K_i^m(x)\mathbb{E}\left[\psi^m\left(Y_i-t_j\right)\mid X_i\right]\mid \mathcal{F}_{i-1}\right\}\\ &\leq \frac{\gamma_3 m!}{\left(\overline{G}\left(\tau_F\right)\right)^{m-1}}\mathbb{E}\left(K_i^m(x)\mid \mathcal{F}_{i-1}\right).\end{aligned}$$

Making use of Lemma (3.2.1), we obtain $\mathbb{E}\left(K_i^m(x)\mid \mathcal{F}_{i-1}\right) = c_0 f_i(x) d h_n^d \int_0^1 k^m(u)u^{d-1}du$. It results then from (H4)(ii) that

$$\mathbb{E}\left(\left|\frac{\delta_i\psi\left(Y_i-t_j\right)}{\overline{G}\left(Y_i\right)}K_i(x)\right|^k \mid \mathcal{F}_{i-1}\right)\mathbb{E}\left(\left|\frac{\delta_i\psi\left(Y_i-t_j\right)}{\overline{G}\left(Y_i\right)}K_i(x)\right|^{p-k} \mid \mathcal{F}_{i-1}\right) \leq \gamma_3 f_i^2(x) h_n^{2d}.$$

Finally, we have $\left|\mathbb{E}\left(S_{n,i}^p\left(x,t_j\right)\mid \mathcal{F}_{i-1}\right)\right| \leq 2^p f_i^2(x) h_n^d$.

By taking $d_i^2 = f_i^2(x)h_n^d$ and $D_n = \sum_{i=1}^n d_i^2 = h_n^d\sum_{i=1}^n f_i^2(x)$, one gets, by assumption (H2), that $n^{-1}D_n = h_n^d f(x)$ as $n \to \infty$. Now one can use (4.6.1) with $D_n = \mathrm{O}_{a.s.}\left(nh_n^d\right)$ and $S_n = \sum_{i=1}^n S_{n,i}\left(x,t_j\right)$ to get, for any $\epsilon_0 > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\mathcal{I}_{n,2}\right| > \epsilon_0\sqrt{\frac{\log n}{nh_n^d}}\right) &\leq \sum_{j=1}^{d_n}\mathbb{P}\left(\left|\widetilde{\Psi}_N\left(x,t_j\right)-\bar{\widetilde{\Psi}}_N\left(x,t_j\right)\right| > \epsilon_0\sqrt{\frac{\log n}{nh_n^d}}\right)\\ &= \sum_{j=1}^{d_n}\mathbb{P}\left(\frac{1}{\mathbb{E}\left(K_1(x)\right)}\left|\sum_{i=1}^n S_{n,i}\left(x,t_j\right)\right| > \epsilon_0\sqrt{\frac{\log n}{nh_n^d}}\right)\\ &\leq \sum_{j=1}^{d_n} 2\exp\left\{-\frac{\left(n\mathbb{E}\left(K_1(x)\right)\epsilon_0\right)^2\frac{\log n}{nh_n^d}}{2D_n + 2\gamma_3 n\mathbb{E}\left(K_1(x)\right)\epsilon_0\sqrt{\frac{\log n}{nh_n^d}}}\right\}\\ &\leq 2d_n\exp\left\{-\gamma_3\epsilon_0^2\log n\right\}\\ &= n^{-\gamma_3\epsilon_0^2}n^{-\alpha_2/2}.\end{aligned}$$

Therefore, by taking ϵ_0 such that, the upper bound becomes a general term of a convergence Riemann series, we get $\sum_{n\geq 1}\mathbb{P}\left(\left|\mathcal{I}_{n,2}\right| > \epsilon_0\sqrt{\frac{\log n}{nh_n^d}}\right) < \infty$. The proof can be achieved by Borel-Cantelli Lemma (4.7.1). □

Lemma below study the uniform asymptotic rate of the quantity $Q_n(x,\theta)$.

Lemma 3.2.5. *[14] Under assumptions (H1)-(H4) and (H5)(iii), we have*

$$\sup_{\theta\in\mathcal{C}_{\psi,\tau}}\left|Q_n(x,\theta)\right| = \mathrm{O}_{a.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right). \tag{3.29}$$

Proof. The proof of this Lemma is easily obtained we use Lemmas (3.2.2) and (3.2.4). □

The following result gives a uniform approximation (with rate) of the estimator $\widehat{\Psi}_n(x,\theta)$ by the pseudo-estimator of $\Psi(x,\theta)$.

Lemma 3.2.6. *[14] Assume that assumptions (H1)-(H6) hold true, then we have*

$$\sup_{\theta\in\mathcal{C}_{\psi,\tau}}\left|\widehat{\Psi}_n(x,\theta)-\widetilde{\Psi}_n(x,\theta)\right| = \mathrm{O}_{a.s.}\left(\sqrt{\frac{\log\log n}{n}}\right).$$

Proof.

$$\begin{aligned}\left|\widehat{\Psi}_n(x,\theta)-\widetilde{\Psi}_n(x,\theta)\right| &\leq \frac{1}{n\mathbb{E}\left(K_1(x)\right)\widehat{\Psi}_D(x)}\sum_{i=1}^{n}\left|K_i(x)\delta_i\psi\left(Y_i,\theta\right)\left(\frac{1}{\overline{G}_n\left(Y_i\right)}-\frac{1}{\overline{G}\left(Y_i\right)}\right)\right| \\ &\leq \frac{\sup_{t\in\mathcal{C}_{\psi,\tau}}\left|\overline{G}_n(t)-\overline{G}(t)\right|}{\overline{G}_n(\zeta)}\widetilde{\Psi}_n(x,\theta).\end{aligned}$$

As $\overline{G}_n(\zeta)>0$, in conjunction with the (SLLN) and the (LIL) on the censoring law, the result is an immediate consequence of decomposition (4.6.1), Lemmas (3.2.2),(3.2.3) and (3.2.5).

□

3.2.2 Asymptotic distribution

The following deals with the asymptotic distribution of the $\widehat{\theta}_{\psi,n}(x)$.

