# Contribution to the non parametric recursive estimation in relation with the extreme value index and the mixing hypothesis

Fatma Ben Khadher

University of Monastir

Faculty of Sciences of Monastir


# PhD Thesis

Presented by


**Fatma BEN KHADHER**


Speciality

**MATHEMATICS AND APPLICATIONS**

# *Contribution to the non parametric estimation in relation with the extreme value index and the mixing hypothesis*

**Defended on November 29, 2021 in front of the committee:**

| | | |
|---|---|---|
| Prof. **Mounir ZILI** | University of Monastir, Tunisia | President |
| Prof. **Jean François DUPUY** | University of Rennes, France | Referee |
| Prof. **Afif MASMOUDI** | University of Sfax, Tunisia | Referee |
| Prof. **Sophie DABO NIANG** | University of Lille, France | Examiner |
| Prof. **Haikel SKHIRI** | University of Monastir, Tunisia | Supervisor |
| Prof. **Yousri SLAOUI** | University of Poitiers, France | Supervisor |

Research laboratory: Analyse, Géométrie et Applications (LR/18/ES/16)

# Acknowledgements

Countless people supported my effort on this thesis. At first, I pray god to give me strength to forge ahead and fulfill my dream.

I express my heartfelt thanks to my supervisors Prof. Yousri SLAOUI and Prof. Haikel SKHIRI for everything they taught me and having faith on me. Being a part of their team, was a great privilege and a big opportunity which boost me to improve myself.

I acknowledge my jury members. Special thanks go to Prof. Jean François DUPUY and Prof. Afif MASMOUDI for the time they spent on reading my manuscript and their valuable feedback. I also thank Prof. Sophie DABO NIANG and Prof. Mounir ZILI. It is a pleasure to accept joining my thesis defense.

I could never thank Prof. Leila BEN ABDELGHANI enough for her guidance. This opportunity would not have been possible without her support. I thank Prof. Salah KHARDANI for helping me improve on the beginning of my thesis research.

I am indebted to Research laboratory "Analyse, Géométrie et Applications (LR/18/ES/16)" for their support with the financial means to complete my research. My sincere thanks also go to LMA Poitiers "Laboratoire de Mathématiques et Applications: UMR7348" for providing me all kind of services I have ever needed. I am fortunate to have been a part of their team.

I will forever beholden to my friends Raja and Salima. Their advice lays between the lines of this manuscript. They were always present to help me improve when I feel hopeless. Many thanks go to all my friends who are considered as my second family at Poitiers: Hadhami, Sana, Fatima zahra, Asma, Chaima, Sami, Kouceila, Omar, Mayssa, Chayma and Ahmed. Special thanks go also to Abir, Carlos, Pietro, Angélique, Cyrine and Sahar for all their support and friendship. Immense gratitude as always to Jihen, Habiba, Rym and Mohamed, then all my professors and classmates at ESSTHS for their encouragement.

I gratefully recognize the unconditional and endless love and support of my family. Mom, thank you for your uncountable prays for me. Dad, your sacrifice is the most powerful source for me to accomplish my research and fulfill your dream. I wish that you will be always proud of me. My brothers: Mustapha, Mokhles and Aymen, thank you for all the kind words and guidance you provided. Similarly other family members are also subjects to special thanks for their encouragement and faith on me during my studies. Last but not least, I am eternally grateful to my cousin Marwa for being patient and supportive through the difficult moments. She has always stood behind me, and been my source of inspiration.

# Abstract

The main objective of this thesis resides in applying the stochastic approximation method to build up a large class of recursive non parametric kernel estimators for dependent and independent variables. First, we define a recursive kernel estimator of the conditional extreme value index. We investigate the properties of the proposed recursive estimator and compare it to Hill's non recursive kernel estimator. We show that using some optimal parameters, the proposed recursive estimator defined by the stochastic approximation algorithm proves to be very competitive to Hill's estimator. Efficiency and feasibility were confirmed by theoretical results and then by applications on simulated real data about Malaria in Senegalese children. Second, we extend the work of Slaoui (2014b) to the case of $\alpha$-mixing data. We study the properties of these estimators and compare them with Nadaraya's non recursive distribution estimator. Using an optimal choice of the bandwidth and an appropriate choice of the stepsize parameter, the recursive estimators allowed us to obtain quite better results compared to the non recursive distribution estimator under $\alpha$-mixing condition in terms of estimation error. We elaborate the central limit theorem and the uniform convergence for the proposed estimators under some mild conditions. The obtained theoretical results are corroborated through simulation study. Finally, we adopt the stochastic approximation algorithms to define a kernel estimator of the mode based on the recursive kernel density estimator developed by Mokkadem et al. (2009a). Additionally, we establish its almost sure convergence under strong mixing hypothesis and we confirm these theoretical results through numerical simulations.

**Keywords:** Asymptotic normality, Bandwidth selection, Extreme value, Non parametric estimation, Mixing Data, Pareto distribution, Recursive estimator, Stochastic approximation algorithm, Strong consistency, Tail index.

# Résumé

L'objectif de cette thèse réside dans l'application de la méthode d'approximation stochastique pour construire une classe d'estimateurs à noyau récursifs non paramétriques pour les variables dépendantes et indépendantes. Dans un premier temps, nous définissons un estimateur récursif à noyau de l'indice conditionnel des valeurs extrêmes. Nous étudions les propriétés de l'estimateur récursif proposé et le comparons à l'estimateur à noyau non récursif de Hill. Nous montrons qu'en utilisant certains paramètres optimaux, l'estimateur récursif proposé défini par l'algorithme d'approximation stochastique s'avère très compétitif par rapport à l'estimateur de Hill. L'efficacité est confirmée par des résultats théoriques puis par des applications sur des données réelles simulées concernant le paludisme chez les enfants sénégalais. Deuxièmement, nous étendons le travail de Slaoui (2014b) au cas des données $\alpha$-mélangeantes. Nous étudions les propriétés de ces estimateurs et les comparons avec l'estimateur de distribution non récursif de Nadaraya. En utilisant un choix optimal de la fenêtre et un choix approprié de pas, les estimateurs récursifs nous permettent d'obtenir de meilleurs résultats que l'estimateur de distribution non récursif dans le cas $\alpha$-mélangeant en termes d'erreur d'estimation. Nous établissons le théorème central limite et la convergence uniforme pour les estimateurs proposés sous certaines conditions. Nous prouvons ces résultats théoriques par une étude de simulation. Enfin, nous adoptons les algorithmes d'approximation stochastique pour définir un estimateur à noyau du mode basé sur l'estimateur récursif de densité à noyau développé par Mokkadem et al. (2009a). En outre, nous établissons sa convergence presque sûre sous l'hypothèse de mélange fort et nous corroborons ces résultats théoriques par des simulations numériques.

**Mot-clés:** Algorithme d'approximation stochastique, Choix de fenêtre, Consistence forte, Distribution de Pareto, Données mélangeantes, Estimation non paramétrique, Estimateur récursif, Indice de queue, Normalité asymptotique, Valeur extrême.

# Contents

# List of Figures

# List of Tables

# General introduction

Estimation theory has been an attractive area of research that has generated significant scientific concern and interest among statisticians. It has led to the development of a wide variety of applied fields such as medicine, biology, public health, epidemiology, astronomy, economics and demography. There are three estimation approaches in literature. The first one stands for the parametric estimation that is the estimation of a finite number of parameter. In this case, the estimators are constructed using either the method of moments, least squared method or maximum likelihood method (See, for instance, Dempster et al (1977) and McLachlan and Peel (2004)). The second one corresponds to the non parametric estimation that is the estimation of an unknown function from observations. An introduction to non parametric methods in Tsybakov (1990). The third estimating model refers to a semi-parametric method which combines both parametric and non parametric aspects. The branch of non parametric estimation has become very attractive in current research. The oldest and most widely used method for non parametric density estimation is the histogram. This method remains insufficient to estimate a smooth density. Hence, the introduction of a kernel technique produces a smooth estimation of the probability density function and rectifies the previous problem. The kernel method of smoothing was introduced by Rosenblatt (1956) and extended by Parzen (1962). It was investigated in several directions. For example, the estimation of the density of probability as well as the distribution function, the regression function and the extreme value index function. The kernel estimators have been improved using stochastic approximation methods. This method has the capacity to facilitate updating estimators when we have new observations. That introduce the notion of the recursivity. As an excellent reference for the stochastic approximation algorithm, we refer the reader to Révész (1973, 1977) and Mokkadem et al. (2009a).

The main objective in this thesis resides in applying the stochastic approximation method to build up a large class of recursive non parametric kernel estimators for dependent and independent variables.

This manuscript is structured in terms of four major chapters. In the first chapter, We provided useful definitions and some asymptotic properties of continuous kernel estimators. Next, we considered various mixing conditions. Subsequently, we introduced the Extreme Value Theory (EVT), we displayed the fundamental theorem in EVT and we recalled certain basic definitions. Additionally, we exhibited the different extreme value distributions and we presented different estimators of extreme value index as well as recalling their asymptotic properties. Finally, we have described the stochastic Robbins-Monro algorithm which allows us to introduce recursive estimators.

In the second chapter, we applied the stochastic approximation method to define a class of recursive kernel estimator of the conditional extreme value index. We investigated the properties of the proposed recursive estimator and compared them to those concerning Hill's non recursive kernel estimator. We demonstrated that using some optimal parameters, the proposed recursive estimator defined by the stochastic approximation algorithm proves to be very competitive compared to Hill's non recursive kernel estimator. Finally, the theoretical results are tackled through simulation experiments and illustrated using real dataset about Malaria in Senegalese children. This research work is actually under review after minor revision Ben Khadher and Slaoui (2021c).

Chapter three is an extension of the work of Slaoui (2014b) to the case of $\alpha$-mixing data. We first examined the properties of these estimators and compared them to Nadaraya's non recursive distribution estimator. We showed that, using some optimal parameters, the recursive estimators allowed us to obtain quite better results compared to the non recursive distribution estimator under $\alpha$-mixing condition in terms of estimation error. Then, we elaborated the central limit theorem and the uniform convergence for the proposed estimators under some mild conditions. Finally, we corroborated these theoretical results through a few simulations. This research work was the subject of the following publication Ben Khadher and Slaoui (2021a).

In chapter four, we identified a kernel estimator of the mode based on the recursive kernel density estimator developed by Mokkadem et al. (2009a). In addition, we established its almost sure convergence under strong mixing hypotheses. This research work was the subject of the following accepted paper Ben Khadher and Slaoui (2021b).

The last part incorporates the closing section which rests upon pertinent concluding remarks as well as promising future perspectives.

# Chapter 1

# Basic concepts

In this chapter, the readers are provided with a brief review of the scientific background of non parametric estimation. First, generalities on non parametric kernel estimation were displayed. Second, methods allowing to obtain the optimal choice of smoothing parameter estimation were proved. Subsequently, a set of definitions related to types of mixing conditions were integrated. Furthermore, the concept of EVT was introduced. Eventually, the stochastic approximation algorithm which create the recursive estimators was presented.

## 1.1 Non parametric kernel estimation

We are interested in this section in classical non parametric estimation. Notably, the properties of kernel approach of the density, mode and distribution functions are reported. In this section, let $X_1, \cdots, X_n$ be independent and identically distributed (iid) $\mathbb{R}$-valued random variables and let $f$ and $F$ denote respectively the probability density and the distribution function of $X_1$.

### 1.1.1 Kernel density estimator

In this subection, we provide some asymptotic properties of continuous kernel density estimators. The first step is to define the notion of the kernel.

**Definition 1.1.1.** *A kernel is a function $K : \mathbb{R} \longrightarrow \mathbb{R}$, which is positive, integrable and satisfies $\int_{\mathbb{R}} K(x)dx = 1$.*

Some classical examples of kernels function are indicated as follows.

**Example 1.1.1.**

$K(x) = \frac{1}{2}\mathbb{1}_{\{|x|\leqslant 1\}}$ *(the rectangular kernel),*

$K(x) = (1-|x|)\mathbb{1}_{\{|x|\leqslant 1\}}$ *(the triangular kernel),*

$K(x) = \frac{3}{4}(1-|x|^2)\mathbb{1}_{\{|x|\leqslant 1\}}$ *(the Epanechnikov kernel or the parabolic kernel),*

$K(x) = \frac{15}{16}(1-|x|^2)^2\mathbb{1}_{\{|x|\leqslant 1\}}$ *(the biweight kernel),*

$K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{x^2}{2}\right)$ *(the Gaussian kernel).*

For further details, we can refer to Tsybakov (1990). The well-known kernel density estimator of $f$ was introduced by Rosenblatt (1956)(see also Parzen (1962)) and defined as

$$\forall x \in \mathbb{R},\ \widehat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right), \tag{1.1.1}$$

where $K : \mathbb{R} \longrightarrow \mathbb{R}$ is a kernel function and $(h_n)$ is a sequence of positive real numbers that goes to zero as $n$ tends to infinity called bandwidth. To investigate the asymptotic behaviors of the estimator (1.1.1), we make the following assumptions:

**Assumption 1.1.1.1.**

***(A1).*** $K(x) = K(-x),\ \forall\, x \in \mathbb{R}$,

***(A2).*** $\int_{\mathbb{R}} xK(x)dx = 0$,

***(A3).*** $\int_{\mathbb{R}} x^2K(x)dx < +\infty$,

***(A4).*** $\int_{\mathbb{R}} K^2(x)dx < +\infty$.

Multiple are the criteria to assess the efficiency of the estimator $\widehat{f}_n$. For instance, we established an asymptotic expression for the mean squared error (MSE) of the kernel estimator for any fixed value of $x$. This is defined as

$$MSE\left[\widehat{f}_n(x)\right] = \mathbb{E}\left[\left(\widehat{f}_n(x) - f(x)\right)^2\right].$$

Developing this expression, we obtain

$$MSE\left[\widehat{f}_n(x)\right] = \mathbb{B}ias^2\left[\widehat{f}_n(x)\right] + \mathbb{V}ar\left[\widehat{f}_n(x)\right],$$

where

$$\mathbb{B}ias\left[\widehat{f}_n(x)\right] = \mathbb{E}\left[\widehat{f}_n(x)\right] - f(x),$$

and

$$\mathbb{V}ar\left[\widehat{f}_n(x)\right] = \mathbb{E}\left[\left(\widehat{f}_n(x) - \mathbb{E}\left[\widehat{f}_n(x)\right]\right)^2\right] = \mathbb{E}\left[\widehat{f}^2{}_n(x)\right] - \mathbb{E}^2\left[\widehat{f}_n(x)\right].$$

To evaluate the MSE of $\widehat{f}_n$, we will need to calculate its bias and variance. Assuming that $f$ is bounded, twice differentiable and $f^{(2)}$ is bounded and substituting that $u = \frac{x-t}{h_n}$, we obtain

$$\mathbb{B}ias\left[\widehat{f}_n(x)\right] = \frac{1}{h_n}\int_{\mathbb{R}} K(u)\left[f(x - uh_n) - f(x)\right]du.$$

and

$$\mathbb{V}ar\left[\widehat{f}_n(x)\right] = \frac{1}{nh_n^2}\left[\int_{\mathbb{R}} K^2\left(\frac{x-t}{h_n}\right)f(t)dt - \left(\int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right)f(t)dt\right)^2\right].$$

By Taylor series expansion, $f(x - uh_n) = f(x) - uh_n f'(x) + \frac{1}{2}u^2h_n^2 f^{(2)}(x) + o(h_n^2)$, and by applying the properties of the kernel $K$, we get

$$\mathbb{B}ias\left[\widehat{f}_n(x)\right] = \frac{1}{2}h_n^2 f^{(2)}(x)\int_{\mathbb{R}} z^2K(z)dz + o(h_n^2)$$

and

$$\mathbb{V}ar\left[\widehat{f}_n(x)\right] = \frac{1}{nh_n}f(x)\int_{\mathbb{R}} K^2(z)dz + o\left(\frac{1}{nh_n}\right).$$

The choice of the bandwidth has an important influence over the quality of kernel estimation. It needs to be carefully determined. The optimal value of $h_n$ is obtained by minimizing the asymptotic Mean Integrated Squared Error ($MISE$).

$$MISE\left[\widehat{f}_n\right] = \int_{\mathbb{R}} MSE\left[\widehat{f}_n(x)\right]dx = \int_{\mathbb{R}} \mathbb{B}ias^2\left[\widehat{f}_n(x)\right]dx + \int_{\mathbb{R}} \mathbb{V}ar\left[\widehat{f}_n(x)\right]dx.$$

Therefore, this optimal value of $h_n$ is expressed by

$$h_{opt,n} = \left(\frac{\|K\|_2^2}{\left(\int_{\mathbb{R}} t^2K(t)dt\right)^2\|f''\|_2^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{1.1.2}$$

where $\|.\|_2$ is the Euclidien norm.

Unfortunately, the optimal smoothing parameter ($h_{opt,n}$) depends on the unknown quantity $\|f''\|_2^2 = \int_{\mathbb{R}}\left(f''(t)\right)^2 dt$. Hence, it cannot be readily applied in practice. There are several methods to estimate the smoothing parameter. Among the most famous and useful ones are the Plug-in approach and the cross-validation criterion (see Tsybakov (1990)).

**Plug-in:**

After defining the theoretical optimal bandwidth as the minimizer of the mean integrated squared error, we estimate the unknown quantities in expression $h_n$. A natural way lies in using non parametric estimator for $\int_{\mathbb{R}} \left(f''(t)\right)^2 dt$. Let us define

$$\widehat{f_n''}(x) = \frac{1}{ng_n^3} \sum_{i=1}^{n} K'' \left( \frac{x - X_i}{g_n} \right),$$

where $g_n$ is a prior bandwidth. Several estimators for $\|f''\|_2^2$ were developed by Hall and Marron (1987). Thus, they determined the bias corrected estimator in terms of

$$\widehat{\|f''\|_2^2} = \widehat{f_n''}(x) - \frac{1}{ng_n^5} \|K''\|_2^2.$$

To obtain an adequate prior bandwidth, Park and Marron (1990) set $g_n$ as the minimizer for the asymptotic mean squared error of $\widehat{\|f''\|_2^2}$. Using (1.1.2), the prior bandwidth is expressed in function of $(h_n)$ as:

$$g_n = I_1(K) I_2(f) h_n^{10/13}$$

where $I_1(K)$ contains the fourth derivative as well as convolutions of $K$, and $I_2(f)$ contains the second and third derivatives of $f$. It follows that, the expression of the optimal bandwith $(h_n)$ is expressed as

$$\widehat{h_{opt,n}} = \left\{ \frac{\|K\|_2^2}{\left(\int_{\mathbb{R}} t^2 K(t) dt\right)^2 \widehat{\|f''\|_2^2}} \right\}^{1/5} n^{-1/5},$$

**Cross-validation:**

The usual method for estimating risk is leave-one-out cross-validation. Recall that the risk of $\widehat{f}_n$ is indicated by $MISE\left[\widehat{f}_n\right] = \mathbb{E}\left(R\right)$ where

$$R\left(\widehat{f}_n\right) = \int_{\mathbb{R}} \left(\widehat{f}_n(x) - f(x)\right)^2 dx$$

is the integrated squared error loss function. The loss function, which we now write as a function of smoothing parameter $(h_n)$, (since $\widehat{f}_n$ depend on $(h_n)$) is

$$L(h_n) \quad = \quad \int_{\mathbb{R}} (\widehat{f}_n(x) - f(x))^2 dx$$

$$= \int_{\mathbb{R}} \widehat{f}_n^2(x)dx - 2\int_{\mathbb{R}} \widehat{f}_n(x)f(x)dx + \int_{\mathbb{R}} f^2(x)dx.$$

The last term does not depend on $(h_n)$. As a matter of fact, minimizing the loss is equivalent to minimizing the expected value of

$$J(h_n) = \int_{\mathbb{R}} \widehat{f}_n^2(x)dx - 2\int_{\mathbb{R}} \widehat{f}_n(x)f(x)dx.$$

We shall refer to $\mathbb{E}(J(h_n))$ as the risk, although it differs from the true risk by the constant term $\int_{\mathbb{R}} f^2(x)dx$. The cross-validation estimator of risk is represented by

$$\widehat{J}(h_n) = \int_{\mathbb{R}} \widehat{f}_n^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i)$$

where$\widehat{f}_{(-i)}(t) = \frac{1}{(n-1)h_n}\sum_{j\neq i}^{n} K\left(\frac{t-X_j}{h_n}\right)$ is the density estimator obtained after removing the $i^{th}$ observation. Next, the optimization is restricted to a range of values of $h_n$ and the one that minimizes $\widehat{J}$ shall be selected.

### 1.1.2 Kernel mode estimator

The mode is often based on a sequence of the density function $f$, defined as the value $\theta$ which maximizes it, as expressed as follows

$$f(\theta) = \sup_{t\in\mathbb{R}} f(t).$$

The kernel estimator of the mode $\theta$ is defined as the random variable $\theta_n$ maximizing the estimator $\widehat{f}_n$ (defined in (1.1.1)), which is expressed as

$$\theta_n := \arg\max_{x\in\mathbb{R}} \widehat{f}_n(x).$$

The majority of properties of mode estimators are related to those of density estimators and have been explored by several authors.

The weak consistency of the kernel sample mode was investigated by Parzen (1962). More precisely, it is assumed that the true probability density function $f(x)$ is uniformly continuous in $x$ and that the mode $\theta$ is unique. Then he reported the following theorem

**Theorem 1.1.1.** [*Parzen* (1962)] *If* $(h_n)$ *is a function of* $n$ *satisfying*

$$\lim_{n\to\infty} nh_n^2 = \infty,$$

*and if the probability density $f(x)$ is uniformly continuous, then for every $\epsilon > 0$*

$$\mathbb{P}\left[\sup_{-\infty<x<\infty} |f_n(x) - f(x)| < \epsilon\right] \longrightarrow 1, \qquad \text{as } n \longrightarrow +\infty.$$

*If $\{\theta_n\}$ are the sample modes, and if the population mode is unique, then for every $\epsilon > 0$*

$$\mathbb{P}\left[|\theta_n - \theta| < \epsilon\right] \longrightarrow 1, \qquad \text{as } n \longrightarrow +\infty.$$

This result was extended in several directions. We can mention for example Chernoff (1964), Eddy (1980, 1982) and Vieu (1996).

The strong consistency was explored by Nadaraya (1965) and Van Ryzin (1969). We recall the following theorem:

**Theorem 1.1.2.** [*Nadaraya* (1965)] *We assume that*

1. *$K$ is a continuous and bounded function such that $K(x) \longrightarrow 0$ as $|x| \longrightarrow +\infty$,*
2. *The series $\sum_{n=1}^{\infty} \exp^{-\gamma n h_n^{-2}}$ converges for any $\gamma > 0$ where $nh_n^{-2} \longrightarrow 0$ as $n \longrightarrow +\infty$,*
3. *$f(x)$ is uniformly continuous,*

*then if the mode is unique, the sample mode $\theta_n$ converges to $\theta$ almost surely (a.s.).*

The asymptotic normality of kernel estimate of the mode was elaborated by Romano (1988). The multidimensional study of the mode was carried out by Samanta (1973) and Konakov (1974).

### 1.1.3 Kernel distribution estimator

There has been a considerable development of methods for smooth estimation of distribution functions. The most popular one is the kernel approach which is identified by Nadaraya (1964) as follows

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h_n}\right) \tag{1.1.3}$$

where $\mathcal{K}(z) = \int_{\infty}^{z} K(x)dx$. In the following, we assume that the kernel function $K$ satisfies **(A1)-(A3)** in (1.1.1.1). We are now ready to state the basic properties of the kernel distribution estimator (3.1.3). In order to measure the quality of our proposed estimator $\widehat{F}_n(x)$, we use the following quantity:

$$\begin{aligned} MWISE\left[\widehat{F}_n\right] &= \mathbb{E}\left[\int_{\mathbb{R}}\left[\widehat{F}_n(x)-F(x)\right]^2 f(x)dx\right] \\ &= \int_{\mathbb{R}} \mathbb{B}ias\left[\widehat{F}_n(x)\right]^2 f(x)dx + \int_{\mathbb{R}} \mathbb{V}ar\left[\widehat{F}_n(x)\right] f(x)dx. \end{aligned}$$

Assuming that $f$ is bounded, differentiable and $f'$ is bounded, the bias and the variance of Nadaraya's estimator $\widehat{F}_n(x)$ are given by:

$$\mathbb{B}ias\left[\widehat{F}_n(x)\right] = \frac{1}{2}h_n^2 f'(x)\mu_2(K) + o(h_n^2),$$

$$\mathbb{V}ar\left[\widehat{F}_n(x)\right] = \frac{1}{n}F(x)(1-F(x)) - \frac{h_n}{n}f(x)\phi(K) + o\left(\frac{h_n}{n}\right),$$

where

$$\mu_2(K) = \int_{\mathbb{R}} t^2 K(t)dt \ \ and \ \ \phi(K) = 2\int_{\mathbb{R}} tK(t)\mathcal{K}(t)dt.$$

It follows that

$$MWISE\left[\widehat{F}_n\right] = \frac{1}{n}V_F - \frac{h_n}{n}I_1\phi(K) + \frac{1}{4}h_n^4 I_2\mu_2^2(K) + o(h_n^4),$$

where

$$I_1 = \int_{\mathbb{R}} f^2(x)dx, \ I_2 = \int_{\mathbb{R}} \left(f'(x)\right)^2 f(x)dx \ \ and \ \ V_F = \int_{\mathbb{R}} F(x)(1-F(x))f(x)dx.$$

To minimize the $MWISE\left[\widehat{F}_n\right]$, the bandwidth $(h_n)$ must be equal to

$$h_n = \left(\left[\frac{I_1\phi(K)}{I_2\mu_2^2(K)}\right]^{1/3} n^{-1/3}\right),$$

and then we get

$$MWISE\left[\widehat{F}_n\right] = n^{-1}V_F\left[1 - \frac{3}{4}\frac{I_1^{4/3}\Theta(K)}{I_2^{1/3}V_F}n^{-1/3} + o(n^{-1/3})\right],$$

where

$$\Theta(K) = \left[\frac{\phi(K)^4}{\mu_2^2(K)}\right]^{1/3}.$$

The properties of $\tilde{F}_n$ have been investigated by several authors. The uniform convergence was elaborated by Nadaraya (1964), Winter (1973), Yamato (1973) and Singh et al (1983). The asymptotic normality was addressed by Watson and Leadletter (1964).

## 1.2 Mixing conditions

Numerous probabilistic tools have been developed for measuring the dependence between variables. For a finite-variance process, elementary measures of dependence are the autocovariances and autocorrelations. Mixing assumptions, introduced by Rosenblatt (1956), are used to convey different ideas of asymptotic independence between the past and future processes. We define now the popular $\alpha$-mixing coefficient.

**Definition 1.2.1.** *Let $X = (X_i)_{i \geqslant 1}$ be a sequence of random variables. Given a positive integer $n$, set*

$$\alpha(n) = \sup_k \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B) \right|, \ A \in \mathcal{F}_1^k(X) \text{ and } B \in \mathcal{F}_{k+n}^{\infty}(X) \right\}, \quad (1.2.1)$$

*where $\mathcal{F}_i^k(X)$ is the $\sigma$-field of events generated by $X_j$, $i \leqslant j \leqslant k$. The sequence is $\alpha$-mixing if the mixing coefficient $\alpha(n) \longrightarrow 0$ as $n \longrightarrow \infty$.*

The $\alpha$-mixing, called also the strong mixing, condition was introduced by Rosenblatt (1956). It is the weakest among the known mixing conditions in the literature. There are numerous examples of stochastic processes satisfying the $\alpha$-mixing condition, such as ARMA processes, the threshold extension, the EXPAR model, the simple ARCH models, their GARCH extension and the bilinear Markovian models. If the mixing condition $\alpha(n) = O\left(\exp^{-an}\right)$ for some $a > 0$, the process is exponentially strongly mixing, where $a$ is the mixing rate and $1/a$ is the mixing time. The process is geometrically strongly mixing when there exists $\rho \in (0,1)$ such that, $\alpha(n) = O(\rho^n)$. Then, if $\alpha(n) = O\left(n^{-k}\right)$ for some $k > 0$, the process is polynomially strongly mixing. The $\alpha$-mixing has many practical applications (see Doukhan (1994), Bosq (1999), Bradley (2007) and Dedecker et al (2007) for more details).

There exist various other mixing conditions used in the literature. We mention, for instance, the $\beta$-mixing condition (see Kolmogorov (1931)), $\phi$-mixing condition (see Ibragimov (1962)), $\psi$-mixing condition (see Blum et al. (1963)) and $\rho$-mixing condition (see Hirschfeld (1935)).

## 1.3 Extreme Value Theory

The asymptotic theory of sample extremes has been developed in parallel with the central limit theory, and in fact both theories bear to a certain extend some resemblance. Let $X_1$,

$X_2, \ldots, X_n$ be independent and identically distributed random variables having a common distribution function $F$. The central limit theory concerns the limit behavior of the sum $X_1 + X_2 + \ldots + X_n$ as $n \longrightarrow \infty$, whereas the theory of sample extremes concerns the asymptotic behavior of the sample extremes $\max(X_1, X_2, \ldots, X_n)$ or $\min(X_1, X_2, \ldots, X_n)$ as $n \longrightarrow \infty$.

In this thesis, we shall consider the maxima of the sample, Knowing that all results obtained can be easy reformulated for sample minima according to the following formula:

$$\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n).$$

### 1.3.1 Extreme value distributions

The main result in EVT was introduced by Fisher and Tippet (1928) and Gnedenko (1943). They proved that the distribution of the extreme values of an iid sample from a cumulative distribution function $F$ can converge only to one distribution from the three possible ones.

**Theorem 1.3.1.** [*FisherandTippet* (1928); *Gnedenko* (1943)] *Under certain regularity conditions on the distribution function $F$, there exist a real parameter $\gamma$ and two normalizing series $(a_n)_{n\geq 1} \subset \mathbb{R}_+^*$ and $(b_n)_{n\geq 1} \subset \mathbb{R}$ such that for all $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\max(X_1, \cdots, X_n) - b_n}{a_n} \leq x\right] \underset{n\to+\infty}{\longrightarrow} \mathcal{H}_\gamma(x),$$

*with,*

$$\text{if } \gamma > 0, \quad \mathcal{H}_\gamma(x) = \begin{cases} 0 & \text{if } x < 0 \\ \exp\left[-x^{-\frac{1}{\gamma}}\right] & \text{if } x \geqslant 0 \end{cases}$$

$$\text{if } \gamma < 0, \quad \mathcal{H}_\gamma(x) = \begin{cases} \exp\left[-(-x)^{-\frac{1}{\gamma}}\right] & \text{if } x < 0 \\ 1 & \text{if } x \geqslant 0 \end{cases}$$

$$\text{if } \gamma = 0, \quad \mathcal{H}_0(x) = \exp\left[\exp(-x)\right] \text{ for all } x \in \mathbb{R}.$$

*The distribution function $\mathcal{H}_\gamma$ is called extreme value distribution. It is indexed by a shape parameter $\gamma$ called the extreme value index. This parameter accounts for the behaviour of the tail of the distribution.*

Three domains of attraction, depending on the sign of $\gamma$ , should be distinguished:

- If $\gamma > 0$, $F$ is said to belong to the **Fréchet** domain of attraction. This domain includes distribution with heavy tails, i.e. their survival distribution function decreases as a power function.

- If $\gamma = 0$, $F$ is said to belong to the **Gumbel** domain of attraction. This domain includes distributions with light tails, i.e. their survival distribution function decreases as an exponential rate.

- If $\gamma < 0$, $F$ is said to belong to the **Weibull** domain attraction. This domain includes distributions with short tails, i.e. they have a finite endpoint $x_F = inf\{x, F(x) \geqslant 1\}$.

**Remark 1.3.1.** *We present, in the table below, some domains of attraction associated with usual distribution:*

| *Fréchet* ($\gamma > 0$) | *Gumbel* ($\gamma = 0$) | *Weibull* ($\gamma < 0$) |
|---|---|---|
| *Pareto* | *Normal* | *Uniform* |
| *Cauchy* | *Exponential* | *Beta* |
| *Student* | *Gamma* | |
| *Fréchet* | *Weibull* | |
| *Burr* | | |

**Remark 1.3.2.** *The normalization sequences $(a_n)_n$ and $(b_n)_n$, in Theorem 1.3.1, are interpreted as a scale and a location parameters. Their choices is not unique.*

**Example 1.3.1.** *If $X_1, \cdots, X_n$ is a sequence of independent standard Fréchet variables, $F(x) = 1 - (x/a)^{-\alpha}$, for $x > a > 0$ and $\alpha > 0$. Let $a_n = a\alpha^{-1}n^{1/\alpha}$, $b_n = an^{1/\alpha}$ and extreme value index $\gamma = 1/\alpha > 0$,*

$$\begin{aligned} \mathbb{P}\left[\frac{\max(X_1, \cdots, X_n) - b_n}{a_n} \leq x\right] &= F^n(a_n x + b_n) \\ &= \left(1 - \frac{(1 + x/\alpha)^{-\alpha}}{n}\right)^n, \end{aligned}$$

*So* $\mathbb{P}\left[\frac{\max(X_1,\cdots,X_n)-b_n}{a_n} \leq x\right] \longrightarrow \exp\left(-(1 + x/\alpha)^{-\alpha}\right) = \mathcal{H}_{1/\alpha}(x)$ *as $n \longrightarrow \infty$ for $x > -\alpha$.*

### 1.3.2 Characterization of the domains of attraction

The characterization of domains of attraction relies on the theory of regularly-varying functions.

**Definition 1.3.1.** *A positive function $L$ is regularly-Vaying with index $\delta \in \mathbb{R}$ at infinity if*

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = t^{\delta}, \forall\, t > 0.$$

*This property is denoted by $L \in \mathcal{RV}_{\delta}$. If $\delta = 0$, the function $L$ is said to be slowly-varying.*

A well known example of a slowly-varying function is $L(x) = \ln x$.
Let us now display the expressions of the distribution function in each domain.

- **Fréchet Domain of attraction**

  **Theorem 1.3.1.** *$F$ is in the domain of attraction of a Fréchet distribution with shape parameter $\xi$ if and only if $\overline{F}$ is regularly varying with index $-1/\xi$*

  $$ie: \;\; \overline{F}(x) = 1 - F(x) = x^{-1/\xi} L(x), \;\; x > 0,$$

  *where $L$ is a slowly-varying function.*

- **Weibull Domain of attraction**
  For all distribution function $F$ with finite endpoint $x_F$, we denote by $F_*$ the distribution function defined by $F_*(x) = F(x_F - 1/x)$ if $x > 0$ and $F_*(x) = 0$ otherwise.

  **Theorem 1.3.2.** *$F$ is in the domain of attraction of a Weibull distribution if and only if $x_F$ is finite and $F_*$ belongs to domain of attraction of Fréchet. Let $\gamma > 0$ be the extreme value index associated with $F_*$, the extreme value index associated with $F$ is then $-\gamma$.*

- **Gumbel Domain of attraction**

  **Theorem 1.3.3.** *$F$ is in the domain of attraction of a Gumbel distribution if and only if there exists $x_0 < x_F \leq \infty$ such that*

  $$\overline{F}(x) = c(x) \exp\left( -\int_{x_0}^{x} \frac{g(t)}{a(t)} dt \right),$$

  *where $a$, $c$ and $g$ three functions verifying $a'(x) \longrightarrow 0$, $c(x) \longrightarrow c > 0$ and $g(x) \longrightarrow 1$ as $x \longrightarrow x_F$.*

### 1.3.3 Estimation of the extreme value index

The extreme value index $\gamma$ plays a central role in terms of searching for the shape of the distribution tail. We need to estimate it in order to better understand the nature of the studied extreme distribution. Several methods for estimating this parameter are proposed in the literature of EVT. The Hill estimator (see Hill (1975)), the Pickands estimator (see Pickands (1975)) and the Dekkers et al. moment estimator (see Dekkers al. (1989)) present the most widely used ones in practice. An extensive discussion of estimation methods for EVT models can be found in Embrechts et al. (1997). We recall below the three most frequently used estimators of the extreme value index and their asymptotic properties.

**The Hill estimator:**

The Hill estimator is defined by:

$$\begin{aligned} \widehat{\gamma}_{t_n}^H(x) &= \frac{1}{t_n}\sum_{i=1}^{t_n} i\left(\log X_{n-i+1,n} - \log X_{n-i,n}\right) \\ &= \frac{1}{t_n}\sum_{i=1}^{t_n} \log X_{n-i+1,n} - \log X_{n-t_n,n}, \end{aligned}$$

where $X_{1,n} \leq \cdots \leq X_{n,n}$ are the associated order statistics to the sample $X_1, \cdots, X_n$ and $t_n$ is the number of the top order statistics (number of extremes) used for the estimation of $\gamma$. The construction of this estimator is based on the maximum likelihood method. It is well known that the Hill estimator displays a very good performance, that is competitive with respect to other EVT methods of estimation. Theoretically, the Hill estimator is favorably considered in view of its asymptotic properties, which are summarized in Embrechts et al. (1997)(Theorem 6.4.6):

- Weak consistency: if $t_n \longrightarrow \infty$ and $t_n/n \longrightarrow 0$ for $n \longrightarrow \infty$, then $\widehat{\gamma}_{t_n}^H \xrightarrow{\mathbb{P}} \gamma$.
- Strong consistency: if $t_n/n \longrightarrow \infty$ and $t_n/\log\log n \longrightarrow \infty$ for $n \longrightarrow \infty$, then $\widehat{\gamma}_{t_n}^H \xrightarrow{a.s.} \gamma$.
- Asymptotic normality: under additional hypotheses, $\sqrt{t_n}\left(\widehat{\gamma}_{t_n}^H - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \gamma^2\right)$,

where $\xrightarrow{\mathbb{P}}$ denotes the convergence in probability, $\xrightarrow{\mathcal{D}}$ the convergence in distribution and $\mathcal{N}$ the gaussian-distribution.