Theorem 3.2.2. *[14] Under Assumptions (H1)-(H8), we have*

$$\sqrt{nh_n^d}\left(\widehat{\theta}_{\psi,n}(x)-\theta_\psi(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\Sigma^2\left(x,\theta(x)\right)\right),\qquad as\qquad n\longrightarrow+\infty$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution,* $\mathcal{N}(\cdot,\cdot)$ *the normal distribution,*

$$\Sigma^2\left(x,\theta_\psi(x)\right)=\frac{M\left(x,\theta_\psi(x)\right)}{\left(\Gamma_1\left(x,\theta_\psi(x)\right)\right)^2}\times\frac{\int_0^1 k^2(u)u^{d-1}du}{f(x)d\left(\int_0^1 k(u)u^{d-1}du\right)^2}, \tag{3.30}$$

and

$$\Gamma_1(x,\theta)=\mathbb{E}\left[\psi'(T-\theta)\mid X=x\right]\ . \tag{3.31}$$

Proof. With a Taylor's expansion and the definition of $\theta_\psi(x)$, we get

$$\widehat{\Psi}_n\left(x,\theta_\psi(x)\right)-\Psi\left(x,\theta_\psi(x)\right)=\left(\theta_\psi(x)-\widehat{\theta}_{\psi,n}(x)\right)\widehat{\Psi}'_n\left(x,\theta_n^\star\right).$$

Then, we have

$$\widehat{\theta}_{\psi,n}(x)-\theta_\psi(x)=-\frac{\widehat{\Psi}_n\left(x,\theta_\psi(x)\right)-\Psi\left(x,\theta_\psi(x)\right)}{\widehat{\Psi}'_n\left(x,\theta_n^\star\right)}.$$

Consequently, the asymptotic normality given in Theorem (3.2.2) is going to be cited using the following Proposition (3.2.2) and Lemma (3.2.8) that give provide, accordingly, the asymptotic normality of the numerator and the convergence in probability of the denominator term $\widehat{\Psi}'_n\left(x,\theta_n^\star\right)$ to $\Gamma_1\left(x,\theta_\psi(x)\right)$, respectively.

Proposition 3.2.2. *[14] Suppose that Assumptions (H1)-(H8) hold true, then we have*

$$\sqrt{nh_n^d}\left(\widehat{\Psi}_n(x,\theta)-\Psi(x,\theta)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma^2(x,\theta)\right),\qquad as\qquad n\longrightarrow+\infty$$

where

$$\sigma^2(x,\theta)=\left[\mathbb{M}(x,\theta)-\Psi(x,\theta)\right]\times\frac{\int_0^1 k^2(u)u^{d-1}du}{f(x)d\left(\int_0^1 k(u)u^{d-1}du\right)^2}. \tag{3.32}$$

Proof. The proof is based on the following decomposition

$$\begin{aligned}&\left(nh_n^d\right)^{1/2}\left[\widehat{\Psi}_n(x,\theta)-\Psi(x,\theta)\right]\\&=\left(nh_n^d\right)^{1/2}\left[\left(\widehat{\Psi}_n(x,\theta)-\widetilde{\Psi}_n(x,\theta)\right)\right)+\left(\widetilde{\Psi}_n(x,\theta)-\bar{\widetilde{\Psi}}_n(x,\theta)\right)+\left(\bar{\widetilde{\Psi}}_n(x,\theta)-\Psi(x,\theta)\right)\right]\\&=\mathcal{D}_{1,n}+\mathcal{D}_{2,n}+\mathcal{D}_{3,n}.\end{aligned}$$

First, by Lemma (3.2.6), under (H3)(ii), we have $\mathcal{D}_{1,n}=\mathrm{O}_{a.s.}\left(nh_n^d\sqrt{\log\log n/n}\right)=\mathrm{o}_{a.s.}(1)$. The term $\mathcal{D}_{3,n}$ is equal to $\sqrt{nh_n^d}B_n(x,\theta)$ which is $\mathrm{o}_{a.s.}(1)$ in view of assumption (H3)(ii) and Lemma (3.2.3). Furthermore, consider that $\mathcal{D}_{n,2}=\sqrt{nh_n^d}\left(Q_n(x,\theta)-R_n(x,\theta)\right)/\widehat{\Psi}_D(x)$. The quantity $\sqrt{nh_n^d}R_n(x,\theta)$ converges almost surely to zero when n goes to infinity, using the second part of Lemma (3.2.3) combined with assumption (H3)(ii). Moreover, since by Lemma (3.2.2), $\lim_{n\to\infty}\widehat{\Psi}_D(x)=1$ almost surely, thus using Slutsky's Theorem, the asymptotic normality is given by the central term $\sqrt{nh_n^d}Q_n(x,\theta)$ that is the subject of the following Lemma(3.2.7). □

Lemma 3.2.7. *[14] Suppose that Assumptions (H1)-(H8) are satisfied, then we have*

$$\left(nh_n^d\right)^{1/2}Q_n(x,\theta)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma^2(x,\theta)\right),\qquad \text{as } n\to\infty$$

where

$$\sigma^2(x,\theta)=\left[\mathbb{M}(x,\theta)-\Psi(x,\theta)\right]\times\frac{\int_0^1 k^2(u)u^{d-1}du}{f(x)d\left(\int_0^1 k(u)u^{d-1}du\right)^2}. \tag{3.33}$$

Proof. Let us consider

$$\eta_{ni}=\left(\frac{h_n^d}{n}\right)^{1/2}\left(\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-\theta\right)-\Psi(x,\theta)\right)\frac{K_i(x)}{\mathbb{E}\left(K_1(x)\right)},$$

and define $\xi_{ni}=\eta_{ni}-\mathbb{E}\left(\eta_{ni}\mid\mathcal{F}_{i-1}\right)$. One can see that

$$\left(nh_n^d\right)^{1/2}Q_n(x,\theta)=\sum_{i=1}^n\xi_{ni}, \tag{3.34}$$

where, for any fixed $x\in\mathbb{R}^d$, the summands in (3.34) form a triangular array of stationary martingale differences with respect to σ-field $\mathcal{F}_{i-1}$. The asymptotic normality of $Q_n(x,\theta)$ is assembled by applying the central limit theorem for discrete-time arrays of real-valued martingales (see Hall and Heyde [29]). Never the less, the establishing of the following statements is a must :

(a) $\sum_{i=1}^n\mathbb{E}\left[\xi_{ni}^2\mid\mathcal{F}_{i-1}\right]\xrightarrow{2}\sigma^2(x,\theta)$,

(b) $n\mathbb{E}\left[\xi_{ni}^2\mathbb{1}_{[|\xi_{ni}|>\epsilon]}\right]=\mathrm{o}(1)$ holds for any $\epsilon>0$ (Lindberg condition).