**Pickands Estimator:**

This estimator is proposed by Pickands (1975) to estimate the shape parameter of any of the three extreme value distributions. It is expressed as

$$\widehat{\gamma}_{t_n}^{P} = \frac{1}{\log 2} \log\left(\frac{\log X_{n-t_n+1,n} - \log X_{n-2t_n+1,n}}{X_{n-2t_n+1,n} - X_{n-4t_n+1,n}}\right).$$

Their asymptotic properties are well studied in Dekkers al. (1989):

- Weak consistency: if $t_n \longrightarrow \infty$ and $t_n/n \longrightarrow 0$ for $n \longrightarrow \infty$, then $\widehat{\gamma}_{t_n}^{P} \xrightarrow{\mathbb{P}} \gamma$.
- Strong consistency: if $t_n/n \longrightarrow \infty$ and $t_n/\log\log n \longrightarrow \infty$ for $n \longrightarrow \infty$, then $\widehat{\gamma}_{t_n}^{P} \xrightarrow{a.s.} \gamma$.
- Asymptotic normality: under additional hypotheses which can be consulted in Dekkers al. (1989),

$$\sqrt{t_n}\left(\widehat{\gamma}_{t_n}^{P} - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\gamma^2\left(2^{2\gamma+1}+1\right)}{4\left(\log 2\right)^2\left(2^{\gamma}-1\right)^2}\right).$$

**The moment Estimator:**

The moment estimator is identified by Dekkers al. (1989) and determined as:

$$\gamma_{t_n}^{M} = M_{t_n}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_{t_n}^{(1)}\right)^2}{M_{t_n}^{(2)}}\right)^{-1}, \quad 1 < t_n < n,$$

where $M_{t_n}^{(r)} = \frac{1}{t_n}\sum_{j=1}^{t_n}\left(\log X_{n_j+1,n} - \log X_{n-t_n,n}\right)^r$, $r = 1, 2$. Note that $M_{t_n}^{(1)}$ corresponds to the Hill estimator. It is called moment estimator since $M_{t_n}^{(1)}$ can be considered as empirical moments of the order $r$. Further more, it is known as the Dekkers-Einmahl-de Haan estimator as an extension to the Hill estimator. Note that the asymptotic properties of $\gamma_{t_n}^{M}$ estimator were investigated in Dekkers al. (1989).

Suppose $F$ belongs to one of the domain of attraction with $\gamma \in \mathbb{R}$, $x_F > 0$ and let $(t_n)_{n\geq 1}$ be a sequence of integers such that $1 \leq t_n < n$, $t_n \longrightarrow \infty$ and $t_n/n \longrightarrow 0$ as $n \longrightarrow \infty$.

- Weak consistency: then, $\widehat{\gamma}_{t_n}^{M} \xrightarrow{\mathbb{P}} \gamma$.
- Strong consistency: if $t_n/(\log n)\delta \longrightarrow \infty$ as $n \longrightarrow \infty$ for $\delta > 0$, then $\widehat{\gamma}_{t_n}^{M} \xrightarrow{a.s.} \gamma$.

- Asymptotic normality: under additional hypotheses on the distribution function $F$ (see Dekkers al. (1989), Theorem 3.1 and Corollary 3.2),

$$\sqrt{t_n}\left(\widehat{\gamma}_{t_n}^M - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\gamma^2\right),$$

where

$$\sigma_\gamma^2 = \begin{cases} 1+\gamma^2 & \text{if } \gamma \geq 0 \\ \dfrac{(1-\gamma)^2(1-2\gamma)\left(1-\gamma+6\gamma^2\right)}{(1-3\gamma)(1-4\gamma)} & \text{if } \gamma < 0 \end{cases}.$$

## 1.4 Stochastic approximation method and recursive estimators

Stochastic algorithms have been widely used in numerous areas including adaptive control, system identification, sequential change detection and transmission systems, see Benveniste et al. (1990) for multiple interesting examples. The stochastic algorithm method allows us to construct a class of recursive estimators. The advantage of recursive estimators lies in the fact that their update, from a sample of size $n$ to one of size $n+1$, requires considerably less computations.

### 1.4.1 The Stochastic approximation method

The general form of stochastic algorithm is:

$$\theta_n = \theta_{n-1} + \gamma_n \phi\left(\theta_{n-1}, W_n\right) + \gamma_n^2 \mu_n\left(\theta_{n-1}, W_n\right), \tag{1.4.1}$$

where $(\theta_n)$ stands for the sequence to be recursively updated, $(\gamma_n)$ corresponds a positive sequence of real numbers decreasing towards zero, $(W_n)$ represents a sequence of random variables representing the on-line observations, $\phi(\theta, W)$ refers to the function which essentially defines how the parameter $\theta$ is updated as a function of new observation and $\mu_n\left(\theta_{n-1}, W_n\right)$ relates to a small perturbation on the algorithm.
The behavior of this algorithm was investigated by Benveniste et al. (1990), the special case when $\mu_n = 0$ was considered by Delyon (1996). Algorithm (1.4.1) coincides with the one analyzed by Kushner (1977), Ljung (1978) and Ruppert (1982):

$$\theta_n = \theta_{n-1} + \gamma_n\left[\phi(\theta_{n-1}) - W_n + \eta_n\right], \tag{1.4.2}$$

where $\eta_n$ stands for a random variables which converges to 0 almost surely and $\phi$ corresponds to a measurable unknown function.

They asserted that (1.4.2) includes the Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) stochastic approximation processes, which allow the search for zero $\theta^\star$ of the function $\phi$. The application of Robbins–Monro's procedure to construct a stochastic approximation algorithm was identified by Révész (1973, 1977) and extended by Tsybakov (1990). Most of the classical results for the Robbins–Monro and Kiefer–Wolfowitz processes require the assumption $\mathbb{E}\left[W_n|\mathcal{F}_{n-1}\right] = 0$, where $\mathcal{F}_{n-1}$ stands for the $\sigma$-algebra of the events occurring up the time $n-1$. Under standard conditions on the function $\phi$ and on the sequence $(\gamma_n)$, Kushner and Yin (2003) highlighted that

$$\theta_n \longrightarrow \theta^\star \ \text{ a.s. } \ \text{ as } \ n \longrightarrow \infty.$$

In the following subsections, two examples of recursive estimators are established using the Robbins and Monro algorithm (See Robbins and Monro (1951)).

### 1.4.2 Recursive kernel estimators

#### Recursive kernel density estimator

In order to construct a stochastic algorithm, which approximates the unknown density function $f$ at a given point $x$, Mokkadem et al. (2009a) defined an algorithm to search for the zero of the function $g : y \longmapsto f(x) - y$ as follows:

(i) $f_0(x) \in \mathbb{R}$,

(ii) $\forall n \geqslant 1$, we set $f_n(x) = f_{n-1}(x) + \gamma_n \mathcal{Z}_n(x)$, where the stepsize $(\gamma_n)$ is a sequence of positive real numbers that go to zero and $(Z_n)$ is a sequence of functions $Z_n : \mathbb{R} \to \mathbb{R}$ defined by $Z_n(x) = g(f_{n-1}(x)) - W_n + \eta_n$. Departing from the fact that $\mathbb{E}(W_n|\mathcal{F}_{n-1}) = 0$, where $\mathcal{F}_{n-1}$ stands for the $\sigma$-algebra of the events occurring at the time $n-1$, it follows that $\mathbb{E}(Z_n(x)) = f(x) - f_{n-1}(x) + \eta_n$. Adapting the approach of Révész (1973, 1977) and noting that

$$\mathbb{E}\left[h_n^{-d} K(h_n^{-1}(x - X_n))\right] = f(x) + \xi_n(x),$$

where $\xi_n(x)$ goes to zero as $n$ goes to infinity, we set $Z_n(x) = h_n^{-d} K(h_n^{-1}(x - X_n)) - f_{n-1}(x)$.

Therefore, the recursive estimator $f_n$ of the density function $f$ at the point $x$ can be stated as

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n K(h_n^{-1}(x - X_n)). \tag{1.4.3}$$

Further more, we suppose that $f_0(x) = 0$. Let $\Pi_n = \prod_{j=1}^{n}(1-\gamma_j)$. As a matter of fact, we infer from Equation (1.4.3) that $f_n$ can be rewritten as

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k K\left(\frac{x-X_k}{h_k}\right).$$

**Recursive kernel distribution estimator**

In order to construct a stochastic algorithm, which approximates the function $F$ at a given point $x$, Slaoui (2014b) defined an algorithm to search for the zero of the function $h : y \to F(x) - y$ as follows:

(i) we set $F_0(x) \in [0,1]$.

(ii) For all $n \geq 1$, we set

$$F_n(x) = F_{n-1}(x) + \gamma_n Q_n(x),$$

where the stepsize $(\gamma_n)$ represents a positive sequence of real numbers decreasing to zero and $(Q_n)$ determins a sequence of functions $Q_n : \mathbb{R} \to \mathbb{R}$ defined by $Q_n(x) = \phi(F_{n-1}(x)) - W_n + \eta_n$. Relying upon the fact that $\mathbb{E}(W_n|\mathcal{F}_{n-1}) = 0$, where $\mathcal{F}_{n-1}$ stands for the $\sigma$-algebra of the events occurring up the time $n-1$, it follows that $\mathbb{E}(Q_n(x)) = F(x) - F_{n-1}(x) + \eta_n$. Based on the approach of Révész (1973, 1977) and noting that

$$\mathbb{E}\left[\mathcal{K}(h_n^{-1}(x-X_n))\right] = F(x) + \xi_n(x),$$

where $\xi_n(x)$ goes to zero as $n$ goes to infinity and $\mathcal{K}(z) = \int_{-\infty}^{z} K(u)du$, we set $Q_n(x) = \mathcal{K}(h_n^{-1}(x-X_n)) - F_{n-1}(x)$.

Hence, the recursive estimator $F_n$ of the distribution function $F$ at the point $x$ can be expressed as

$$F_n(x) = (1-\gamma_n)F_{n-1}(x) + \gamma_n \mathcal{K}(h_n^{-1}(x-X_n)). \tag{1.4.4}$$

Further more, we suppose that $F_0(x) = 0$. Let $\Pi_n = \prod_{j=1}^{n}(1-\gamma_j)$. Thus, we infer from Equation (1.4.4) that $F_n$ can be rewritten as

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k \mathcal{K}\left(\frac{x-X_k}{h_k}\right).$$

Recall that all definitions and theoretical concepts, introduced in this chapter, allows us to better understand the following chapters.

# Chapter 2

# The stochastic approximation method for recursive kernel estimation of the conditional extreme value index

***Abstract**: In this research chapter, our central focus is upon applying the stochastic approximation method to define a class of recursive kernel estimator of the conditional extreme value index. We investigate the properties of the proposed recursive estimator and compare them to those pertaining to Hill's non recursive kernel estimator. We attempt to demonstrate that using some optimal parameters, the proposed recursive estimator defined by the stochastic approximation algorithm proves to be very competitive to Hill's non recursive kernel estimator. Finally, the theoretical results are explored through simulation experiments and illustrated using real data about Malaria in Senegalese children.*

***Keywords**:* Extreme value, Pareto distribution, Stochastic approximation algorithm, Tail index.

## 2.1 Introduction

The EVT is a branch of statistics which studies the asymptotic distributions of extreme values. It can be a maximum or a minimum of a set of random variables. This theory was developed by Emil Julius Gumbel (1958). It is widely applied in many research areas like climate changes, environmental risks, insurance and financial banking (see Beirlant et al. (2004) for a list of interesting examples). Estimation of the tail index, associated with a random variable $Y$, is one of the main problems in the area of EVT. Therefore, a lot of research, aiming to estimate this parameter, carried out during last decades (see for example Embrechts et al. (1997), Beirlant et al. (2004), De Haan and Fereira (2006), Reiss and Thomas (2007), Gardes and Girard (2008), Gardes et al. (2010) and Stupfler (2013)). We denote by $\gamma$ the tail index which characterizes the distribution tail heaviness of $Y$. For example when $\gamma$ is positive, the survival function of Y decreases polynomially to zero. It can be estimated parametrically using the Hill (1975) estimator and non parametrically using a kernel version of the Hill's estimator proposed by Goegebeur et al. (2014). Improved approaches have recently appeared in literature. Among them, we can mention Beirlant et al. (2004) to proposed the generalized Hill estimator;Brilhante et al. (2013) to defined a moment of order $p$ estimator which reduces to the Hill estimator for $p = 0$; Beran et al. (2014) who proposed a harmonic moment tail index estimator; Paulauskas and Vaičiulis (2013, 2017) who elaborated parametric families of functions of the order statistics.

In many practical applications, it is often the case that the variable of interest $Y$ can be linked to a covariate $X$. In this case, the extreme-value index of the conditional distribution of $Y$ given $X = x$ can depend on $x$; the problem is then to estimate the conditional extreme-value index $x \longrightarrow \gamma(x)$. Motivating examples in the literature include the estimation of the maximal production level as a function of the quantity of labor (see Daouia et al. (2010)), studying extreme temperatures as a function of various topological parameters (see Ferrez et al. (2011)), or analyzing extreme earthquakes as a function of the location (see Pisarenko and Sornette (2003)). Let $(X_i, Y_i)$, $i = 1, \cdots, n$, be independent realizations of the random vectors $(X, Y) \in \mathbb{R}^d \times \mathbb{R}_0^+$, where $X$ is a $d$-dimensional covariate with joint density function $g$, $d \geqslant 1$. The probability density function of $Y$ given $X = x$ is defined as $f(y|x) = \mathbb{P}(Y = y|X = x)$ and the conditional survival function of $Y$ given $X = x$ is denoted by $\overline{F}(y|x) = \mathbb{P}(Y > y|X = x)$. Now, we define the kernel version of Hill's estimator of the conditional extreme value index proposed by Goegebeur et al. (2014), which is expressed as follows

$$\tilde{\gamma}_n(x) = \frac{\frac{1}{n}\sum_{i=1}^n K_{h_n}(x - X_i)\left[\ln Y_i - \ln t_n\right] \mathbb{1}_{\{Y_i > t_n\}}}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(x - X_i)\, \mathbb{1}_{\{Y_i > t_n\}}}, \tag{2.1.1}$$

where $K_h(x) := h^{-d} K\left(h^{-1}x\right)$ with K is a kernel function and $(t_n)$ is a nonrandom threshold sequence tending to $\infty$ as $n \longrightarrow \infty$.

Recently, recursive estimation has drawn the attention and whetted the interest of multiple researchers. Recursivity means that the estimator calculated from the first $n$ observations, say $\theta_n$, is a function of $\theta_{n-1}$. More precisely, we can easily update the estimator value with each additional observation specialy in large sample sizes. The basic objective of the present work lies in applying the stochastic approximation method to construct a recursive kernel estimator of the conditional extreme value index defined in (2.1.1). To the best of our knowledge, this tail index estimator construction was not previously considered in literature and it aims to improve the estimation accuracy. It turns out that this estimator depends on two important parameters, which are the bandwidth and the stepsize of the stochastic algorithm. By making an adequate choice of the two parameters, the proposed recursive estimator can be very competitive to Hill's non recursive kernel estimator in terms of estimation error and much better in terms of computational costs.

The remainder of the chapter is organized as follows. In Section 2, we identify our estimator and we set forward its asymptotic properties. Simulation experiments and investigation of real data are presented in Section 3. Finally, the last section wraps up the conclusion and provides new perspectives for future works.

## 2.2 Construction of the estimator and asymptotic properties

### 2.2.1 The proposed estimator

In this chapter, we are basically interested in heavy tails. More precisely, we assume that the conditional survival function of $Y$ given $X = x$ satisfies

**(C1)**: $\overline{F}(y|x) = y^{-\frac{1}{\gamma(x)}} l(y|x)$,

where $\gamma(\cdot)$ is an unknown positive continuous function of the covariate $x$ called the tail function and for a fixed $x$, $l(\cdot|x)$ is a function that varies slowly at infinity, i.e for all $\lambda > 0$,

$$\lim_{y \to \infty} \frac{l(\lambda y|x)}{l(y|x)} = 1.$$

Condition $(C1)$ means that the conditional distribution of $Y$ given $X = x$ is in the Frechet maximum domain of attraction. The tail function $\gamma(x)$ is the conditional extreme value index function which needs to be adequately estimated from the available data.

**(C2)**: $l(\cdot|x)$ is normalized.

The Karamata representation (Theorem 1.3.1 given in Bingham et al. (1987)) of the slowly-varying function, $l(\cdot|x)$, can be written as

$$l(y|x) = c(x) \exp\left( \int_1^y \frac{\varepsilon(z|x)}{z} dz \right),$$

where $c(\cdot)$ is a positive function and $\varepsilon(z|x) \longrightarrow 0$ as $z \longrightarrow \infty$. Thus, $l(\cdot|x)$ is differentiable and the function $\varepsilon(\cdot|x)$ is given by $\varepsilon(z|x) = z\frac{l'(z|x)}{l(z|x)}$.

**(C3)**: There exists a strictly negative function $\rho(\cdot)$, a strictly positive function $\gamma(\cdot)$ and a rate function $b(\cdot|x)$, $b(y|x) \longrightarrow 0$ as $y \longrightarrow \infty$, of constant sign for large values of $y$ such that for all $\upsilon > 0$

$$\lim_{y\to\infty} \frac{\frac{\overline{F}(\upsilon y|x)}{\overline{F}(y|x)} - \upsilon^{-\frac{1}{\gamma(x)}}}{b(y|x)} = \upsilon^{-\frac{1}{\gamma(x)}} \frac{\upsilon^{\frac{\rho(x)}{\gamma(x)}} - 1}{\rho(x)\gamma(x)}.$$

Additional conditions are needed for ensuring the asymptotic properties of the estimators. Let $d(x, y)$ denote the Euclidean distance between $x$ and $y$, for all $x, y \in \mathbb{R}^d$.

**(C4)**: There exists $c_g > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$|g(x) - g(y)| \leq c_g d(x, y).$$

**(C5)**: There exists $c_{\overline{F}} > 0$ and $y_0 > 1$ such that for all $x, z \in \mathbb{R}^d$,

$$\sup_{y\geq y_0} \left| \frac{\ln \overline{F}(y|x)}{\ln \overline{F}(y|z)} - 1 \right| \leq c_{\overline{F}} d(x, z).$$

Moreover, we impose a condition on the kernel function $K$.

**(C6)**: $K$ is a bounded density function on $\mathbb{R}^d$, with support $\Omega$ included in the unit hypersphere of $\mathbb{R}^d$.