Proof of part (a). Observe that

$$\left|\sum_{i=1}^n\mathbb{E}\left[\eta_{ni}^2\mid\mathcal{F}_{i-1}\right]-\sum_{i=1}^n\mathbb{E}\left[\xi_{ni}^2\mid\mathcal{F}_{i-1}\right]\right|\leq\sum_{i=1}^n\left(\mathbb{E}\left[\eta_{ni}\mid\mathcal{F}_{i-1}\right]\right)^2.$$

By double conditioning with respect to $(\mathcal{G}_{i-1}, T_i)$ and $\mathcal{F}_{i-1}$ and using Assumptions (H2) and (H5)(iii) and Lemma (3.2.1), we obtain

$$\sum_{i=1}^{n} \left(\mathbb{E}\left[\eta_{ni} \mid \mathcal{F}_{i-1}\right]\right)^2 = \mathrm{O}_{a.s.}\left(h_n^{(2\alpha_1+1)d}\right). \tag{3.35}$$

Therefore, the statement of (a) follows then if we show that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[\eta_{ni}^2 \mid \mathcal{F}_{i-1}\right] \xrightarrow{\mathbb{P}} \sigma^2(x,\theta). \tag{3.36}$$

To prove(3.36), observe that (by double conditioning) that

$$\sum_{i=1}^{n} \mathbb{E}\left[\eta_{ni}^2 \mid \mathcal{F}_{i-1}\right] = \frac{h_n^d/n}{\left(\mathbb{E}\left(K_1(x)\right)\right)^2} \sum_{i=1}^{n} \mathbb{E}\left\{K_i^2(x)\mathbb{E}\left[\left(\frac{\delta_i}{\overline{G}\left(Y_i\right)}\psi\left(Y_i-\theta\right) - \Psi(x,\theta)\right)^2 \mid X_i\right] \mid \mathcal{F}_{i-1}\right\}.$$

Using the definition of the conditional variance, one gets

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\delta_i\psi\left(Y_i-\theta\right)}{\overline{G}\left(Y_i\right)} - \Psi(x,\theta)\right)^2 \mid X_i\right] &= \mathrm{Var}\left\{\frac{\delta_i\psi\left(Y_i-\theta\right)}{\overline{G}\left(Y_i\right)} \mid X_i\right\} + \left\{\mathbb{E}\left(\frac{\delta_i\psi\left(Y_i-\theta\right)}{\overline{G}\left(Y_i\right)} \mid X_i\right) - \Psi(x,\theta)\right\}^2 \\ &= \mathcal{L}_{n,1} + \mathcal{L}_{n,2}.\end{aligned}$$

Using again a double conditioning with respect to $(\mathcal{G}_{i-1}, T_i)$, Assumptions (H5)(iii) and (H2) and Lemma (3.2.1), we obtain

$$\frac{h_n^d/n}{\left(\mathbb{E}\left(K_1(x)\right)\right)^2} \sum_{i=1}^{n} \mathbb{E}\left[K_i^2(x)\mathcal{L}_{n,2} \mid \mathcal{F}_{i-1}\right] = \mathrm{O}_{a.s.}\left(h_n^{2\alpha_1 d}\right).$$

Likewise we can show, by Assumptions (H2), (H3), (H5)(iii), (H7) and Lemma (3.2.1), that

$$\begin{aligned}\lim_{n\to\infty} \frac{h_n^d/n}{\left(\mathbb{E}\left(K_1(x)\right)\right)^2} \sum_{i=1}^{n} \mathbb{E}\left[K_i^2(x)\mathcal{L}_{n,1} \mid \mathcal{F}_{i-1}\right] &= \left(\mathbb{M}(x,\theta) - \Psi(x,\theta)\right) \times \frac{\int_0^1 k^2(u)u^{d-1}du}{f(x)d\left(\int_0^1 k(u)u^{d-1}du\right)^2} \\ &= \sigma^2(x,\theta).\end{aligned}$$

Proof of part (b)
The Lindeberg's condition results from Corollary 9.5.2 in Chow and Teicher (1998) which implies that

$$n\mathbb{E}\left[\xi_{ni}^2 \mathbb{1}_{[|\xi_{ni}|>\epsilon]}\right] \leq 4n\mathbb{E}\left[\eta_{ni}^2 \mathbb{1}_{[|\eta_{ni}|>\epsilon/2]}\right].$$

Let $a > 1$ and $b > 1$ such that $1/a + 1/b = 1$. Through Hölder's and Markov's inequalities (4.6.2) (4.6.3) , we write, for all $\epsilon > 0$ $\mathbb{E}\left[\eta_{ni[|\eta_{ni}|>\epsilon/2]}^2\right] \leq \frac{\mathbb{E}|\eta_{ni}|^{2a}}{(\epsilon/2)^{2a/b}}$. Taking C_0 a positive constant and $2a = 2 + \varsigma$, and by Assumption (H8) we get that

$$4n\mathbb{E}\left[\eta_{ni}^2 \mathbb{1}_{[|\eta_{ni}|>\epsilon/2]}\right] \leq C_0 \left(\frac{h_n^d}{n}\right)^{(2+\varsigma)/2} \frac{n\mathbb{E}\left[K_1^{2+\varsigma}(x)\right]}{\left(\mathbb{E}\left(K_1(x)\right)\right)^{2+\varsigma}} \left[\widetilde{\mathbb{V}}_{2+\varsigma}(\theta \mid x) + \mathrm{o}(1)\right].$$

Conclusively Lemma (3.2.1) allows us to state $4n\mathbb{E}\left[\eta_{ni}^2 \mathbb{1}_{[|\eta_{ni}|>\epsilon/2]}\right] = \mathrm{O}_{a.s.}\left(\left(nh_n^d\right)^{-\varsigma/2}\right)$ which complete the proof by using Assumption (H3). □

Lemma 3.2.8. *[14] Under assumptions of Theorem (3.2.2) , we have, uniformly in θ,*

$$\widehat{\Psi}'_n(x,\theta) \to \Gamma_1(x,\theta) \text{ in probablility as } n \to \infty.$$

Proof. The proof of this lemma is based on the following decomposition

$$\left|\widehat{\Psi}'_n(x,\theta^\star) - \Gamma_1(x,\theta(x))\right| \leq \left|\widehat{\Psi}'_n(x,\theta^\star) - \widehat{\Psi}'_n(x,\theta(x))\right| + \left|\widehat{\Psi}'_n(x,\theta(x)) - \Gamma_1(x,\theta(x))\right| \quad (3.37)$$

Concerning the first term, using the fact that δ_i is bounded by one and $\overline{G}_n(Y_i)$ is dominated by $\overline{G}_n(\zeta)$, then one can write

$$\left|\widehat{\Psi}'_n(x,\theta^\star) - \widehat{\Psi}'_n(x,\theta_\psi(x))\right| \leq \sup_{t\in\mathcal{C}_{\psi,\tau}} \left|\frac{\partial\psi(t-\theta^\star)}{\partial\theta} - \frac{\partial\psi(t-\theta_\psi(x))}{\partial\theta}\right| \frac{1}{\overline{G}_n(\zeta)}.$$

Since $\partial\psi(T-\theta)/\partial\theta$ is continuous at $\theta_\psi(x)$ uniformly in t, the use of Theorem (3.2.1) and the convergence in probability of $\overline{G}_n(\zeta_F)$ to $\overline{G}_n(\tau_F)$ permits to conclude that the first term of (3.37) converges in probability to zero.

Considering Assumptions (H1)-(H4), (H5)(iii) and (H6)(i), we demonstrate, by using similar arguments as in the proof of Lemma (3.2.3) , that the second term in the right side of the inequality (3.37) converges almost surely to zero. □

3.2.3 Confidence intervals

Regarding the fact that Theorem (3.2.2) is useless in practice as many quantities in the variance are unknown. By Assumption (H1), the term $f(x)h_n^d$ can be interpreted as the value of the probability that X_i belongs to the sphere of radius h_n and centered at x. In practice this probability might be estimated by $\widehat{\mathbb{P}}_{X_i}(S_{h_n,x}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\in S_{h_n,x}\}}$. Moreover, $\mathbb{M}(x,\theta_\psi(x))$ and $\Gamma_1(x,\theta_\psi(x))$ could be replaced, in practice, by their nonparametric estimators, defined by:

$$\widehat{\mathbb{M}}_n\left(x,\widehat{\theta}_{\psi,n}(x)\right) = \frac{\sum_{i=1}^{n}\psi^2\left(Y_i - \widehat{\theta}_{\psi,n}(x)\right)\left(\overline{G}_n(Y_i)\right)^{-1}K\left(h^{-1}(x-X_i)\right)}{\sum_{i=1}^{n}K\left(h^{-1}(x-X_i)\right)},$$

and

$$\widehat{\Gamma}_{1,n}\left(x,\widehat{\theta}_{\psi,n}(x)\right) = \frac{\sum_{i=1}^{n}\psi'\left(Y_i - \widehat{\theta}_{\psi,n}(x)\right)K\left(h^{-1}(x-X_i)\right)}{\sum_{i=1}^{n}K\left(h^{-1}(x-X_i)\right)}.$$

From these quantities we get the following

Corollary 3.2.1. *[14] Under the assumptions of Theorem (3.2.2), we have, as $n \to \infty$,*

$$\widehat{\Gamma}_{1,n}\left(x,\widehat{\theta}_{\psi,n}(x)\right)\left\{\frac{n\widehat{\mathbb{P}}_{X_i}(S_{h_n,x})\,d}{\widehat{\mathbb{M}}_n\left(x,\widehat{\theta}_{\psi,n}(x)\right)\int_0^1 k^2(u)u^{d-1}du}\right\}^{1/2}\int_0^1 k(u)u^{d-1}du \times \left(\widehat{\theta}_{\psi,n}(x) - \theta_\psi(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Corollary (3.2.1) permits to establish the 100(1 - a)% confidence intervals for the ψ-regression $\theta_\psi(x)$, which are given by

$$\widehat{\theta}_{\psi,n}(x) \pm q_{\alpha/2}\widehat{\Gamma}_{1,n}\left(x,\widehat{\theta}_{\psi,n}(x)\right)\left\{\frac{n\widehat{\mathbb{P}}_{X_i}\left(S_{h_n,x}\right)d}{\widehat{\mathbb{M}}_n\left(x,\widehat{\theta}_{\psi,n}(x)\right)\int_0^1 k^2(u)u^{d-1}du}\right\}^{1/2}\int_0^1 k(u)u^{d-1}du,$$

where $\pm q_{\alpha/2}$ is the upper $\alpha/2$ quantile of the Gaussian distribution. All quantities appearing in the confidence bands are known which make the confidence interval useful in practice.

Proof. Observe that

$$\begin{aligned}\widehat{\Gamma}_{1,n}\left(x,\widehat{\theta}_{\psi,n}(x)\right)\left\{\frac{n\widehat{\mathbb{P}}_{X_i}\left(S_{h_n,x}\right)d}{\widehat{\mathbb{M}}_n\left(x,\widehat{\theta}_{\psi,n}(x)\right)\int_0^1 k^2(u)u^{d-1}du}\right\}^{1/2}\int_0^1 k(u)u^{d-1}du\times\left(\widehat{\theta}_{\psi,n}(x)-\theta_\psi(x)\right) = \\ \left\{\frac{\widehat{\mathbb{P}}_{X_i}\left(S_{h_n,x}\right)}{h_n^d f(x)}\right\}^{1/2}\left\{\frac{\mathbb{M}\left(x,\theta_\psi(x)\right)}{\widehat{\mathbb{M}}_n\left(x,\widehat{\theta}_{\psi,n}(x)\right)}\right\}^{1/2}\left\{\frac{\widehat{\Gamma}_{1,n}\left(x,\widehat{\theta}_{\psi,n}(x)\right)}{\Gamma_1\left(x,\theta_\psi(x)\right)}\right\}\times \\ \Gamma_1\left(x,\theta_\psi(x)\right)\left\{\frac{nh_n^d f(x)d}{\mathbb{M}\left(x,\theta_\psi(x)\right)\int_0^1 k^2(u)u^{d-1}du}\right\}^{1/2}\int_0^1 k(u)u^{d-1}du\times\left(\widehat{\theta}_{\psi,n}(x)-\theta_\psi(x)\right).\end{aligned}$$