Our idea rests upon to construct a recursive estimator for the conditional tail index $\gamma(x)$. This recursive version is based on the estimator proposed by Goegebeur et al. (2014) which is a rational function. Therefore, it will be presented as a ratio of two estimators $a_n(x)$ and $b_n(x)$. The denominator $b_n(x)$ is an estimator of the function $b(x) = g(x)\overline{F}(t_n|x)$. The nominator $a_n(x)$ is an estimator of the function $a(x) = \gamma(x)\overline{F}(t_n|x)C_x g(x)$, where

$$C_x = 1 + \frac{b(t_n|x)}{\gamma(x)\rho(x)} \left[ \frac{1}{1-\rho(x)} - 1 + r_{n,x} \right]$$

and $(r_{n,x})$ is a non-random sequence, tending to 0 as $n \longrightarrow \infty$, defined as

$$r_{n,x} = \frac{\rho(x)}{\gamma^2(x)} \int_1^\infty z^{-\frac{1}{\gamma(x)}-1} \left( \gamma^2(x) \frac{z^{\frac{1}{\gamma(x)}} \frac{\overline{F}(t_n z|x)}{\overline{F}(t_n|x)} - 1}{b(t_n|x)} - \frac{z^{\frac{\rho(x)}{\gamma(x)}} - 1}{\rho(x)\gamma(x)} \right) dz.$$

**Remark 2.2.1.** *Since $C_x$ tends to 1, we can remove it safely from the expression of $a(x)$. Thus it can be written as $a(x) = \gamma(x)\overline{F}(t_n|x)g(x)$.*

**Remark 2.2.2.** *Based on a deterministic threshold as in the article Goegebeur et al. (2014) and Ndao et al. (2016), we use the deterministic threshold. It is possible also to consider a random threshold* $(t_n)$ *as in the article of Stupfler (2013). Additionally, we can even make the comparison between two results.*

**Construction of a recursive estimator of the function $a(x)$:**
Let us introduce the stochastic algorithm to estimate the function $a(\cdot)$ at a point $x$. It is based on searching the zero of the function $f_1 : y \longmapsto a(x) - y$. Following Robbins-Monro's procedure, this algorithm is defined as follows:

(i) $a_0(x) \in \mathbb{R}$,

(ii) $\forall n \geqslant 1$, we set $a_n(x) = a_{n-1}(x) + \gamma_n \mathcal{Z}_n(x)$, where the stepsize $(\gamma_n)$ is a sequence of positive real numbers that goes to zero and $\mathcal{Z}_n(x)$ is a sequence of function $\mathcal{Z}_n : \mathbb{R}^d \to \mathbb{R}$ defined by $\mathcal{Z}_n(x) = f_1(a_{n-1}(x)) - \mathcal{W}_n + \zeta_n$, where $\zeta_n$ is a random variables converges to 0 almost surely.

To construct $\mathcal{W}_n(x)$, we follow the approach of Révész (1973, 1977), Tsybakov (1990) and Slaoui (2013, 2014a,b, 2018) which are based on the classical property of stochastic algorithms (which is $\mathbb{E}[\mathcal{W}_n(x)|\mathbb{F}_{n-1}] = 0$, where $\mathbb{F}_{n-1}$ stands for the $\sigma$ -algebra of the events occurring at the time $n-1$). Then, it comes $\mathbb{E}(\mathcal{Z}_n(x)) = a(x) - a_{n-1}(x) + \zeta_n$. In addition, we introduce a kernel $K$ (which is a function satisfying $\int_{\mathbb{R}^d} K(z)dz = 1$), and a bandwidth $(h_n)$ (which is a sequence of positive real numbers that goes to zero when $n \longrightarrow \infty$) . Moreover, we have $\mathbb{E}\left[K_{h_n}(x - X_n)[\ln Y_n - \ln t_n] \mathbb{1}_{\{Y_n > t_n\}}\right] = a(x) + \eta_n(x)$, where $\eta_n(x)$ goes to zero as $n$ goes to $\infty$. Then, we set $\mathcal{Z}_n(x) = K_{h_n}(x - X_n)[\ln Y_n - \ln t_n] \mathbb{1}_{\{Y_n > t_n\}} - a_{n-1}(x)$. The stochastic approximation algorithm introduced in Mokkadem et al. (2009a) which estimates recursively the function $a$ at the point $x$ is defined as follows:

$$a_n(x) = (1 - \gamma_n) a_{n-1}(x) + \gamma_n K_{h_n}(x - X_n)[\ln Y_n - \ln t_n] \mathbb{1}_{\{Y_n > t_n\}}. \tag{2.2.1}$$

Considering $a_0(x) = 0$, the estimator $a_n$ defined in (2.2.1) can be rewritten as

$$a_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k K_{h_k}(x - X_k)[\ln Y_k - \ln t_n] \mathbb{1}_{\{Y_k > t_n\}}, \tag{2.2.2}$$

with

$$\Pi_n = \prod_{k=1}^{n} (1 - \gamma_k). \tag{2.2.3}$$

**Construction of a recursive estimator of the function $b(x)$:**
We apply the stochastic algorithm to estimate the function $b(\cdot)$ at a point $x$. It is based on searching the zero of the function $f_2 : y \longmapsto b(x) - y$. Following Robbins-Monro's procedure, this algorithm is defined as follows:

(i) $b_0(x) \in \mathbb{R}$,

(ii) $\forall n \geqslant 1$, we set $b_n(x) = b_{n-1}(x) + \beta_n \mathcal{T}_n(x)$, where the stepsize $(\beta_n)$ is a sequence of positive real numbers that goes to zero and $\mathcal{T}_n(x)$ is an observation of the function $f_2$ at the point $b_{n-1}(x)$.

Based on the same previously used approach, we consider $\mathcal{T}_n(x) = K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}} - b_{n-1}(x)$, with the same bandwidth $(h_n)$ and kernel function $K_h$ previously defined. Then, the stochastic approximation algorithm to estimate recursively the function $b$ at the point $x$ is defined as follows:

$$b_n(x) = (1 - \beta_n) b_{n-1}(x) + \beta_n K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}. \tag{2.2.4}$$

Considering $b_0(x) = 0$, the estimator $b_n$ defined by (2.2.4) can be rewritten as

$$b_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k K_{h_k}(x - X_k) \mathbb{1}_{\{Y_k > t_n\}}, \tag{2.2.5}$$

with

$$Q_n = \prod_{k=1}^{n} (1 - \beta_k). \tag{2.2.6}$$

Then, our proposed recursive estimator for the conditional tail index $\gamma(x)$ is defined as:

$$\widehat{\gamma}_n(x) := \frac{a_n(x)}{b_n(x)} = \frac{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k K_{h_k}(x - X_k) [\ln Y_k - \ln t_n] \mathbb{1}_{\{Y_k > t_n\}}}{Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k K_{h_k}(x - X_k) \mathbb{1}_{\{Y_k > t_n\}}}. \tag{2.2.7}$$

The second objective of our chapter is to study the properties of the recursive estimator defined by (2.2.7) and to compare them with the kernel version of Hill's estimator of the conditional extreme value index defined in (2.1.1).

The asymptotic properties of $\widehat{\gamma}_n$ are investigated in the next subsection.

### 2.2.2 Asymptotic results

In order to obtain the bias and the variance of the recursive estimator $\widehat{\gamma}_n$ defined by (2.2.7), we first calculate those of the recursive estimator $a_n$ defined by (2.2.2). Then, we calculate those of the recursive estimator $b_n$ defined by (2.2.5). Throughout this chapter, stepsizes and bandwidths are considered to belong to the following regularly varying sequences class.

**Definition 2.2.1.** *Let $u \in \mathbb{R}$ and $(u_n)_{n\geqslant 1}$ be a nonrandom positive sequence. We say that $u_n \in \mathcal{GS}(u)$ if*

$$\lim_{n\to\infty} n\left[1-\frac{u_{n-1}}{u_n}\right] = u.$$

This condition was introduced by Galambos and Seneta (1973). The acronym $\mathcal{GS}$ stands for (**G**alambos and **S**eneta). Typical sequences in $\mathcal{GS}(u)$ are, for $b \in \mathbb{R}$, $n^u(\log n)^b$, $n^u(\log\log n)^b$ and so on.

Finally, we impose the following additional conditions:

**(C7)**:

(i) $\gamma_n \in \mathcal{GS}(-\alpha)$ with $\alpha \in (1/2, 1]$.

(ii) $h_n \in \mathcal{GS}(-p)$ with $p \in (0, \alpha/d)$.

(iii) $\lim_{n\to\infty} n\gamma_n \in (\min(p, (\alpha - pd)/2), \infty]$.

(iv) $\beta_n \in \mathcal{GS}(-b)$ with $b \in (1/2, 1]$.

(v) $\lim_{n\to\infty} n\beta_n \in (\min(p, (b - pd)/2), \infty]$.

(vi) $nh_n^{d+2}\ln^2 t_n \underset{n\to\infty}{\longrightarrow} \infty$.

The following notations will be often used in this chapter:

$$\begin{aligned}
\varepsilon &= \lim_{n\to\infty}(n\gamma_n)^{-1}. && (2.2.8)\\
\varepsilon_1 &= \lim_{n\to\infty}(n\beta_n)^{-1}. && (2.2.9)\\
c'_{\overline{F}} &= c_{\overline{F}}\|z\|_2 \text{ such as } z \in \mathbf{B}^*_{\mathbb{R}^d}(0,1) = \{x \in \mathbb{R}^d\ ;\ 0 < \|x\|_2 \le 1\}. \\
C &= \left(-\frac{1}{\gamma(x)} + o(1)\right) c_{\overline{F}}\|u\|_2 \text{ for all } u \in \Omega. \\
\widetilde{m}_n(x) &= \mathbb{E}\left[K_{h_n}(x - X_n)(\ln Y_n - \ln t_n)\mathbb{1}_{\{Y_n > t_n\}}\right]. \\
m_n(x) &= \mathbb{E}\left[(\ln Y_n - \ln t_n)\mathbb{1}_{\{Y_n > t_n\}} | X_n = x\right].
\end{aligned}$$

Since we are interested in the asymptotic behavior of the estimator $\widehat{\gamma}_n$, we shall start by giving the asymptotic behavior of the estimator $a_n$.

**Theorem 2.2.1.** *(Bias and variance of the estimator $a_n$)*
*Let Assumptions **(C1)** – **(C7)** hold.*

1. *If $p \in (0, \alpha/(d+2)]$, then*

$$\mathbb{E}(a_n(x)) = a(x) + O(h_n \ln t_n). \tag{2.2.10}$$

*If $p \in (\alpha/(d+2), 1/d)$, then*

$$\mathbb{E}(a_n(x)) = a(x) + O(\sqrt{\gamma_n h_n^{-d}}). \tag{2.2.11}$$

2. *If $p \in (0, \alpha/(d+2))$, then*

$$\mathbb{V}ar(a_n(x)) = o(h_n^2 \ln^2 t_n). \tag{2.2.12}$$

*If $p \in [\alpha/(d+2), 1/d)$, then*

$$\begin{aligned}\mathbb{V}ar(a_n(x)) \quad &= \quad \frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1 g(x)\,\gamma^2(x)\overline{F}(t_n|x)\gamma_n h_n^{-d} \\ &\quad +o(\gamma_n h_n^{-d}).\end{aligned} \tag{2.2.13}$$

Departing from the above, we infer that the bias and the variance of the estimator $a_n$ heavily depend on the choice of the stepsize $(\gamma_n)$. We consider an example of choices of $(\gamma_n)$based on the minimization of the variance.

**Choices of $(\gamma_n)$ minimizing the variance of the estimator $a_n$:**
As mentioned in Mokkadem et al. (2009a), by considering the point of view of estimation by confidence intervals, it is recommended to minimize the variance of the proposed estimator for confidence interval estimation (see also Hall (1992)).

**Corollary 2.2.1.** *Let the assumptions of Theorem 2.2.1 hold. To minimize the asymptotic variance of the estimator $a_n$, $\alpha$ must be chosen equal to 1, $(\gamma_n)_n$ must satisfy $\lim_{n\to\infty} n\gamma_n = 1 - pd$, and we then have*

$$\mathbb{V}ar(a_n(x)) = 6\,(1-pd)\,\|K^2\|_1 g(x)\,\gamma^2(x)\frac{\overline{F}(t_n|x)}{nh_n^d} + o\left(\frac{1}{nh_n^d}\right).$$

The proof of Corollary 2.2.1 follows immediately from (2.2.13).

The following proposition provides the $MISE$ of the estimator $a_n$. First, we have

$$MISE(a_n) = \int_{\mathbb{R}^d} MSE(a_n(x))dx = \int_{\mathbb{R}^d} \left[(\mathbb{E}(a_n(x)) - a(x))^2 + \mathbb{V}ar(a_n(x))\right] dx.$$

**Proposition 2.2.1.** *Let Assumptions **(C1)-(C7)** hold.*

1. *If $p \in (0, \alpha/(d+2))$,*

$$MISE(a_n) = O(h_n^2 \ln^2 t_n).$$

2. *If $p = \alpha/(d+2)$,*

$$\begin{aligned} MISE(a_n) &= \frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1 \int_{\mathbb{R}^d} g(x)\gamma^2(x)\overline{F}(t_n|x)dx\gamma_n h_n^{-d} \\ &\quad + o\left(\gamma_n h_n^{-d}\right) + O\left(h_n^2 \ln^2 t_n\right). \end{aligned}$$

3. *If $p \in (\alpha/(d+2), 1/d)$,*

$$MISE(a_n) = \frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1 \int_{\mathbb{R}^d} g(x)\gamma^2(x)\overline{F}(t_n|x)dx\gamma_n h_n^{-d} + O(\gamma_n h_n^{-d}).$$

Now, we treat the asymptotic behavior of the estimator $b_n$, in order to deduce the one of the estimator $\widehat{\gamma}_n$.

**Theorem 2.2.2.** *(Bias and variance of the estimator $b_n$)*
*Let Assumptions **(C1)-(C7)** hold.*

1. *If $p \in (0, b/(d+2)]$, then*

$$\mathbb{E}(b_n(x)) = b(x) + O(h_n \ln t_n). \tag{2.2.14}$$

   *If $p \in (b/(d+2), 1/d)$, then*

$$\mathbb{E}(b_n(x)) = b(x) + O\left(\sqrt{\frac{\beta_n}{h_n^d}}\right). \tag{2.2.15}$$

2. *If $p \in (0, b/(d+2))$, then*

$$\mathbb{V}ar(b_n(x)) = O(h_n^2 \ln^2 t_n). \tag{2.2.16}$$

   *If $p \in [b/(d+2), 1/d)$, then*

$$\mathbb{V}ar(b_n(x)) = \frac{1}{2-(b-pd)\varepsilon_1}\|K\|_2^2 g(x)\overline{F}(t_n|x)\frac{\beta_n}{h_n^d} + O\left(\frac{\beta_n}{h_n^d}\right). \tag{2.2.17}$$

The bias and the variance of the estimator $b_n$ defined by the stochastic approximation algorithm (2.2.4), then heavily depend on the choice of the stepsize $(\beta_n)$. For an adequate choice, we consider an example of choices of $(\beta_n)$ based on the minimization of the variance.

**Choices of $(\beta_n)_n$ minimizing the variance of the estimator $b_n$:**
As mentioned in Mokkadem et al. (2009a), it is recommended to minimize the variance of the proposed estimator for confidence interval estimation.

**Corollary 2.2.2.** *Let the assumptions of Theorem 2.2.2 hold. To minimize the asymptotic variance of the estimator $b_n$, $b$ must be chosen equal to 1, $(\beta_n)_n$ must satisfy $\lim_{n\to\infty} n\beta_n = 1-pd$, and we then have*

$$\mathbb{V}ar(b_n(x)) = (1-pd)\|K\|_2^2 g(x)\frac{\overline{F}(t_n|x)}{nh_n^d} + O\left(\frac{1}{nh_n^d}\right).$$

The proof of Corollary 2.2.2 follows immediately from (2.2.17).

The following proposition provides the $MISE$ of the estimator $b_n$.

**Proposition 2.2.2.** *Let Assumptions **(C1)-(C7)** hold.*

1. *If $p \in (0, b/(d+2))$,*

$$MISE(b_n) = O(h_n^2 \ln^2 t_n).$$

2. *If $p = b/(d+2)$,*

$$MISE(b_n) = \frac{1}{2-(b-pd)\varepsilon_1}\|K\|_2^2\int_{\mathbb{R}^d} g(x)\overline{F}(t_n|x)dx\frac{\beta_n}{h_n^d} + O\left(h_n^2\ln^2 t_n\right) + O\left(\frac{\beta_n}{h_n^d}\right).$$

3. *If $p \in (b/(d+2), 1/d)$,*

$$MISE(b_n) = \frac{1}{2-(b-pd)\varepsilon_1}\|K\|_2^2\int_{\mathbb{R}^d} g(x)\overline{F}(t_n|x)dx\frac{\beta_n}{h_n^d} + O\left(\frac{\beta_n}{h_n^d}\right).$$

Now we present the bias and the variance of $\widehat{\gamma}_n$.

**Theorem 2.2.3.** *(Bias and variance of $\widehat{\gamma}_n$)*
*Let Assumptions **(C1)-(C7)** hold, and suppose that the stepsize $(\beta_n)_n = (n^{-1})_n$.*

1. *If $p \in (0, \alpha/(d+2)]$, then*

$$\mathbb{E}\left(\widehat{\gamma}_n(x)\right) - \gamma\left(x\right) \quad = \quad O\left(h_n \ln t_n\right). \tag{2.2.18}$$

   *If $p \in (\alpha/(d+2)), 1/d)$, then*

$$\mathbb{E}\left(\widehat{\gamma}_n(x)\right) - \gamma\left(x\right) = O\left(\sqrt{\frac{\gamma_n}{h_n^d}}\right). \tag{2.2.19}$$

2. *If $p \in (0, \alpha/(d+2))$, then*

$$\mathbb{V}ar\left(\widehat{\gamma}_n(x)\right) = o\left(h_n^2 \ln^2 t_n\right). \tag{2.2.20}$$

   *If $p \in [\alpha/(d+2), 1/d)$, then*

$$\mathbb{V}ar\left(\widehat{\gamma}_n(x)\right) = \frac{1}{b^2(x)} \frac{6}{2-(\alpha-pd)\varepsilon} \|K^2\|_1 g\left(x\right) \gamma^2(x) \overline{F}(t_n|x) \frac{\gamma_n}{h_n^d} + o\left(\frac{\gamma_n}{h_n^d}\right). \tag{2.2.21}$$

Clearly, the bias and the variance of the estimator $\widehat{\gamma}_n$ depend on the choice of the two stepsizes $(\gamma_n)_n$ and $(\beta_n)_n$.

Let us state now the following Theorem, which gives the weak convergence rate of the proposed recursive estimator $\widehat{\gamma}_n$ defined in (2.2.7) in the special case of $(\beta_n)_n = (n^{-1})_n$.

**Theorem 2.2.4.** *Let Assumptions **(C1)-(C7)** hold, and suppose that $(\beta_n)_n = (n^{-1})_n$.*

1. *If there exists $r > 0$ such that $\overline{F}(t_n|x)^{-1}\gamma_n^{-1}h_n^{d+2}\ln^2 t_n \underset{n\to\infty}{\longrightarrow} r$ then*

$$\sqrt{\overline{F}(t_n|x)^{-1}\gamma_n^{-1}h_n^d}\left(\widehat{\gamma}_n(x) - \gamma(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{r}\mathcal{B}(x), \mathcal{V}ar(x)\right),$$

   *where*

$$\begin{aligned} \mathcal{B}(x) &= -\left(\frac{C}{1-p\varepsilon} + \frac{C}{1-p}\right)\gamma\left(x\right), \\ \mathcal{V}ar(x) &= \frac{1}{b^2(x)} \frac{6}{2-(\alpha-pd)\varepsilon} \|K^2\|_1 g\left(x\right) \gamma^2(x). \end{aligned}$$

2. *If $\overline{F}(t_n|x)^{-1}\gamma_n^{-1}h_n^{d+2}\ln^2 t_n \underset{n\to\infty}{\longrightarrow} \infty$, then*

$$\frac{1}{h_n \ln t_n}\left(\widehat{\gamma}_n(x) - \gamma(x)\right) \xrightarrow{\mathbb{P}} \mathcal{B}(x),$$

**Corollary 2.2.3.** *Under the same assumptions as the previous theorem and if $r = 0$ then*

$$\sqrt{\overline{F}(t_n|x)^{-1}\gamma_n^{-1}h_n^d}\left(\widehat{\gamma}_n(x) - \gamma(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathcal{V}ar(x)\right),$$

*with*

$$\mathcal{V}ar(x) = \frac{1}{b^2(x)}\frac{6}{2-(\alpha - pd)\varepsilon}\|K^2\|_1 g\left(x\right)\gamma^2(x).$$

We can consider the case when the stepsize $(\beta_n)$ is chosen to minimise the variance of the estimator $b_n$. Similarly, we can obtain the weak convergence rate of the estimator $\widehat{\gamma}_n$.

The following corollary is a consequence of the previous theorem which gives an asymptotic confidence interval of the index function $\gamma$.

**Corollary 2.2.4.** *The asymptotic $100(1-\alpha)\%$ confidence interval for $\gamma(x)$ is given by*

$$\left(\widehat{\gamma}_n + \sqrt{\overline{F}(t_n|x)\gamma_n h_n^{-d}}\sqrt{r}\mathcal{B}(x) \pm u_{1-\frac{\alpha}{2}}\sqrt{\overline{F}(t_n|x)\gamma_n h_n^{-d}}\sqrt{\mathcal{V}ar(x)}\right),$$

*where $u_{1-\frac{\alpha}{2}}$ is the normal $\left(1-\frac{\alpha}{2}\right)$ quantile.*

## 2.3 Simulation study

The target of our applications is to compare the performance of the proposed recursive kernel estimator of the conditional extreme value index given in (2.2.7) to that of Hill's non recursive estimator defined in (2.1.1) using the "Leave One Out" cross-validation bandwidth selection.

### 2.3.1 The study design

We use the following simulation design: we consider the unidimensional case $d = 1$ and we simulate $N = 500$ samples of size $n$ $(n = 50, 250)$ of independent replicates $(X_i, Y_i)$ where $X$ is uniformly distributed on $[0, 1]$ and the conditional distribution of $Y_i$ given $X_i = x$ is Pareto with parameter $\gamma(x) = 0.5(0.1 + \sin(\pi x) \times (1.1 - 0.5\exp(-64(x - 0.5)^2)))$ (this function was proposed by Daouia et al. (2011)), it was also used in Goegebeur et al. (2014) and in Ndao et al. (2016). The pattern of $\gamma$ is given in Figure 2.1.

For each of the $N$ simulated samples, we estimate $\gamma(\cdot)$ at $x = (0.1, 0.2, 0.3, \cdots, 0.8, 0.9)$ using the estimator (2.2.7) with a biquadratic kernel $K(x) = \frac{15}{16}(1-x^2)^2\mathbb{1}_{[-1,1]}$. As mentioned in previous papers (see Slaoui (2014a,b)), there is no big influence on the choice of the kernel $K$ in our setup when the observations are not contamined.

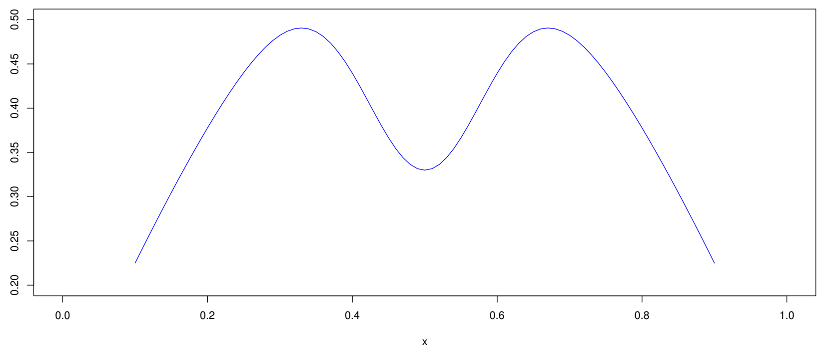


Figure 2.1: Pattern of $\gamma(\cdot)$ on $[0,1]$

In order to calculate our estimator, we need to choose the bandwidth $(h_n)_n$ and the threshold $(t_n)_n$. We take $t_n$ to be the $(n-k)$ th order statistic $Y_{(n-k)}$ as is usual in extreme value statistics.
Moreover, we propose an algorithm for choosing $(h_n, k)$. This algorithm adapted from Goegebeur et al. (2014), was considered recently by Ndao et al. (2016). The purpose is then to select the bandwidth $(h_n)$ using the following cross-validation criterion

$$h_{cv} = \arg\min_{h\in\mathcal{H}} \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathbb{1}_{\{Y_i\leqslant Y_j\}} - \widehat{F}_{n,-i}(Y_j|X_i)\right)^2,$$

where $\mathcal{H} = \{h_n = cn^{-v}; n \geqslant 1 \textit{ and } (c,v) \in \{0.1, 0.2, \cdots, 0.9\}\}$ is a grid of values for $(h_n)$ and

$$\widehat{F}_{n,-i}(y|x) := \frac{\sum_{j=1,j\neq i}^{n} K_h(x-X_j)\mathbb{1}_{\{Y_j\leqslant y\}}}{\sum_{j=1,j\neq i}^{n} K_h(x-X_j)}.$$

This criterion was introduced in Yao (1999), implemented by Gannoun et al. (2002), and established in an extreme value context by Daouia et al. (2011, 2013), Goegebeur et al. (2014) and Ndao et al. (2016). Using this bandwidth selection, we consider the following procedure to determine the number of threshold excesses $k$. This procedure rests on considering for each point $x$, the following steps:

Step 1: we compute the estimates for $\widehat{\gamma}_{n,Y_{(n-k)}}(x)$ with $k = 1, \cdots, n-1$.

Step 2: we construct several successive "blocks" of the estimates $\widehat{\gamma}_{n,Y_{(n-k)}}(x)$ (one block for $k \in \{1, \cdots, 15\}$, a second block for $k \in \{16, \cdots, 30\}$ and so on).

Step 3: we calculate the standard deviation of the estimate in each block.

Step 4: we determine the $k$-value (denoted by $k_1$) from the block with minimal standard deviation (in particular, we take the median of the $k$-values in that block).

Finally, we estimate $\gamma(x)$ by using the estimator $\widehat{\gamma}_n$ (2.2.7) by taking $(h_n, k) = (h_{cv}, k_1)$

### 2.3.2 Results

For each configuration of the simulation design parameters (sample size $n$, stepsize parameters $(\gamma_n, \beta_n)$ and covariate value $x$), we calculate the average $IAE$ (Integrated Absolute Error), the average $ISE$ (Integrated Squared Error); and $L_\infty$ of the estimators over $N =$ 500 trials; $\overline{IAE} = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}}\left|\gamma_n^{[i]}(x) - \gamma(x)\right| dx$, $\overline{ISE} = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}}\left(\gamma_n^{[i]}(x) - \gamma(x)\right)^2 dx$ and $L_\infty = \max_{i=1,\cdots,N}\int_{\mathbb{R}}\left|\gamma_n^{[i]}(x) - \gamma(x)\right| dx$, where $\gamma_n^{[i]}$ corresponds to the estimator computed from the ith sample. In order to investigate the comparison estimators, we consider the stepsizes $(\gamma_n, \beta_n)$ equal to $(n^{-1}, n^{-1})$, $((2/3)\, n^{-1}, n^{-1})$, $(n^{-1}, (2/3)\, n^{-1})$ and $((2/3)\, n^{-1}, (2/3)\, n^{-1})$ respectively. These four choices of parameters of the recursive estimator are referred to as $R1$, $R2$, $R3$ and $R4$ respectively. Results are highlighted in Table 2.1. We point out that the major merit of our proposed estimator lies in its update aspect. Indeed, when new sample points are available, it requires less computational cost than non recursive estimator. Moreover, Table 2.1 reveals that our proposed recursive estimator can provide better results in some specific situations that are very close in general to the reference values, which proves the effectiveness of our proposed recursive estimator in terms of the estimation error. Figure 2.2 discloses that all the considered estimators yield good results since the values of $\gamma$ at each point $x \in \{0.1, 0.2, \ldots 0.9\}$ are very close to the median.

### 2.3.3 Real data application

We considered a Malaria dataset of 176 families in Senegal, totaling 505 children between 2 and 19 years old, living in two villages of Niakhar (Toucar and Diohine). The number of observations was 6986. We measured Plasmodium falciparum Parasite Load (PL) from thick blood smears obtained by finger-prick during two different seasons and regularly over a three-year observation period (2001-2003). The number of measurements per child ranged from 1 to 15. We refer readers to consult Milet et al. (2010) for more details about data. These data were used also in Slaoui and Nuel (2014c) in a parametric context and more recently in Slaoui (2019a) in a non parametric context. Considering this real data, the recursive estimator (2.2.7) and the non recursive estimator (2.1.1) are compared with

| $n$ | $\gamma(0.1) = 0.225$ | | | | | $\gamma(0.2) = 0.3777$ | | | | | $\gamma(0.3) = 0.4824$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ |
| | **0.2095** | 0.1829 | 0.1589 | 0.2089 | 0.1814 | 0.4012 | **0.3901** | 0.3446 | 0.4057 | 0.3583 | 0.3566 | 0.3647 | 0.3324 | **0.3843** | 0.3503 |
| 50 | **0.0224** | 0.0238 | 0.0335 | 0.023 | 0.0227 | **0.023** | 0.0239 | 0.0326 | 0.0301 | 0.0249 | **0.0223** | 0.0241 | 0.0344 | 0.0306 | 0.0259 |
| | **0.0977** | 0.1003 | 0.1087 | 0.1012 | 0.0994 | **0.0995** | 0.1006 | 0.1083 | 0.1083 | 0.1015 | **0.0982** | 0.1015 | 0.1082 | 0.1105 | 0.1036 |
| | 0.4173 | 0.4028 | 0.6902 | **0.3668** | 0.4225 | 0.41 | **0.3877** | 0.5269 | 0.5114 | 0.4687 | **0.3381** | 0.4909 | 0.7606 | 0.6507 | 0.46 |
| | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ |
| | 0.3189 | 0.3166 | **0.2971** | 0.3312 | 0.3108 | 0.3355 | 0.3115 | **0.3391** | 0.3 | 0.3253 | **0.4204** | 0.3709 | 0.3857 | 0.3744 | 0.3893 |
| 250 | **0.0036** | 0.0036 | 0.0042 | 0.0042 | 0.0038 | **0.0037** | 0.0038 | 0.0043 | 0.0042 | 0.004 | **0.0038** | 0.0039 | 0.0044 | 0.0044 | 0.004 |
| | **0.0422** | 0.0425 | 0.0457 | 0.0453 | 0.0438 | **0.0427** | 0.0435 | 0.0465 | 0.0457 | 0.0447 | **0.0433** | 0.0436 | 0.0466 | 0.0466 | 0.045 |
| | **0.088** | 0.0883 | 0.1025 | 0.1016 | 0.0957 | 0.0995 | 0.1039 | **0.0926** | 0.1069 | 0.1237 | 0.1265 | 0.1253 | **0.1156** | 0.1352 | 0.1177 |
| $n$ | $\gamma(0.4) = 0.4395$ | | | | | $\gamma(0.5) = 0.33$ | | | | | $\gamma(0.6) = 0.4395$ | | | | |
| | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ |
| | 0.3936 | 0.3723 | 0.357 | **0.3983** | 0.3819 | 0.4181 | **0.4306** | 0.3539 | 0.5089 | 0.4183 | 0.3578 | 0.3581 | 0.3029 | **0.4345** | 0.3676 |
| 50 | **0.0248** | 0.0259 | 0.0352 | 0.0335 | 0.0282 | 0.0245 | **0.0239** | 0.0314 | 0.031 | 0.026 | **0.0243** | 0.0249 | 0.0336 | 0.031 | 0.0265 |
| | **0.1011** | 0.1029 | 0.1102 | 0.1124 | 0.1057 | 0.1021 | **0.1011** | 0.1071 | 0.1095 | 0.1029 | **0.1017** | 0.1018 | 0.1088 | 0.1091 | 0.1035 |
| | **0.4756** | 0.4909 | 0.7606 | 0.6464 | 0.5454 | **0.3842** | 0.3877 | 0.5269 | 0.5114 | 0.4114 | **0.4756** | 0.4909 | 0.7606 | 0.5158 | 0.5454 |
| | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ |
| | **0.4269** | 0.5024 | 0.4624 | 0.523 | 0.4814 | 0.382 | 0.359 | **0.346** | 0.3736 | 0.3605 | 0.3848 | 0.3862 | **0.4339** | 0.382 | 0.4292 |
| 250 | **0.0036** | 0.0037 | 0.0043 | 0.0043 | 0.004 | **0.0038** | 0.0039 | 0.0045 | 0.0043 | 0.004 | **0.0037** | 0.0038 | 0.0044 | 0.0044 | 0.004 |
| | **0.0425** | 0.0428 | 0.0462 | 0.0458 | 0.0449 | **0.0432** | 0.0435 | 0.0472 | 0.046 | 0.0448 | **0.0425** | 0.0431 | 0.0463 | 0.046 | 0.0444 |
| | 0.0952 | 0.0895 | **0.0941** | 0.1173 | 0.1039 | **0.0925** | 0.1099 | 0.1283 | 0.1006 | 0.1125 | 0.098 | 0.0894 | 0.1218 | 0.1216 | **0.0968** |
| $n$ | $\gamma(0.7) = 0.4824$ | | | | | $\gamma(0.8) = 0.3777$ | | | | | $\gamma(0.9) = 0.2250$ | | | | |
| | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ |
| | 0.4085 | 0.4032 | 0.3457 | **0.4837** | 0.4147 | 0.3496 | 0.3495 | 0.3223 | **0.3680** | 0.3394 | 0.5872 | 0.5987 | 0.4448 | **0.3635** | 0.4021 |
| 50 | **0.0239** | 0.0257 | 0.0348 | 0.0291 | 0.0255 | **0.0217** | 0.0225 | 0.0305 | 0.0285 | 0.0257 | **0.0205** | 0.0231 | 0.0318 | 0.027 | 0.0234 |
| | **0.0999** | 0.1024 | 0.1097 | 0.1081 | 0.1032 | **0.0982** | 0.0995 | 0.1062 | 0.1079 | 0.1034 | **0.0964** | 0.0987 | 0.1078 | 0.105 | 0.1005 |
| | **0.4801** | 0.4909 | 0.7606 | 0.5158 | 0.5192 | **0.3284** | 0.3396 | 0.5352 | 0.3997 | 0.39 | **0.3037** | 0.4909 | 0.7606 | 0.5114 | 0.3913 |
| | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ | $NR$ | $R1$ | $R2$ | $R3$ | $R4$ |
| | 0.4431 | **0.4751** | 0.5599 | 0.4351 | 0.5128 | **0.3708** | 0.3571 | 0.405 | 0.3071 | 0.3483 | 0.2428 | 0.2439 | **0.2223** | 0.2372 | 0.2162 |
| 250 | **0.0038** | 0.0039 | 0.0045 | 0.0044 | 0.004 | **0.0037** | 0.0037 | 0.0043 | 0.0041 | 0.0039 | **0.0034** | 0.0036 | 0.0041 | 0.0041 | 0.0037 |
| | **0.0433** | 0.0436 | 0.0469 | 0.0462 | 0.0449 | **0.0426** | 0.043 | 0.0464 | 0.0453 | 0.0441 | **0.0414** | 0.0421 | 0.0455 | 0.0447 | 0.0433 |
| | **0.092** | 0.1177 | 0.1152 | 0.12 | 0.0994 | 0.0994 | **0.0959** | 0.1482 | 0.0986 | 0.1484 | 0.0925 | **0.0825** | 0.0888 | 0.0922 | 0.0899 |

Table 2.1: Simulation results for $\gamma(x)$. $N = 500$ trials are considered. For each configuration of the simulated parameters $(n, \gamma_n, \beta_n, x)$, the first line indicates the value of each estimator in different points $x$. Lines two, three and four provide the averages $IAEs$, $ISEs$ and $L_\infty$ respectively of the five different kernel density estimators. NR corresponds to the non recursive estimator and $R1$, $R2$, $R3$ and $R4$ correspond to the proposed recursive estimators when $(\gamma_n, \beta_n)$ takes $(n^{-1}, n^{-1})$, $((2/3)\, n^{-1}, n^{-1})$, $(n^{-1}, (2/3)\, n^{-1})$ and $((2/3)\, n^{-1}, (2/3)\, n^{-1})$ respectively.