Using Theorem (3.2.2), we get

$$\Gamma_1\left(x,\theta_\psi(x)\right)\left\{\frac{nh_n^d f(x)d}{\mathbb{M}\left(x,\theta_\psi(x)\right)\int_0^1 k^2(u)u^{d-1}du}\right\}^{1/2}\int_0^1 k(u)u^{d-1}du\times\left(\widehat{\theta}_{\psi,n}(x)-\theta_\psi(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Therefore, the Corollary (3.2.1) is established if we show that

$$\left\{\frac{\widehat{\mathbb{P}}_{X_i}\left(S_{h_n,x}\right)}{h_n^d f(x)}\right\}^{1/2}\left\{\frac{\mathbb{M}\left(x,\theta_\psi(x)\right)}{\widehat{\mathbb{M}}_n\left(x,\widehat{\theta}_{\psi,n}(x)\right)}\right\}^{1/2}\left\{\frac{\widehat{\Gamma}_{1,n}\left(x,\widehat{\theta}_{\psi,n}(x)\right)}{\Gamma_1\left(x,\theta_\psi(x)\right)}\right\}\xrightarrow{\mathbb{P}}1,\quad \text{as } n\to\infty.$$

By the consistency of the empirical distribution function of X and the decomposition given by (3.11), we obtain

$$\frac{\widehat{\mathbb{P}}_{X_i}\left(S_{h_n,x}\right)}{h_n^d f(x)}\xrightarrow{\mathbb{P}}1,\quad \text{as } n\to\infty.$$

Since $\widehat{\theta}_{\psi,n}(x)$ is a consistent estimator of $\theta_\psi(x)$ (see Theorem (3.2.1)), then we have only to show that $\widehat{\mathbb{M}}_n(x,\theta)\xrightarrow{\mathbb{P}}\mathbb{M}(x,\theta)$ and $\widehat{\Gamma}_{1,n}(x,\theta)\xrightarrow{\mathbb{P}}\Gamma_1(x,\theta)$ as $n\to\infty$ which are consequence of the previous results. □

Chapter 4

Simulation

The main objective of this chapter is to show the superiority of our prediction method by varying a number of parameters and conditions. We vary the sample size and the Censoring Rate. We carry out a simulation to compare the finite-sample performance of the robust regression $\widehat{\theta}_{\psi,n}(x)$ and the classical regression, say $\widehat{\theta}_n(x)$, when both the response and the covariate are one-dimensional scalar random variables.

4.1 Data

First, Consider a stationary ergodic process generated in the following way: for $i = 1, \ldots, n$, $X_i = 0.4X_{i-1} + \eta_i$, where $\eta_i \sim$ Bernoulli(0.5). Observe that, since the η_i's are Bernoulli-distributed then the processes , described above, do not satisfy the α-mixing condition whereas they are ergodic.

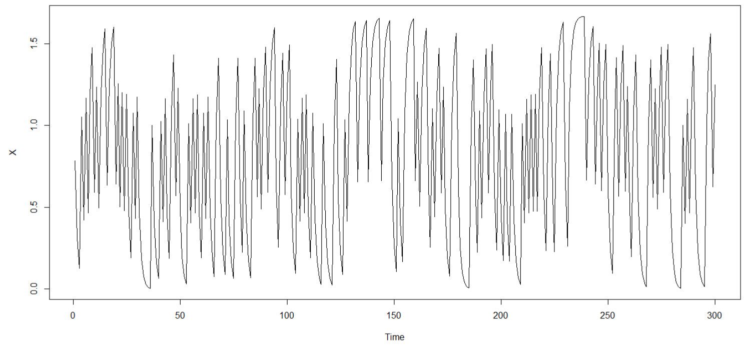

Figure 4.1: Plot of AR(1) process.

Concerning the response variables T_i, we collect them according to the following regression model: for $i = 1, \ldots, n, T_i = m(X_i) + \sigma\epsilon_i$, where $m(x) = x + 2\exp\{-16x^2\}$, $\epsilon_i \sim \mathcal{N}(0, 1)$ and $\sigma = 0.01$

We also, simulate n i.i.d. censoring variables $(C_i)_{i=1,\ldots,n}$ is such that $C_i = X_{i+1} + \lambda$ with λ is a varying constant which allows to adapt the censoring rate. In this simulation study, two sample sizes are considered $n = 300$ and 800 and for each sample size different Censoring Rates (CR) are taken $CR = 32\%, 63\%$ and 79%.

We propose to compare two methods namely, for the robust case we used the score function defined by $\psi_1(y, t) = \dfrac{y - t}{\sqrt{1 + (y - t)^2}}$ to calculate $\widehat{\theta}_{\psi,n}(x)$, while for the classical case $\widehat{\theta}_n(x)$ we have $\psi_2(y,t) = y - t$. It should be noted that for simulations we choose Gaussian kernel.

4.2 Algorithm

The robust estimator computation algorithm is as follows:
Step 1
Generate an n-sample $(X_i, Y_i, \delta_i)_{i=1,\cdots,n}$ from the regression model.
Step 2:
Calculate the Kaplan-Meier estimator based on the data generated in Step 1.
Step 3:
Calculate the estimator of $\theta_{\psi,n}(x)$.
Step 4:
We compute the MSE of $\widehat{\theta}_{\psi,n}$ as follows: $MSE\left(\widehat{\theta}_{\psi,n}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_{\psi,n}(X_i) - m(X_i)\right)^2$. Same thing for MSE of $\widehat{\theta}_n(x)$.

4.3 Result

The MSE values found for the Robuste estimator, $\widehat{\theta}_{\psi,n}$, as well as the classical estimator, $\widehat{\theta}_n$ are shown in the table below, when different sample sizes and censoring rates are taken into account. One can observe, that the two estimators perform better when the sample size n increases. It is also clear that the estimators accuracy is affected by the censoring rate, the quality of the estimators fell sharply when the censoring rate law.