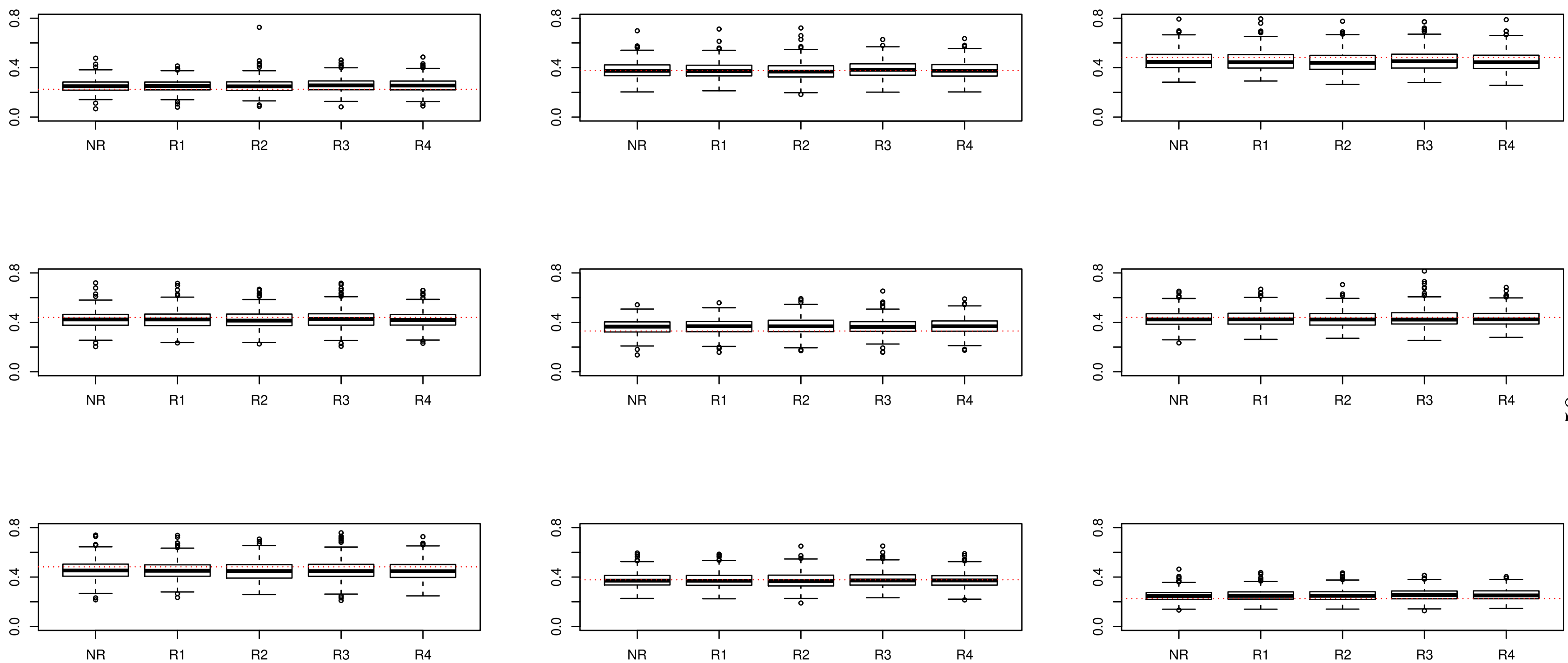


Figure 2.2: Boxplots of the $N = 500$ estimates of our five considered estimators ($NR$, $R1$, $R2$, $R3$, $R4$) in points $x = 0.1$, 0.2, 0.3 (1st line), $x = 0.4$, 0.5, 0.6 (2nd line) and $x = 0.7$, 0.8, 0.9 (3rd line), and dashed lines represent the values of reference estimator $\gamma(\cdot)$ in each point $x$ as mentioned above.

the reference index function, considered by Lekina (2010) and defined as follows:

$$\gamma(x_i) = 0.3\frac{x_i^2 - \min\limits_j x_j^2}{\max\limits_j x_j^2 - \min\limits_j x_j^2} + 0.2,$$

where $x_i$ is the ith value in the data vector. Once a variable change has been taken into consideration, the data vector should be in the same interval [0.2,0.5] proposed by Lekina (2010), so that the previous function $\gamma(x)$ can be used. Therefore, for any considered estimator $\gamma_n$ of the index function $\gamma$, we propose to compute $IAE$ and $ISE$ defined as:

$$IAE(\gamma_n) = \int_{\mathbb{R}} |\gamma_n(x) - \gamma(x)|\, dx$$

and

$$ISE(\gamma_n) = \int_{\mathbb{R}} (\gamma_n(x) - \gamma(x))^2\, dx,$$

Departing from Table 2.2 and Figure 2.3, we infer that the $IAE$ and the $ISE$ of the proposed recursive estimator are smaller than those of the non recursive estimator set forward by Goegebeur et al. (2014). Thus, demonstrating the effectiveness of our considered estimator.

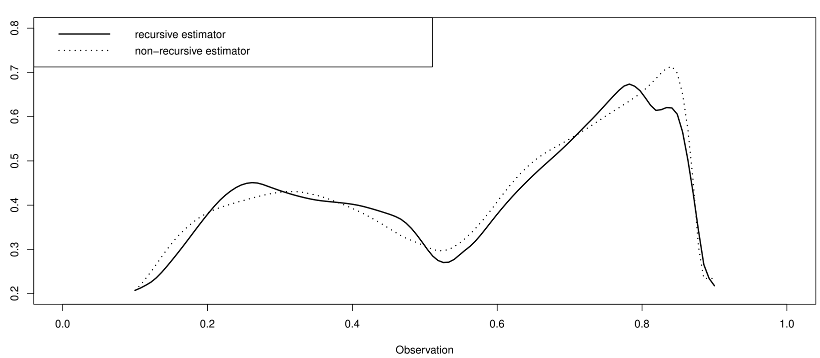


Figure 2.3: Qualitative comparaison between the non recursive estimator (2.1.1) and the proposed recursive one (2.2.7).

## 2.4 Conclusion

In this chapter, we tackled the estimation of the conditional extreme value index $\gamma(x)$ of a heavy-tailed distribution when some random covariate information is available. We

| | ISE | IAE |
|---|---|---|
| Recursive estimator | **0.04173758** | **0.1660733** |
| Non recursive estimator | 0.0454373 | 0.1722541 |

Table 2.2: The comparaison between errors of the non recursive estimator (2.1.1) and the proposed recursive estimator (2.2.7).

elaborate recursive kernel estimator of the extreme value index function based on the stochastic approximation algorithm. The proposed estimator asymptotically follows normal distribution. We subsequently compared the proposed estimator to Hill's non recursive extreme value index estimator. We demonstrated that using some particular stepsizes and a specific bandwidth selection through a cross-validation procedure, the proposed recursive estimator could be very competitive to the non recursive version. Moreover, we highlighted that the proposed estimator is much better in terms of computational costs. Numerical results illustrate the effectiveness of our recursive approach. To this extent, we would state that although our work is an extension of a wealthy historical background, it may be taken further, extended and built upon since it offers different perspectives and opens new horizons for future research. We can extend our recursive extreme value index estimator to the case of censored data. We can also propose a new estimator of the conditional extreme quantile using our recursive estimator defined by (2.2.7) and compare it to the classical Weissman estimator. Another direction is to investigate the almost sure convergence and the large and moderate deviation principles of the proposed estimator, which requires non trivial mathematics. This would go well beyond the scope of the present chapter.

## 2.5 Proofs

We introduce the following Lemmas that will enable us to obtain the asymptotic expansion of $a_n$.

**Lemma 2.5.1.** *Let assumption **(C3)** holds. Then, for $t_n \longrightarrow \infty$ as $n \longrightarrow \infty$ we have*

$$m_n(x) = \gamma(x)\overline{F}(t_n|x).$$

The proof of Lemma 2.5.1 is presented in Goegebeur et al. (2014).

**Lemma 2.5.2.** *Let assumptions **(C1)-(C6)** hold. Then, for all $x \in \mathbb{R}^d$ such that $g(x) > 0$ we have for $t_n \underset{n\to\infty}{\longrightarrow} \infty$ and $h_n \underset{n\to\infty}{\longrightarrow} 0$ with $h_n \ln t_n \underset{\to\infty}{\longrightarrow} 0$*

$$\widetilde{m}_n(x) \quad = \quad m_n(x) g(x) \left(1 + O(h_n \ln t_n)\right). \tag{2.5.1}$$

The proof of Lemma 2.5.2 is presented in Goegebeur et al. (2014).

**Lemma 2.5.3.** *Let assumptions **(C1)** and **(C4)-(C6)** hold. Then, for all $x \in \mathbb{R}^d$ such that $g(x) > 0$, we have for $t_n \underset{n\to\infty}{\longrightarrow} \infty$ and $h_n \underset{n\to\infty}{\longrightarrow} 0$ with $h_n \ln t_n \underset{n\to\infty}{\longrightarrow} 0$*

$$\mathbb{E}\left[K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}\right] = g(x) \overline{F}(t_n|x) \left(1 + O(h_n \ln t_n)\right).$$

## 2.5.1 Proof of Lemma 2.5.3

Since $(X_i, Y_i)$, $i = 1, \cdots, n$ are independent and identically distributed, we have under the assumption **(C6)**

$$\begin{aligned}
\mathbb{E}\left[K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}\right] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} \frac{1}{h_n^d} K\left(\frac{x-t}{h_n}\right) \mathbb{1}_{\{y > t_n\}} f(y|t) g(t) dt dy \\
&= \int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{x-t}{h_n}\right) \overline{F}(t_n|t) g(t) dt \\
&= \int_{\Omega} K(u) \overline{F}(t_n|x - uh_n) g(x - uh_n) du.
\end{aligned}$$

Now, we consider

$$\begin{aligned}
\left|\mathbb{E}\left[K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}\right] - \overline{F}(t_n|x) g(x)\right| &\leq \overline{F}(t_n|x) \int_{\Omega} K(u) \left|g(x - h_n u) - g(x)\right| du \\
&+ \overline{F}(t_n|x) \int_{\Omega} K(u) \left|\frac{\overline{F}(t_n|x - uh_n)}{\overline{F}(t_n|x)} - 1\right| g(x - h_n u) du \\
&:= \tilde{\mathbb{J}}_1 + \tilde{\mathbb{J}}_2.
\end{aligned}$$

Under the assumption **(C5)**, and since $g(x) > 0$, we have

$$\tilde{\mathbb{J}}_1 \leq \overline{F}(t_n|x) c_g h_n \int_{\Omega} \|u\|_2 K(u) du = \overline{F}(t_n|x) g(x) O(h_n). \tag{2.5.2}$$

Concerning $\tilde{\mathbb{J}}_2$, under **(C5)** and using this equation

$$\frac{\overline{F}(y|x - h_n z)}{\overline{F}(y|x)} = \exp\left[\ln \overline{F}(y|x) \left(\frac{\ln \overline{F}(y|x - h_n z)}{\ln \overline{F}(y|x)} - 1\right)\right],$$

it comes that

$$\left|\frac{\overline{F}(t_n|x - uh_n)}{\overline{F}(t_n|x)} - 1\right| \quad \leq \quad \left|\exp\left[C h_n \ln t_n\right] - 1\right|.$$

Applying Taylor, we get:

$$\sup_{u\in\Omega}\left|\frac{\overline{F}(t_n|x-uh_n)}{\overline{F}(t_n|x)}-1\right| = O(h_n\ln t_n).$$

and therefore, in view of (2.5.2),

$$\begin{aligned}\tilde{\mathbb{J}}_2 &= g(x)\,\overline{F}(t_n|x)O(h_n\ln t_n)\int_\Omega K(u)\frac{g(x-h_nu)}{g(x)}du\\ &= g(x)\,\overline{F}(t_n|x)O(h_n\ln t_n)(1+o(1))\\ &= g(x)\,\overline{F}(t_n|x)O(h_n\ln t_n).\end{aligned}$$

Then, we get

$$\mathbb{E}\left[K_{h_n}(x-X_n)\mathbb{1}_{\{Y_n>t_n\}}\right] = g(x)\,\overline{F}(t_n|x)\left(1+O(h_n\ln t_n)\right).$$

We state now the following technical lemma, which is proved in Mokkadem et al. (2009a), and which will be used throughout the demonstrations.

**Lemma 2.5.1.**
*Let $(v_n)\in\mathcal{GS}(v^*)$, $(\gamma_n)\in\mathcal{GS}(-\alpha)$ and $m>0$ such that $m-v^*\varepsilon>0$ where $\varepsilon$ is defined in* (4.2.1)*, and $\Pi_n$ in* (2.2.3)*. Then,*

$$\lim_{n\to\infty} v_n\Pi_n^m\sum_{k=1}^n\Pi_k^{-m}\frac{\gamma_k}{v_k} = \frac{1}{m-v^*\varepsilon}.$$

*Moreover, for all positive sequences $(\alpha_n)$ such that $\lim_{n\to\infty}\alpha_n=0$, and all $C\in\mathbb{R}$,*

$$\lim_{n\to\infty} v_n\Pi_n^m\left[\sum_{k=1}^n\Pi_k^{-m}\frac{\gamma_k}{v_k}\alpha_k + C\right] = 0.$$

## 2.5.2 Proof of Theorem 2.2.1

1. The application of Lemma 2.5.2, ensures that

$$\mathbb{E}(a_n(x)) = \Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_k\tilde{m}_k(x) = \Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_k m_k(x)g(x)\left(1+O(h_k\ln t_n)\right).$$

In the case $p\in(0,\alpha/(d+2)]$, we have $\lim_{n\to\infty} n\gamma_n > p$; the application of lemma 2.5.1 ensures that

$$\mathbb{E}(a_n(x)) \;=\; a(x)+O(h_n\ln t_n),$$

and (2.2.10) follows. In the case $p\in(\alpha/(d+2),1/d)$, we have $h_n\ln t_n = o\left(\sqrt{\gamma_n h_n^{-d}}\right)$, Lemma 2.5.1 ensures $\mathbb{E}(a_n(x))-a(x) = O\left(\sqrt{\gamma_n h_n^{-d}}\right)$, which gives (2.2.11).

2. Now, we have

$$\begin{aligned}\mathbb{V}ar(a_n(x)) &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left[\mathbb{E}\left[h_k^{-2d} K^2\left(\frac{x-X_k}{h_k}\right)[\ln Y_k - \ln t_n]^2 \mathbb{1}_{\{Y_k>t_n\}}\right]\right. \\ &\left. - \mathbb{E}^2\left[h_k^{-d} K\left(\frac{x-X_k}{h_k}\right)[\ln Y_k - \ln t_n] \mathbb{1}_{\{Y_k>t_n\}}\right]\right].\end{aligned}$$

Following the same steps of the proof of Theorem 1 in Goegebeur et al. (2014), we obtain

$$\begin{aligned}\mathbb{V}ar(a_n(x)) &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left[6\frac{\|K^2\|_1}{h_k^d}\gamma^2(x)\overline{F}(t_n|x)g(x)(1+o(1))\right] \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k \frac{\gamma_k}{h_k^d}\left[6\|K^2\|_1\gamma^2(x)\overline{F}(t_n|x)g(x)(1+o(1))\right].\end{aligned}$$

In the case when $p \in [\alpha/(d+2), 1/d)$, we have $\lim_{n\to\infty} n\gamma_n > \frac{\alpha - pd}{2}$, and the application of Lemma 2.5.1 ensures that

$$\begin{aligned}\mathbb{V}ar(a_n(x)) &= \frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1 g(x)\gamma^2(x)\overline{F}(t_n|x)\gamma_n h_n^{-d} \\ &+ \frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1\overline{F}(t_n|x)g(x)\gamma^2(x)o(\gamma_n h_n^{-d}),\end{aligned}$$

which proves (2.2.13). In the case when $p \in (0, \alpha/(d+2))$, we have $\gamma_n h_n^{-d} = o(h_n^2 \ln^2 t_n)$, Lemma 2.5.1 ensures that $\mathbb{V}ar(a_n(x)) = o(h_n^2 \ln^2 t_n)$, which yields (2.2.12).

### 2.5.3 Proof of Theorem 2.2.2

1. First, the application of Lemma 2.5.3 provides

$$\mathbb{E}(b_n(x)) = Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k g(x)\overline{F}(t_n|x)(1+O(h_k \ln t_n)).$$

Now, in the case when $p \in (0, b/(d+2)]$, we have $\lim_{n\to\infty} n\beta_n > p$; the application of Lemma 2.5.1 ensures that

$$\mathbb{E}(b_n(x)) = b(x) + O(h_n \ln t_n),$$

and (2.2.14) follows. In the case when $p \in (b/(d+2), 1/d)$, we have $h_n \ln t_n = o(\sqrt{\beta_n h_n^{-d}})$, Lemma 2.5.1 ensures $\mathbb{E}(b_n(x)) = b(x) + O\left(\sqrt{\beta_n h_n^{-d}}\right)$, which gives (2.2.15).

2. Now, we have

$\mathbb{V}ar(b_n(x))$

$$
\begin{aligned}
&= Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 \left[\mathbb{E}\left[h_k^{-2d} K^2\left(\frac{x-X_k}{h_k}\right)\mathbb{1}_{\{Y_k>t_n\}}\right] - \mathbb{E}^2\left[h_k^{-d} K\left(\frac{x-X_k}{h_k}\right)\mathbb{1}_{\{Y_k>t_n\}}\right]\right] \\
&= Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 \left[\frac{\|K\|_2^2}{h_k^d}\mathbb{E}\left[h_k^{-d} H\left(\frac{x-X_k}{h_k}\right)\mathbb{1}_{\{Y_k>t_n\}}\right] - \mathbb{E}^2\left[h_k^{-d} K\left(\frac{x-X_k}{h_k}\right)\mathbb{1}_{\{Y_k>t_n\}}\right]\right],
\end{aligned}
$$

with $H(\cdot) =: \frac{K^2(\cdot)}{\|K\|_2^2}$ also satisfying assumption **(C6)**. Using Lemma 2.5.3, we get

$$
\begin{aligned}
\mathbb{V}ar(b_n(x)) &= Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 \left[\frac{\|K\|_2^2}{h_k^d}\left[g(x)\,\overline{F}(t_n|x)\,(1+O(h_k \ln t_n))\right]\right. \\
&\quad \left. - g^2(x)\overline{F}^2(t_n|x)\,(1+O(h_k \ln t_n))\right],
\end{aligned}
$$

then, we have

$$
\begin{aligned}
\mathbb{V}ar(b_n(x)) &= \|K\|_2^2 g(x)\,\overline{F}(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2}\frac{\beta_k^2}{h_k^d} + \|K\|_2^2 g(x)\,\overline{F}(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 O\left(\frac{\ln t_n}{h_k^{d-1}}\right) \\
&\quad - g^2(x)\overline{F}^2(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 - g^2(x)\overline{F}^2(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2}\beta_k^2 O(h_k \ln t_n).
\end{aligned}
$$

In the case when $p \in [b/(d+2), 1/d)$, we have $\lim_{n\to\infty} n\beta_n > (b-pd)/2$, and the application of Lemma 2.5.1 gives

$$
\begin{aligned}
&\mathbb{V}ar(b_n(x)) \\
&= \frac{1}{2-(b-pd)\varepsilon_1}\|K\|_2^2 g(x)\,\overline{F}(t_n|x)\frac{\beta_n}{h_n^d} + O\left(\frac{\beta_n \ln t_n}{h_n^{d-1}}\right) \\
&\quad - \frac{1}{2-b\varepsilon_1} g^2(x)\overline{F}^2(t_n|x)\beta_n + O(\ln t_n \beta_n h_n),
\end{aligned}
$$

which proves (2.2.17). In the case when $p \in (0, b/(d+2))$, we have $\beta_n h_n^{-d} = o(h_n^2 \ln^2 t_n)$, Lemma 2.5.1 ensures that $\mathbb{V}ar(b_n(x)) = O(h_n^2 \ln^2 t_n)$, which gives (2.2.16).

### 2.5.4 Proof of Theorem 2.2.3

Let us first note that, for $x$ such that $b_n(x) \neq 0$, we have

$$
\widehat{\gamma}_n(x) - \gamma(x) = D_n(x)\frac{b(x)}{b_n(x)}, \tag{2.5.3}
$$

with

$$
D_n(x) = \frac{1}{b(x)}(a_n(x) - a(x)) - \frac{\gamma(x)}{b(x)}(b_n(x) - b(x)). \tag{2.5.4}
$$

It follows from (2.5.3), that the asymptotic behavior of $\widehat{\gamma}_n(x) - \gamma(x)$ can be deduced from the one of $D_n(x)$. Then, (2.2.18) follows from (2.2.10), (2.2.14) and (2.5.3) whereas (2.2.19) follows from (2.2.11), (2.2.15) and (2.5.3). Now it follows from (2.5.4) that

$$\mathbb{V}ar(D_n(x)) = \frac{1}{b^2(x)}\mathbb{V}ar(a_n(x)) - \frac{2\gamma(x)}{b^2(x)}\mathbb{C}ov(a_n(x), b_n(x)) + \frac{\gamma^2(x)}{b^2(x)}\mathbb{V}ar(b_n(x)). \tag{2.5.5}$$

By using Lemma 2.5.1 and choosing the stepsize $(\gamma_n) = (n^{-1})$, computations provide

$$\mathbb{C}ov(a_n(x), b_n(x)) = \frac{1}{n}Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k \left(\mathcal{J}_1 - \mathcal{J}_2\mathcal{J}_3\right), \tag{2.5.6}$$

with

$$\mathcal{J}_1 = \mathbb{E}\left[K_{h_k}^2\left(x - X_k\right)\left[\ln Y_k - \ln t_n\right]\mathbf{1}_{\{Y_k > t_n\}}\right], \ \mathcal{J}_2 = \widetilde{m}_n(x) \text{ and } \mathcal{J}_3 = \mathbb{E}\left[K_{h_k}\left(x - X_k\right)\mathbb{1}_{\{Y_k > t_n\}}\right].$$

Following similar steps as Lemma 2 in Goegebeur et al. (2014) and Lemma 2.5.2, we infer that

$$\mathcal{J}_1 = m_n(x)g\left(x\right)\frac{\|K\|_2^2}{h_k^d}\left(1 + O(h_k \ln t_n)\right),$$

$\mathcal{J}_2$ and $\mathcal{J}_3$ are already calculated in Lemmas 2.5.2 and 2.5.3. Then, the combination of (2.5.4), (2.5.5), (2.2.13), (2.2.17) and (2.5.6), gives (2.2.21), and the combination of (2.5.4), (2.5.5), (2.2.12), (2.2.16) and (2.5.6), gives (2.2.20).