CR	n	$\widehat{\theta}_{\psi,n}$	$\widehat{\theta}_n$
$\lambda = 2$, $CR \simeq 32\%$	300	0.4529446	1.068348
	800	0.4356878	1.091404
$\lambda = 5$, $CR \simeq 63\%$	300	0.5147869	1.103946
	800	0.4563536	1.108231
$\lambda = 8$, $CR \simeq 79\%$	300	0.4425118	1.100906
	800	0.4441645	1.117007

4.4 R Code

****

```
> n = 300
> X = arima.sim(list(ar = 0.4), innov = rbinom(n, 1, 0.5), n)
> E = rnorm(n)
> T = X + 2 * exp(-16 * X^2) + 0.01 * E
> plot.ts(X)
> W = window(X, start = 5, end = 10)
> plot.ts(W)
> K = function(t)(1/sqrt(2 * pi)) * exp(-0.5 * t^2)
> psi = function(y, t)(y)/(sqrt(1 + (y)^2))
> h = n^-.3      (the bandwidth)
```

$> s = 300$ (Initiation)
$> a = min(X)$
$> b = max(X)$
$> x = seq(a, b, length = s)$
$> V = numeric(n)$
$> fn = numeric(s)$
$> W = numeric(n)$
$> Hn = numeric(s)$ (Fonction Hn(.))
$> c = X + 8$
$> Y = pmin(T, c)$
$> d = as.numeric(T = c)$
$> summary(Y)$
$> df = data.frame(x = X, y = Y, Rn)$
$> z = survfit(Surv(Y, d)\ 1, data = df)$
$> summary(z)$
$> plot(z, mark.time = TRUE, conf.int = F)$
$> z1 = function(t) z\$surv$
$> for(jin1 : s) > for(iin1 : n)V[i] = K((x[j] - X[i])/h) > fn[j] = sum(V * z1(Y[i]))$
$> for(jin1 : s) > for(iin1 : n)W[i] = d[i] * K((x[j] - X[i])/h) * (Y[i]) > Hn[j] = sum(W)$
$> Rn = Hn/fn$ (Régression Rn(.))
$> Se.reg = abs(Rn - T)$
$> mes1 = mean(Se.reg)$
$> op = par(mfrow = c(1, 3))$
$> plot.ts(Rn)$
$> W = window(Rn, start = 1, end = 3)$
$> W1 = window(Y, start = 1, end = 3)$
$> lines(W1, col = "red")$
****

Conclusion

In this work, we treated a robust nonparametric estimation of the regression function. The model considered here is the right censored model which is the most used in different practical fields. The principal aim is to establish the asymptotic properties of this estimator for a stationary ergodic process with any use of traditional mixing conditions. Notice that the ergodic setting covers and completes various situations as compared to the mixing case and stands as more convenient to use in practice.

From a theoretical standpoint, this model is a generalization to the right censored case of the robust estimator of the regression function proposed by Gueriballah et al.(2013). As asymptotic results we have established the strong consistency (with rate) and the asymptotic distribution of the estimator this asymptotic property will make it possible to determine the confidence intervals. However, from a practical standpoint, our model provides an alternative estimate of the classical regression which has more advantage in sense of robustness and incomplete data. The superiority of our model compared to the classical and robust regression in the presence of values atypical censored is confirmed by our simulation study.

furthermore, the results that were obtained on the robust censored regression estimation in finite dimension can be generalized to the case of functional data.

Appendix

4.5 Definitions

Definition 4.5.1. *(Stationary process) The process $\{x_t; t \in \mathbb{Z}\}$ is* ***strongly stationary*** *if*

$$F_{t_1+k,t_2+k,\cdots,t_s+k}(b_1, b_2, \cdots, b_s) = F_{t_1,t_2,\cdots,t_s}(b_1, b_2, \cdots, b_s)$$

for any finite set of indices $\{t_1, t_2, \cdots, t_s\} \subset \mathbb{Z}$ with $s \in \mathbb{Z}^+$, and any $K \in \mathbb{Z}$.
Thus the process is strongly stationary if the joint distribution function of the vector $(x_{t_1+k}, x_{t_2+k}, \ldots, x_{t_s+k})$ is equal with the one of $(x_{t_1}, x_{t_2}, \ldots, x_{t_s})$ for any finite set of indices $\{t_1, t_2, \cdots, t_s\} \subset \mathbb{Z}$ with $s \in \mathbb{Z}^+$, and any $K \in \mathbb{Z}$.

Definition 4.5.2. *(Filtration) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family $(\mathcal{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathcal{F}$. In other words, for each t, $\mathcal{F}_t$ is a σ-algebra included in $\mathcal{F}$ and if $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_t)_{t\geq 0}$ is called a filtered probability space.*

Definition 4.5.3. *(Martingale) A process $M = (M_n)_{n=0}^{\infty}$ is martingale if*

1. *if (M_n) is adapted,*
2. *$M_n \in L^1$ for all n,*
3. *$\mathbb{E}[M_{n+1} \mid \mathcal{F}_{n-1}] = M_n$ almost surely, for all n.*

Definition 4.5.4. *(Martingale differences). A sequence of random variables $(Z_n)_{n\geq 1}$ is said to be a sequence of martingale differences with respect to the sequence of σ-fields $(\mathcal{F}_n)_{n\geq 1}$ whenever Z_n is $\mathcal{F}_n$ measurable and*

$$\mathbb{E}(Z_n \mid \mathcal{F}_{n-1}) = 0 \qquad \textit{almost surely.}$$

4.6 Some Inequalities

4.6.1 Exponential inequality.

Lemma 4.6.1. *[44] Let $(Z_n)_{n\geq 1}$ be a sequence of real martingale differences with respect to the sequence of σ-fields $(\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n))_{n\geq 1}$, where $\sigma(Z_1, ..., Z_n)$ is the σ-field generated by the random variables $Z_1, ..., Z_n$. Set $S_n = \sum_{i=1}^{n} Z_i$. For any $p \geq 2$ and any $n \geq 1$, assume that there exist some nonnegative constants C and d_n such that*

$$\mathbb{E}(Z_n^p \mid \mathcal{F}_{n-1}) \leq C^{p-2} p! d_n^2 \qquad \textit{almost surely.} \tag{4.1}$$

Then, for any $\varepsilon > 0$, we have

$$\mathbb{P}(|S_n| > \varepsilon) \leq 2 \exp\left\{-\frac{\varepsilon^2}{2(D_n + C\varepsilon)}\right\},$$

where $D_n = \sum_{i=1}^{n} d_i^2$.