### 2.5.5 Proof of Theorem 2.2.4

Let us at first assume that, if $p \geqslant \alpha/(d+2)$, then

$$\sqrt{\overline{F}(t_n|x)^{-1}}\gamma_n^{-1}h_n^d\left(\widehat{\gamma}_n(x) - \mathbb{E}\left(\widehat{\gamma}_n(x)\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathcal{V}ar(x)\right). \tag{2.5.7}$$

In the case when $p > \alpha/(d+2)$, Part 1 of the theorem follows from the combination of (2.2.19) and (2.5.7). In the case when $p = \alpha/(d+2)$, Parts 1 and 2 of the Theorem follow from the combination of (2.2.18) and (2.5.7). In the case $p < b/(d+2)$, (2.2.20) implies that

$$\frac{1}{h_n \ln t_n}\left(\widehat{\gamma}_n(x) - \mathbb{E}\left(\widehat{\gamma}_n(x)\right)\right) \xrightarrow{\mathbb{P}} 0,$$

and the application of (2.2.18) gives Part 2 of Theorem. Now (2.5.7) is proved. Relying on (2.5.4), we have

$$D_n(x) - \mathbb{E}[D_n(x)] = \frac{1}{b(x)}\Pi_n \sum_{k=1}^{n}\left(\mathcal{Y}_k(x) - \mathbb{E}[\mathcal{Y}_k(x)]\right),$$

where

$$\mathcal{Y}_k(x) = \Pi_k^{-1}\left(\gamma_k \mathcal{Z}_k(x) - \gamma(x)\eta_n\eta_k^{-1}\beta_k\mathcal{W}_k(x)\right),$$

with $\mathcal{Z}_n(x) = K_{h_n}(x - X_n)[\ln Y_n - \ln t_n]\mathbb{1}_{\{Y_n > t_n\}}$, $\mathcal{W}_n(x) = K_{h_n}(x - X_n)\mathbb{1}_{\{Y_n > t_n\}}$ and $\eta_n = \Pi_n^{-1} Q_n$. Now, in the case when $(\beta_n) = \left(n^{-1}\right)$, we have $\eta_n = (n\Pi_n)^{-1}$ and $\eta_k^{-1}\beta_k = \Pi_k$. Then,

$$\mathcal{Y}_k(x) = \Pi_k^{-1}\gamma_k \mathcal{Z}_k(x) - \gamma(x)(n\Pi_n)^{-1}\mathcal{W}_k(x).$$

Set

$$T_k(x) = \mathcal{Y}_k(x) - \mathbb{E}\left[\mathcal{Y}_k(x)\right]. \tag{2.5.8}$$

Moreover, we have

$$\begin{aligned}
s_n^2 &= \sum_{k=1}^n \mathbb{V}ar\left(T_k(x)\right) \\
&= \sum_{k=1}^n \Pi_k^{-2}\gamma_k^2 \mathbb{V}ar\left(\mathcal{Z}_k(x)\right) + \gamma^2(x)(n\Pi_n)^{-2}\sum_{k=1}^n \mathbb{V}ar\left(\mathcal{W}_k(x)\right) \\
&\quad -2\gamma(x)(n\Pi_n)^{-1}\sum_{k=1}^n \Pi_k^{-1}\gamma_k \mathbb{C}ov\left(\mathcal{Z}_k(x), \mathcal{W}_k(x)\right) \\
&:= \Gamma_1 + \Gamma_2 + \Gamma_3.
\end{aligned}$$

In addition, classical computations and applications of Lemma 2.5.1 ensure that

$$\begin{aligned}
\Gamma_1 &= \Pi_n^{-2}\gamma^2(x)\left[\frac{6}{2 - (\alpha - pd)\varepsilon}\|K^2\|_1 g(x)\,\overline{F}(t_n|x)\frac{\gamma_n}{h_n^d} + o\left(\frac{\gamma_n}{h_n^d}\right)\right], \\
\Gamma_2 &= \Pi_n^{-2}\gamma^2(x)\left[\frac{1}{1 + pd}\|K\|_2^2 g(x)\,\frac{\overline{F}(t_n|x)}{nh_n^d} + o\left(\frac{1}{nh_n^d}\right)\right], \\
\Gamma_3 &= \Pi_n^{-2}\gamma^2(x)\left[\frac{2}{1 + pd\varepsilon}\|K\|_2^2 g(x)\,\frac{\overline{F}(t_n|x)}{nh_n^d} + o\left(\frac{1}{nh_n^d}\right)\right].
\end{aligned}$$

As a matter of fact, we infer that

$$s_n^2 = \Pi_n^{-2} b^2(x)\overline{F}(t_n|x)\frac{\gamma_n}{h_n^d}\left[\mathcal{V}ar(x) + o(1)\right].$$

On the other side, we have, for all $q > 0$,

$$\mathbb{E}\left[\mid \mathcal{Y}_k(x)\mid^{2+q}\right] = O\left(\frac{1}{h_k^{(1+q)d}}\right),$$

and, since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - pd)/2$, there exists $q > 0$ such that $\lim_{n\to\infty} n\gamma_n > \frac{1+q}{2+q}(\alpha - pd)$. Applying Lemma 2.5.1, we get

$$\sum_{k=1}^n \mathbb{E}\left[\mid T_k(x)\mid^{2+q}\right] = O\left(\sum_{k=1}^n \Pi_k^{-2-q}\gamma_k^{2+q}\mathbb{E}\left[\mid \mathcal{Y}_k(x)\mid^{2+q}\right]\right) = O\left(\frac{\gamma_n^{1+q}}{\Pi_n^{2+q}h_n^{(q+1)d}}\right),$$

and we thus obtain

$$\frac{1}{s_n^{2+q}}\sum_{k=1}^n \mathbb{E}\left[\mid T_k(x)\mid^{2+q}\right] = \frac{1}{s_n^{2(1+q/2)}} O\left(\frac{\gamma_n^{1+q}}{\Pi_n^{2+q}h_n^{(q+1)d}}\right) = O\left(\gamma_n^{\frac{q}{2}} h_n^{-\frac{dq}{2}}\right) = o(1).$$

The convergence in (2.5.7) then follows from the application of Lyapounov's Theorem.

# Chapter 3

# The Stochastic Approximation Method for Estimation of a Distribution Function under $\alpha$-mixing condition

***Abstract***: *In this chapter, we extend the work of Slaoui (2014b) [The stochastic approximation method for the estimation of a distribution function. Math. Methods Statist., 23, 306-325] to the case of $\alpha$-mixing data. Then, we study the properties of these estimators and compare them with Nadaraya's non recursive distribution estimator. We show that, using some optimal parameters, the recursive estimators allowed us to obtain quite better results compared to the non recursive distribution estimator under $\alpha$-mixing condition in terms of estimation error. We establish the central limit theorem and the uniform convergence for the proposed estimators under some mild conditions. Finally, we corroborate these theoretical results through a few simulations.*

***Keywords***: Asymptotic normality, Bandwidth selection, Mixing Data, Recursive distribution estimator, Stochastic approximation algorithm.

## 3.1 Introduction

Non parametric distribution function methods have a central position in statistics, and an enormous literature exists in this subject. Non parametric kernel type methods have been widely used in estimating distribution function. We can list without to trying to be exhaustive Nadaraya (1964), Azzalini (1981), Reiss (1981), Sarda (1993), Bowman et al (1998) and Slaoui (2014b). This estimation has been widely applied in many disciplines such as economics, finances, medicine, biology and various other situations.

In the current chapter, we consider the case of non parametric estimation of the distribution function, using a recursive kernel estimator version. In this way, the estimator can be updated with each additional new observation. This recursive scheme offers many advantages to recursive estimators: they do not require extensive storage of data and they are fast to compute. In particular cases, they also appear as more efficient than classical estimators. Let $X_1, \cdots, X_n$ a sequence of random variables satisfy the $\alpha$-mixing dependency property (see Definition 1.2.1) having a common unknown distribution function $F$ with associated density $f$. Let us introduce a kernel function $K$ defined on $\mathbb{R}$ (that is, a function satisfying $\int_{\mathbb{R}} K(x)dx = 1$), a function $\mathcal{K}$ (defined by $\mathcal{K}(z) = \int_{-\infty}^{z} K(u)du$) and a bandwidth $(h_n)$ (that is, a sequence of positive real number tending to zero when $n$ goes to $\infty$). Let us recall that the usual kernel estimate of $F(x)$ is given by $\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{x-X_i}{h_n}\right)$.

Whereas to construct a stochastic algorithm, which approximates the function $F$ at a given point $x$, we define an algorithm of search of the zero of the function $\phi : y \mapsto F(x) - y$. We thus proceed in the following way:

(i) We set $F_0(x) \in [0, 1]$;

(ii) For all $n \geqslant 1$, we set $F_n(x) = F_{n-1}(x) + \gamma_n T_n(x)$, where the stepsize $(\gamma_n)$ is a positive sequence of real numbers decreasing to zero and $T_n$ is an observation of the function $\phi$ at the point $F_{n-1}(x)$.

Now, to define $T_n(x)$, we follow the approach of Révész (1973, 1977), Tsybakov (1990) and more recently Slaoui (2014a,b) and we set $T_n(x) = \mathcal{K}\left(h_n^{-1}[x - X_n]\right) - F_{n-1}(x)$.
Then, the estimator $F_n$ to recursively estimate the distribution function $F$ at the point $x$ can thus be written as

$$F_n(x) = (1 - \gamma_n)F_{n-1}(x) + \gamma_n \mathcal{K}\left(\frac{x - X_n}{h_n}\right). \tag{3.1.1}$$

This estimator was proposed by Slaoui (2014b) in the case of independent data and whose large and moderate deviation principles was obtained by Slaoui (2019b). Moreover, we consider for simplicity that $F_0(x) = 0$ and $\Pi_n = \prod_{j=1}^{n}(1 - \gamma_j)$. We can also consider the problem of bias reduction (see for instance the recent work of Slaoui (2018b), this would go well beyond the scope of the present chapter. By iteration, the estimator $F_n$ defined by (3.1.1) can be rewritten as

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k \mathcal{K}\left(\frac{x - X_k}{h_k}\right). \tag{3.1.2}$$

Moreover, in the case when the stepsize $(\gamma_n)$ is chosen equal to $(n^{-1})$, the estimator $F_n$ defined by (3.1.2) can be rewritten as

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathcal{K}\left(\frac{x - X_k}{h_k}\right). \tag{3.1.3}$$

This estimator was considered by Isogai and Hirose (1994). The choice of such stepsize belongs to the subclass of the recursive kernel estimators of density, which have a minimum MSE (Mean Squared Error) or MISE (Mean Integrated Squared Error) (see Mokkadem et al. (2009a)).
The aim of this chapter is to study the properties of the recursive distribution function estimator (3.1.3) in the case $\alpha$-mixing data and its comparison with the kernel distribution estimator defined as

$$\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}\left(\frac{x - X_i}{h_n}\right). \tag{3.1.4}$$

A number of authors have studied the properties of the estimator (3.1.4) (see Nadaraya (1964), Reiss (1981) and Hill (1985)).

In recent years, data flows have become increasingly important in the field of research. In this situation, the data arrives so fast that it is impossible to store it in a traditional database. In such a situation, the construction of a recursive estimator that does not require the storage of all the data in memory and that can be easily updated to process the online data is of great interest. This recursive estimator shows good theoretical properties, from the point of view of Mean Weighted Integrated Squared Error (MWISE) and almost sure convergence.

The purpose of this chapter is to generalize the recursive estimators proposed by Slaoui (2014b) to the case of $\alpha$-mixing data. We first compute the bias and the variance of the estimator $F_n$ defined by (3.1.3). It turns out that they heavily depend on the choice of the stepsize $(\gamma_n)$. We show that using an adequate choice of the bandwidth $(h_n)$, the expansion of the MWISE of the proposed estimator $F_n$ will be smaller than that of Nadaraya's estimator (3.1.4). We show also that estimator (3.1.3) can be very competitive to the estimator (3.1.4) in terms of estimation error and much better in terms of computational costs, especially for large $n$. This chapter is organized into five sections. We study asymptotic properties in Section 2, while Section 3 is devoted to our application results. We conclude the chapter in Section 4, whereas the technical details are deferred to Section 5.

## 3.2 Assumptions and Main Results

Throughout this chapter, we consider stepsizes and bandwidths belonging to the following class of regularly varying sequences.

**Definition 3.2.1.** *Let $u \in \mathbb{R}$ and $(u_n)_{n\geqslant 1}$ be a nonrandom positive sequence. We say that $u_n \in \mathcal{GS}(u)$ if*

$$\lim_{n\to\infty} n\left[1 - \frac{u_{n-1}}{u_n}\right] = u.$$

This condition was introduced by Galambos and Seneta (1973) and by Mokkadem et Pelletier. (2007) in the context of stochastic approximation algorithms. Note that the acronym

$\mathcal{GS}$ stand for (**G**alambos and **S**eneta). Typical sequences in $\mathcal{GS}(u)$ are, for $b \in \mathbb{R}$, $n^u(\log n)^b$, $n^u(\log\log n)^b$ and so on.

The assumptions to which we shall refer are the following:

**(H1)** $K : \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous bounded function satisfying $\int_{\mathbb{R}} K(x)dx = 1$, $\int_{\mathbb{R}} xK(x)dx = 0$ and $\int_{\mathbb{R}} x^2K(x)dx < \infty$.

**(H2)**

(i) The stepsize $\gamma_n \in \mathcal{GS}(-\alpha)$ with $\alpha \in (1/2, 1]$.

(ii) $h_n \in \mathcal{GS}(-a)$ with $a \in (0, 1)$.

(iii) $\lim_{n\to\infty} n\gamma_n \in (\min(2a, (\alpha + a)/2), \infty]$.

**(H3)** The density $f$ is bounded, differentiable and $f'$ is bounded.

**(H4)** The stepsize $(\gamma_n)$ is a decreasing sequence and $\gamma_n \longrightarrow 0$ as $n \longrightarrow \infty$.

**(H5)** The mixing coefficient of the sequence $(X_i)$ is geometry-dependent and satisfies $\alpha(n) = O(n^{-\rho})$ for some $\rho > 3$.

**(H6)** There exist integer sequences $(p_n)_n$ and $(q_n)_n$ going to $\infty$ along with $n$ such that

$$q_n/p_n \longrightarrow 0, \quad q_nh_n \longrightarrow 0, \quad p_nh_n \longrightarrow \infty \quad as \quad n \longrightarrow \infty.$$

Moreover, for $w := w_n := \left[\frac{n}{p_n+q_n}\right]$ (where [.] is the integer of the formula), we have

$$w_nq_n^{-\rho} \longrightarrow 0 \quad as \quad n \longrightarrow \infty; \quad w_n\gamma_n \longrightarrow 0, \left(\frac{w\gamma_n^{-1}\Pi_n^2}{(p_n+q_n)^{\rho}}\right) \longrightarrow 0, \quad as \quad n \longrightarrow \infty.$$

**(H7)** $\sum_{k=1}^{n} \frac{\theta_k}{\theta_n} < \infty$, $\sum_{1\leqslant k\leqslant n} \left(\frac{\theta_k}{\theta_n}\right)^2 < \infty$, $\sum_{1\leqslant i<j\leqslant n} \frac{\theta_i\theta_j}{\theta_n^2} < \infty$ and $\sum_{1\leqslant k\leqslant n} \left(\frac{\theta_k}{\theta_n}\right)^2 h_k < \infty$ where $\theta_k^2 = \Pi_k^{-2}\gamma_k^2$.

Moreover, we use the following notations:

$$\xi = \lim_{n\to\infty} (n\gamma_n)^{-1}, \tag{3.2.1}$$

$$Z_n(x) = \mathcal{K}\left(\frac{x - X_n}{h_n}\right), \tag{3.2.2}$$

$$\mu_2(K) = \int_{\mathbb{R}} z^2K(z)dz, \quad \phi(K) = 2\int_{\mathbb{R}} zK(z)\mathcal{K}(z)dz,$$

$$C_1 = \int_{\mathbb{R}} f^2(x)dx, \quad C_2 = \int_{\mathbb{R}} \left(f'(x)\right)^2 f(x)dx,$$

$$V_F = \int_{\mathbb{R}} F(x)(1 - F(x))f(x)dx.$$

Our first result are given in the following propositions, which give the bias and the variance of $F_n$ respectively.

**Proposition 3.2.1.** *(Bias of $F_n$)*
*Let assumptions **(H1) – (H3)** hold, and assume that $f'$ is continuous at $x$. Then*

$$\begin{aligned}\mathbb{E}[F_n(x)] - F(x) &= \left[\frac{1}{2(1-2a\xi)}f'(x)\mu_2(K)h_n^2 + o(h_n^2)\right]\mathbb{1}_{\{a\in(0,\alpha/3]\}} \\ &\quad + o\left(\sqrt{\gamma_n h_n}\right)\mathbb{1}_{\{a\in(\alpha/3,1)\}} \qquad (3.2.3)\end{aligned}$$

**Proposition 3.2.2.** *(Variance of $F_n$)*
*Let Assumptions **(H1)-(H6)** hold. Then*

$$\begin{aligned}\mathbb{V}ar\left[F_n(x)\right] &= \frac{1}{2-\alpha\xi}F(x)(1-F(x))\gamma_n\mathbb{1}_{\{a\in[\alpha/4,1]\}} \\ &\quad - \frac{1}{2-(a+\alpha)\xi}f(x)\phi(K)\gamma_n h_n\mathbb{1}_{\{a\in[\alpha/3,1]\}} \\ &\quad + o(\gamma_n h_n)\mathbb{1}_{\{a\in[\alpha/3,1]\}} + o\left(h_n^4\right)\mathbb{1}_{\{a\in(0,\alpha/4)\}}. \qquad (3.2.4)\end{aligned}$$

Moreover, in order to measure the quality of our proposed estimator $F_n$ defined in (3.1.3), we consider the Mean Weighted Integrated Squared Error (MWISE):

$$\begin{aligned}MWISE(F_n) &= \mathbb{E}\int_{\mathbb{R}}(F_n(x) - F(x))^2 f(x)dx \\ &= \int_{\mathbb{R}}(\mathbb{E}(F_n(x)) - F(x))^2 f(x)dx + \int_{\mathbb{R}}\mathbb{V}ar\left(F_n(x)\right)f(x)dx.\end{aligned}$$

The following proposition gives the MWISE of the estimator $F_n$ for the mixing case.