4.6.2 Jensen's inequality.

Proposition 4.6.1. *[49] Let X be a real valued random variable such that $\mathbb{E}|X| < \infty$, and let $g : \mathbb{R} \to \mathbb{R}$ be convex function such that $\mathbb{E}|g(X)| < \infty$. Then*

$$g(\mathbb{E}X) \leq \mathbb{E}(g(X)).$$

4.6.3 Hölder's inequality.

Proposition 4.6.2. *[49]*

$$\mathbb{E}[|XY|] \leq \mathbb{E}^{1/p}\left[|X|^p\right]\mathbb{E}^{1/q}\left[|Y|^q\right] \equiv \|X\|_p\|Y\|_q, \quad \frac{1}{p}+\frac{1}{q}=1.$$

4.6.4 Markov's inequality.

Proposition 4.6.3. *[49] Let $X : S \to \mathbb{R}$ be a non-negative random variable. Then, for any $a > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

4.6.5 C_r-inequality.

Proposition 4.6.4. *[49]*

$$\mathbb{E}|X+Y|^r \leq c_r\mathbb{E}|X|^r + c_r\mathbb{E}|Y|^r \text{ where } c_r = 1 \text{ for } 0 < r \leq 1 \text{ and } c_r = 2^{r-1} \text{ for } r \geq 1.$$

4.6.6 Burkholder's inequality

Theorem 4.6.1. *[49] If $\{S_i, \mathcal{F}_i, 1 \leqslant i \leqslant n\}$ is a martingale and $1 < p < \infty$, then there exist constants C_1 and C_2 depending only on p such that*

$$C_1\mathbb{E}\left|\sum_{i=1}^n X_i^2\right|^{p/2} \leqslant \mathbb{E}\left|S_n\right|^p \leqslant C_2\mathbb{E}\left|\sum_{i=1}^n X_i^2\right|^{p/2}.$$

4.7 Convergences

Definition 4.7.1. *(Convergence in Probability). A sequence of random variables $X_1, X_2, X_3, \cdots$ converges* ***in probability*** *to a random variable X, shown by $X_n \overset{p}{\to} X$, if*

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \epsilon\right) = 0, \qquad \text{for all } \epsilon > 0$$

Definition 4.7.2. *Let X_n, Y_n be two sequences of random variables. The following $\left(\frac{X_n}{Y_n}\right)_{n\in\mathbb{N}} \longrightarrow 0$ a.co. if*

$$X_n \longrightarrow 0 \qquad a.co.$$

And

$$\exists\delta > 0, \quad \sum_{n=0}^{\infty}\mathbb{P}\left(|Y_n| < \delta\right) < \infty.$$

Definition 4.7.3. *(Almost complete convergence). Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of real random variables; we say that Z_n* ***converges almost completely (a.co.)*** *to zero if, and only if, $\forall\epsilon > 0$, $\sum_{n=1}^{\infty}\mathbb{P}(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of almost complete convergence of Z_n to zero is of order u_n (with $u_n \to 0$) and we write $Z_n = O_{a.co.}(u_n)$ if, and only if, $\exists\epsilon > 0$, $\sum_{n=1}^{\infty}\mathbb{P}(|z_n| > \epsilon(u_n)) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.*

Definition 4.7.4. *(Almost Sure Convergence). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X_1, X_2, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$). We say that X_n* ***converges almost surely*** *(or,* ***with probability 1****) to X if*

$$\lim_{n\to\infty}\mathbb{P}\left(\{\omega : X_n(\omega) = X(\omega)\}\right) = 1.$$

In this case, we write

$$X_n \stackrel{a.s.}{\to} X.$$

This is a really strong type of convergence for random variables in the sense that

$$X_n \stackrel{a.s.}{\to} X \Rightarrow X_n \stackrel{\mathbb{P}}{\to} X \Rightarrow X_n \stackrel{d}{\to} X.$$

Remark 4.7.1. *Almost complete convergence is stronger than almost sure convergence and convergence in probability.*

Definition 4.7.5. *(Convergence in Distribution). A sequence of random variables $X_1, X_2, X_3, \cdots$ converges* ***in distribution*** *to a random variable X, shown by $X_n \stackrel{d}{\to} X$, if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for all x at which $F_X(x)$ is continuous.

Lemma 4.7.1. *[49](Borel-Cantelli Lemma.) Let $\{E_n\}$ be a sequence of events in sample space Ω. Then*
(a) If

$$\sum_{n=1}^{\infty}\mathbb{P}(E_n) < \infty, \quad \Longrightarrow \quad \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}, E_m\right) = 0$$

that is,

$$\mathbb{P}\left[E_n \text{ occurs infinitely often}\right] = 0.$$

(b) If the events $\{E_n\}$ ***are independent***

$$\sum_{n=1}^{\infty}\mathbb{P}(E_n) = \infty \quad \Longrightarrow \quad \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} E_m\right) = 1,$$

that is,

$$\mathbb{P}\left[E_n \text{ occurs infinitely often}\right] = 1.$$

This result is useful for assessing almost sure convergence. For a sequence of random variables $\{X_n\}$ and limit random variable X, suppose, for $\epsilon > 0$, that $A_n(\epsilon)$ is the event

$$A_n(\epsilon) \equiv \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$$

The BC Lemma says

(a) *if* $\sum_{n=1}^{\infty} \mathbb{P}\left(A_n(\epsilon)\right) = \sum_{n=1}^{\infty} \mathbb{P}\left[|X_n - X| > \epsilon\right] < \infty$ ***then*** $X_n \xrightarrow{a.s} X$

(b) *if* $\sum_{n=1}^{\infty} \mathbb{P}\left(A_n(\epsilon)\right) = \sum_{n=1}^{\infty} \mathbb{P}\left[|X_n - X| > \epsilon\right] = \infty$ *with the* X_n ***independent then*** $X_n \overset{a.s}{\nrightarrow} X$.

Bibliography

[1] Aalen OO. (1975). Statistical inference for a family of counting processes. PhD thesis, Univ. of California, Berkeley.

[2] Abdelkader Gheriballah, Ali Laksaci, Soumeya Sekkal, (2013). Nonparametric M-regression for functional ergodic data. Statistics and Probability Letters 83, 902-908

[3] Andersen PK, Borgan O, Gill RD, Keiding N. (1993). Statistical Models Based on Counting Processes. Springer, New York.

[4] Andrews, D. F. (1974). A robust method for multiple linear regression. Thechnometrics, 16, 523-531.

[5] Azzedine, N., Laksaci, A., Ould Saïd, E., (2008). On the robust nonparametric regression estimation for functional regressor. Statist. Probab. Lett. 78, 3216-3221.

[6] Beran, R. (1981). Nonparametric Regression with Randomly Censored Survival Data. Technical Report. University of California, Berkeley.

[7] Benziadi Fatima, Abdelkader Gheriballah and Ali Laksaci, 2016. Asymptotic normality of kernel estimator of ψ-regression function for functional ergodic data. New Trends in Mathematical Sciences 4, No. 1, 268-282

[8] Boente, G. and Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations. Ann. Statist., 17, 1242-1256.

[9] Boente, G. and Fraiman, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. Ann. Statist., 18, 891-906.

[10] Box. G. E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. The Annals of Mathematical Statistics, 25, No.2, 290-302.

[11] Cai, Z.W., and Roussas, G.G. (1992). Uniform strong estimation under α-mixing, with rates. Statist. Probab. Lett., 15, 47-55.

[12] Cai, Z., Ould Saïd, E., (2003). Local M-estimator for nonparametric time series. Statist. Probab. Lett. 65, 433-449.

[13] Carbonez, A., Gyorfi, L., and van der Meulen, E.C. (1995). Partitioning-estimates of a regression function under random censoring. Statist. Decisions, 13, 21-37.

[14] Chaouch, M., Laib, N. and Louani, D. (2016). Nonparametric M-estimation for right censored regression model with stationary ergodic data, Statistical Methodology, DOI:http://dx.doi.org/10.1016/j.stamet.2016.10.002

[15] Chaouch, M., Laib, N. and Louani, D. (2016). Rate of uniform consistency for a class of mode regression on functional stationary ergodic data. To appear Statistical Methods & Applications, DOI: 10.1007/s10260-016-0356-9.

[16] Chaouch, M., and Khardani, S. (2015). Randomly censored regression quantile estimation using functional stationary ergodic data. J. Nonparametr. Stat., 27, 1, 65-87.

[17] Collomb.G, W. Härdle. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction : Kernel regression estimation from dependent observations. Stoch. Proc. Appl., 23, 77-89.

[18] Cox, D.R. (1972). Regression models and life-tables (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 34, 187-220.

[19] Crambes, C., Delsol, L., Laksaci, A., (2008). Robust nonparametric estimation for functional data. J. Nonparametr. Stat. 20, 573-598.

[20] Dabrowska, D.M. (1987). Nonparametric regression with censored survival data. Scand. J. Statist. 14, 181-197.

[21] Dabrowska, D.M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. Ann. Statist. 17, 1157-1167.

[22] Fan.J, T.C. Hu, Y.K. Truong. (1994). Robust nonparametric function estimation. Scand. J. Statist., 21 ,433-446.

[23] Ferraty, F., Vieu, P., (2006). Nonparametric Functional Data Analysis Theory and Practice. Springer, New York.

[24] Ferraty, F., Vieu, P., (2009). Additive prediction and boosting for functional data. Comput. Statist. Data Anal. 53, 1400-1413.

[25] Fleming TR and Harrington DP. (1991). Counting Processes and Survival Analysis. Wiley, New York.

[26] Ghement.I, M. Ruiz, R. Zamar (2008). Robust Estimation of Error Scale in Nonparametric Regression Models. J. of Statist. Plan. and Inf., in press.

[27] Györfi, L., Härdle, W., Sarda, P., Vieu, P., (1989). Nonparametric Curve Estimation for Time Series. In: Lecture Notes in Statistics, vol. 60. Springer-Verlag.

[28] Hampel, F. R. (1971). A general qualitative definition of robustness. Ann. Math. Stat., 42, 1887-1896.

[29] Hall, P., and Heyde, C.C., Martingale limit theory and its application, Probability and Mathematical Statistics, New York: Academic Press (1980).

[30] Hardle.w (1984). Robust and nonlinear time series analysis. Lecture Notes in Statistics, SpringerVerlag, New York, 26.

[31] Hardle.W , A. B. Tsybakov.(1988). Robust nonparametric regression with simultaneous scale curve estimation. Ann. Statist., 16, No.1, 120-135.

[32] Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. Proc. Natl. Acad. Sci. USA 33 25-31.

[33] Huber, P.J., (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35, 73-101.

[34] Huber, P.J. (1973). Robust regression. Asymptotics, conjonctures and monte Carlo. Ann. Stat., 43, 799-821.

[35] Huber, P.J. (1981). Robust statistics. John Wiley.

[36] Jin, Z. (2007). M-estimation in regression modes for censored data. J. Statist. Plann. Inference, 12, 3894-3903.

[37] Kaplan, E.M. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc., 53, 457-481.

[38] Kalbfleisch JD and Prentice RL. (2002). The Statistical Analysis of Failure Time Data. Wiley, New York.

[39] Kohler, M., Mathe, K., and Pinter, M. (2002). Prediction from randomly right censored data. J. Multivariate Anal., 80, 73-100.

[40] Koul, H., Susarla, V. and Van Ryzin, J. (1981) Regression analysis with randomly right-censored data. Ann. Statist. 9, 1276-1288.

[41] Koul, H. and Stute, W. (1998). Regression model fitting with long memory errors. J. Statist. Plann. Inference, 71, 35-56.

[42] Krasker, W. S. Welsch, R. E. (1982). Efficient bounded influence regression estimation. J. A. S. A., 77, 595-604.

[43] Krengel, U. (1985). Ergodic Theorems, Walter de Gruyter et Co. Berlin.

[44] Laïb, N., Louani, D., (2011). Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. J. Statist. Plann. Inference 141,359-372.

[45] Laïb, N. and Ould Saïd, E. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. Canad. J. Statist. 28, 817-828.

[46] Laïb, N. and Louani, D. (2010). Nonparametric kernel regression estimate for functional stationary ergodic data: Asymptotic properties. J. Multivariate Anal., 101, 2266-2281

[47] Laksaci, A., Lemdani, M., Ould Saïd, E., 2009. L^1-norm kernel estimator of conditional quantile for functional regressors: consistency and asymptotic normality. Statist. Probab. Lett. 79, 1065-1073.

[48] Lawless JF. (2003). Statistical Models and Methods for Lifetime Data. New York: John Wiley and Sons.

[49] Loève.M (1963). Probability Theory. Third Edition. Van Nostrand Princeton.

[50] Ren, J.J. and Gu, M. (1997). Regression M-estimators with doubly censored data. Ann. Statist., 25, 2638-2664.

[51] Robinson, R (1984). Robust nonparametric autoregression. Lecture Notes in Statistics, 26, 247-255. Springer-Verlag, New York.

[52] Stone, C.J., 2005. Nonparametric M-regression with free knot splines. J. Statist. Plann. Inference 130, 183-206.