**Proposition 3.2.3.** *Let Assumptions **(H1)-(H6)** hold, and assume that $f'$ is continuous at $x$.*

$$\begin{aligned}MWISE(F_n) &= \frac{1}{2-\alpha\xi}\gamma_n V_F\mathbb{1}_{\{a\in[\alpha/4,1]\}} - \frac{1}{2-(a+\alpha)\xi}\gamma_n h_n C_1\phi(K)\mathbb{1}_{\{a\in[\alpha/3,1]\}} \\ &\quad + \frac{1}{4(1-2a\xi)^2}h_n^4 C_2^2\mu_2^2(K)\mathbb{1}_{\{a\in(0,\alpha/3]\}} + o(\gamma_n h_n)\mathbb{1}_{\{a\in[\alpha/3,1]\}} \\ &\quad + o\left(h_n^4\right)\mathbb{1}_{\{a\in(0,\alpha/3]\}}. \qquad (3.2.5)\end{aligned}$$

The following remark follows immediately from the previous proposition.

**Remark 3.2.1.** *One can infer from* (3.2.3) *and* (3.2.4)*, that in the special case $a = \alpha/3$, we have an asymptotic expression of the bias and the variance of $F_n$, the same remark can be done from* (3.2.5)*, then, under the assumptions of the proposition 3.2.3, we have in the case when $a = \alpha/3$*

$$\begin{aligned}MWISE(F_n) &= \frac{1}{2-\alpha\xi}\gamma_n V_F - \frac{1}{2-(a+\alpha)\xi}\gamma_n h_n C_1\phi(K) + \frac{1}{4(1-2a\xi)^2}h_n^4 C_2^2\mu_2^2(K) \\ &\quad + o\left(h_n^4\right).\end{aligned}$$

*Then it comes that we can obtain an optimal bandwidth in the special case $a = \alpha/3$, which will be very helpful for practice.*

The following corollary is an immediate consequence of the previous remark. Now, we explicit the choices of $(h_n)$ which minimize the $MWISE$ of our proposed recursive estimator defined by (3.1.3).

**Corollary 3.2.1.** *Let Assumptions* ***(H1)-(H5)*** *hold. To minimize the* $MWISE$ *of* $F_n$*, the stepsize* $(\gamma_n)$ *must be chosen in* $\mathcal{GS}(-1)$*,* $\lim_{n\to\infty} n\gamma_n = \gamma_0$*, the bandwidth* $(h_n)$ *must equal*

$$(h_n)_n = \left(2^{-1/3}\left(\gamma_0 - \frac{2}{3}\right)^{1/3}\left(\frac{C_1\phi(K)}{C_2\mu_2^2(K)}\right)^{1/3} n^{-1/3}\right), \tag{3.2.6}$$

*and then the corresponding* $MWISE$ *is equal to*

$$MWISE[F_n] = n^{-1}V_F\left(\frac{\gamma_0^2}{2\gamma_0 - 1} - \frac{3}{4}\frac{1}{2^{4/3}}\frac{\gamma_0^2}{(\gamma_0 - 2/3)^{2/3}}\frac{C_1^{4/3}\phi(K)}{C_2^{1/3}V_F}n^{-1/3} + o(n^{-1/3})\right).$$

## 3.3 Practical bandwidth selection

In order to give more details on the practical implementation of the proposed algorithm, we give first a data driven bandwidth selection procedure to estimate the optimal bandwidth (3.2.6), we must estimate $C_1$ and $C_2$. We followed the approach proposed in Slaoui (2014a,b), and we use the following kernel estimator of $C_1$ and $C_2$ respectively :

$$\widehat{C}_1 = \frac{\Pi_n}{n}\sum_{i,k=1}^{n}\Pi_k^{-1}\gamma_k b_k^{-1}K_b\left(\frac{X_i - X_k}{b_k}\right), \tag{3.3.1}$$

$$\widehat{C}_2 = \frac{\Pi_n^2}{n}\sum_{i,j,k=1}^{n}\Pi_j^{-1}\Pi_k^{-1}\gamma_j\gamma_k b_j'^{-2}b_k'^{-2}K_{b'}^{(1)}\left(\frac{X_i - X_j}{b_k}\right)K_{b'}^{(1)}\left(\frac{X_i - X_k}{b_k}\right), \tag{3.3.2}$$

where $K_b$ is a kernel and $b_n$ is the associated bandwidth, $K_{b'}^{(1)}$ is the first derivative of a kernel $K_{b'}$ and $b_n'$ the associated bandwidth. It was shown in Slaoui (2014a,b) that in order to minimize the $MISE$ (Mean Integrated Squared Error) of $\widehat{C}_1$ (resp. of $\widehat{C}_2$), $b_n$ (resp. $b_n'$) should belongs to $\mathcal{GS}(-2/5)$ (resp. $\mathcal{GS}(-3/10)$). In practice, we take

$$b_n = n^{-\beta}\min\left\{\widehat{s}, \frac{Q_3 - Q_1}{1.349}\right\}, \quad \beta \in (0,1) \tag{3.3.3}$$

(see Silverman (1986)) with $\widehat{s}$ the sample standard deviation, and $Q_1$, $Q_3$ denoting the first and third quartiles, respectively, here we take $\beta = 2/5$ to estimate $b_n$ and $\beta = 3/10$ to estimate $b_n'$. Moreover, in order to make a choice of the stepsize $(\gamma_n)$, we choose $\gamma_0 = 2/3 + \varepsilon$ such that $\varepsilon$ is close to zero to ensure that the MISE of the recursive estimator $F_n$ is smaller than the one of the non recursive estimator $\widetilde{F}_n$ (see Slaoui (2014b)).

We state the following theorem, which gives the weak convergence rate of our estimator (3.1.3).

**Theorem 3.3.1.** *Let Assumptions **(H1)**-**(H6)** hold.*

1. *If there exists $d \geqslant 0$ such that $\gamma_n^{-1} h_n^3 \longrightarrow d$, then*

$$\begin{aligned} &\sqrt{\gamma_n^{-1}}\left(F_n(x) - F(x)\right) \\ &\quad \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{d^{1/2}}{2(1-2a\xi)} f'(x)\mu_2(K), \frac{1}{2-\alpha\xi} F(x)(1-F(x))\right). \end{aligned} \tag{3.3.4}$$

2. *If $\gamma_n^{-1} h_n^3 \longrightarrow \infty$, then*

$$\frac{1}{h_n^2}\left(F_n(x) - F(x)\right) \xrightarrow{\mathbb{P}} \frac{1}{2(1-2a\xi)} f'(x)\mu_2(K).$$

**Remark 3.3.1.** *The asymptotic bias and the asymptotic variance of the considered estimator $F_n$ in the case of of dependent data ($\alpha$-mixing) are exactly the same as in the case of independent data, and consequently the convergence rate of $F_n$ in the two cases are the same, the main difference between the two cases are certainly linked to the considered assumptions, the ones used in the case of dependent data are much more stronger than the ones used in the case of independent data. Moreover, the data-driven bandwidth procedures are the same in the two cases.*

Let us underline that, when the bandwidth $(h_n)_n$ is chosen such that $\lim_{n\to+\infty} \gamma_n^{-1} h_n^3 = 0$ (which corresponds to undersmoothing) and using the stepsize $(\gamma_n)_n = \left(\gamma_0 n^{-1}\right)_n$, we infer from (3.3.4), that the considered estimator $F_n$ fulfils the following central limit theorem

$$\sqrt{n}\left(F_n(x) - F(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\gamma_0^2}{2\gamma_0 - 1} F(x)(1-F(x))\right).$$

We let $\phi$ denote the standard normal distribution function $\mathcal{N}(0,1)$, and $t_{\alpha/2}$ be such that $\phi\left(t_{\alpha/2}\right) = 1 - \alpha/2$ (where $\alpha \in (0,1)$). Then, the asymptotic confidence interval of $F(x)$ with level $1-\alpha$, is given by

$$\left[F_n(x) \pm t_{\alpha/2}\sqrt{C(\gamma_0)}\sqrt{\frac{F(x)(1-F(x))}{n}}\right],$$

where, $C(x) = x^2(2x-1)^{-1}$, this function reaching its minimum at the point $x = 1$. Then the best choice in point of view of estimation by confidence intervals is obtained by considering the stepsize $(\gamma_n)_n = \left(n^{-1}\right)_n$, using this choice, the estimator $F_n$ fulfils the central limit theorem

$$n^{1/2}\left(F_n(x) - F(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, F(x)(1-F(x))\right).$$

It comes that, the asymptotic confidence interval of $F(x)$ with level $1-\alpha$, in this special choice is given by

$$\left[F_n(x) \pm t_{\alpha/2}\sqrt{\frac{F(x)(1-F(x))}{n}}\right].$$

**Remark 3.3.2.** *We can observe that in the special case when the bandwidth $(h_n)_n$ is chosen such that $\lim_{n\to+\infty}\gamma_n^{-1}h_n^3 = 0$, and the stepsize $(\gamma_n)_n$ is chosen to be equal to $\left(n^{-1}\right)_n$, the estimator $F_n$ fulfills the same limit theorem as the one obtained for the empirical distribution, and consequently the two estimators ($F_n$ and the empirical distribution) have the same asymptotic confidence interval of $F$.*

**Theorem 3.3.2** (Uniform convergence)**.** *Let assumptions **(H1)** – **(H3)** hold, $F$ is uniformly continuous and there exists $\eta > 0$ such that $z \longrightarrow \|z\|^{\eta}|F(x)|$ is a bounded function. We let $\mathcal{C}$ be a compact set of $\mathbb{R}$. Then, we have*

$$\sup_{x\in\mathcal{C}}|F_n(x) - F(x)| = o(1) \quad a.s. \quad as\, n \longrightarrow \infty.$$

## 3.4 Simulation study

The aim of our applications is to compare the performance of Nadaraya's estimator defined in (3.1.4) with the proposed recursive kernel distribution function estimator under $\alpha$-mixing condition (1.2.1), defined in (3.1.3), using the Plug-in method of bandwidth selection.

### 3.4.1 The study design

We consider the following simulation design, we simulate $N = 500$ samples of sizes, $n = 50$, $n = 100$, $n = 150$ and a sequence of m-dependent variables

$$X_i = \sum_{j}^{i+m}\sqrt{|Z_j|},$$

where $(Z_j)_j$ are generated from the following mixture distribution:

$$Z \sim \frac{1}{2}\mathcal{N}(2.5, 6) + \frac{1}{2}\mathcal{N}(9, 1).$$

In order to calculate the MWISE of the proposed recursive kernel distribution function $F_n$, we need to use the following quantities:

- The Fonction $K$, we use the normal kernel.
- The stepsize $(\gamma_n) = (\gamma_0 n^{-1})$, where $\gamma_0 = 2/3 + c$, with $c \in [0, 1/3]$.
- The bandwidth $(h_n)_n$, we consider the plug-in method, given in Slaoui (2014b), that there is chosen to be equal to

$$\left(2^{-1/3}\left(\gamma_0 - \frac{2}{3}\right)^{1/3}\left(\frac{\widehat{C}_1\phi(K)}{\widehat{C}_2\mu_2^2(K)}\right)^{1/3} n^{-1/3}\right),$$

| | | $Y \sim \frac{1}{2}\mathcal{N}(2.5, 6) + \frac{1}{2}\mathcal{N}(9, 1)$ | | |
|---|---|---|---|---|
| | | $n = 50$ | $n = 100$ | $n = 150$ |
| | | $MWISE$ | $MWISE$ | $MWISE$ |
| $m = 2$ | | | | |
| | Non recursive | 0.003061 | 0.001641 | 0.001123 |
| | Recursive | **0.002391** | **0.001453** | **0.001040** |
| $m = 4$ | | | | |
| | Non recursive | 0.002877 | 0.001565 | 0.001087 |
| | Recursive | **0.002272** | **0.001396** | **0.001016** |
| $m = 6$ | | | | |
| | Non recursive | 0.002740 | 0.001510 | 0.001030 |
| | Recursive | **0.002176** | **0.001357** | **0.000972** |
| $m = 8$ | | | | |
| | Non recursive | 0.002666 | 0.001462 | 0.001025 |
| | Recursive | **0.002146** | **0.001319** | **0.000968** |

Table 3.1: MWISE (approximated using $N = 500$ trials) of the non recursive estimator and the recursive estimator.

with $\widehat{C}_1$ and $\widehat{C}_2$ are given in Slaoui (2014b) (see respectively equations (3.3.1) and (3.3.2)). Moreover, some numerical results of $\phi(K)$ and $\mu_2(K)$ are given for some standard kernels (see Table 1 in Slaoui (2014b)).

In order to calculate the MWISE of the non recursive kernel distribution function $\widetilde{F}_n$, we need to use these quantities:

- The Fonction $K$, we use the normal kernel.
- The bandwidth $(h_n)$, we consider the plug-in method (see Slaoui (2014b)).

### 3.4.2 Results

For each configuration of the simulation design parameters, we calculate the $MWISE$ of the non recursive estimator (3.1.4) and the recursive estimator (3.1.3). From Table 1, the proposed recursive estimator of $F_n(x)$ outperformed the non recursive estimator $\widetilde{F}(x)$ in all the considered situations. We can observe that the MWISE decrease as $m$ increase. We can observe also that the MWISE decrease as the sample size $n$ increase.

## 3.5 Conclusion

This chapter proposes an automatic bandwidth selection of the recursive density estimators under $\alpha$-mixing condition (1.2.1). The proposed estimators asymptotically follow normal distribution. The proposed estimators are compared to the non recursive distribution function estimator under $\alpha$-mixing condition. We showed that using a specific plug-in bandwidth selection method and some particularly stepsizes, the proposed recursive estimators can give better results compared to the non recursive distribution function estimator under $\alpha$-mixing condition in terms of estimation error. However, the main advantage of the recursive method is considerably faster than the classical one; see, for instance, Mokkadem et al. (2009a) and Slaoui (2014a) in the framework of density of probability estimation, Slaoui (2014b) in the framework of distribution estimation, Slaoui (2015, 2016a) in the framework of regression estimation and Slaoui (2016b) in the framework of hazard function. In conclusion, the proposed recursive estimators allowed us to obtain quite better results compared to the non recursive density estimator under $\alpha$-mixing condition in terms of estimation error and much better in terms of computational costs.

## 3.6 Proofs

Before giving the outlines of the proofs, we introduce the following technical lemma, which is proved in Mokkadem et al. (2009a), and which will be used throughout the demonstrations.

**Lemma 3.6.1.**
*Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$ and $m > 0$ such that $m - v^*\varepsilon > 0$ where $\varepsilon$ is defined in* (4.2.1)*, then*

$$\lim_{n\to\infty} v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\varepsilon}.$$

*Moreover, for all positive sequence $(\alpha_n)$ such that $\lim_{n\to+\infty} \alpha_n = 0$, and all $C \in \mathbb{R}$,*

$$\lim_{n\to\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + C\right] = 0.$$

### 3.6.1 Proof of Proposition 3.2.1

In view of (3.1.1) and (3.2.2), we have

$$\begin{aligned} F_n(x) - F(x) &= (1-\gamma_n)F_{n-1}(x) + \gamma_n Z_n(x) - F(x) \\ &= (1-\gamma_n)(F_{n-1}(x) - F(x)) + \gamma_n [Z_n(x) - F(x)] \end{aligned}$$

$$\begin{aligned}
&= \prod_{j=1}^{n}(1-\gamma_j)(F_0(x)-F(x)) + \sum_{i=1}^{n-1}\prod_{j=i+1}^{n}(1-\gamma_j)\gamma_i[Z_i(x)-F(x)] \\
&\quad +\gamma_n[Z_n(x)-F(x)] \\
&= \Pi_n \sum_{k=1}^{n}\Pi_k^{-1}\gamma_k(Z_k(x)-F(x)) + \Pi_n(F_0(x)-F(x)).
\end{aligned}$$

It implies that

$$\mathbb{E}[F_n(x)] - F(x) = \Pi_n \sum_{k=1}^{n}\Pi_k^{-1}\gamma_k(\mathbb{E}[Z_k(x)]-F(x)) + \Pi_n(F_0(x)-F(x)).$$

Then, an integration by parts ensures that

$$\begin{aligned}
\mathbb{E}[Z_k(x)] &= \int_{\mathbb{R}} \mathcal{K}\left(\frac{x-t}{h_k}\right) f(t)dt \\
&= \int_{\mathbb{R}} K(z)F\left(x+h_kz\right)dz. \qquad (3.6.1)
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{E}[Z_k(x)] - F(x) &= \int_{\mathbb{R}} K(z)\left[F(x+h_kz)-F(x)\right]dz \\
&= \frac{h_k^2}{2}f'(x)\mu_2(K) + \beta_k(x)
\end{aligned}$$

with

$$\beta_k(x) = \int_{\mathbb{R}} K(z)\left[F(x+h_kz)-F(x)-zh_kf(x)-\frac{1}{2}z^2h_k^2f'(x)\right]dz,$$

and, since $f'$ is bounded and continuous, we have $\lim_{k\to\infty}\beta_k(x)=0$.
In the case $a>\alpha/3$, we have $h_n^2 = o\left(\sqrt{\gamma_nh_n}\right)$ and $\lim_{n\to\infty}(n\gamma_n) > (a+\alpha)/2$, then Lemma 4.5.1 ensures that

$$\begin{aligned}
\mathbb{E}[F_n(x)] - F(x) &= \Pi_n \sum_{k=1}^{n}\Pi_k^{-1}\gamma_k o\left(\sqrt{\gamma_kh_k}\right) + O(\Pi_n) \\
&= o\left(\sqrt{\gamma_nh_n}\right). \qquad (3.6.2)
\end{aligned}$$

Moreover, in the case when $a \leqslant \alpha/3$, we have $\lim_{n\to\infty}(n\gamma_n) > 2a$; the application of Lemma 4.5.1 ensures that

$$\begin{aligned}
\mathbb{E}[F_n(x)] - F(x) &= \frac{1}{2}f'(x)\mu_2(K)\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k[h_k^2+o(1)] + \Pi_n(F_0(x)-F(x)) \\
&= \frac{1}{2(1-2a\xi)}f'(x)\mu_2(K)[h_n^2+o(1)]. \qquad (3.6.3)
\end{aligned}$$

Then, the combination of (3.6.2) and (3.6.3) give (3.2.3).

### 3.6.2 Proof of Proposition 3.2.2

First, we have

$$\begin{aligned}
\mathbb{V}ar\left[F_n(x)\right] &= \mathbb{V}ar\left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k Z_k\left(x\right)\right] \\
&= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \mathbb{V}ar\left(Z_k(x)\right) + 2\Pi_n^2 \sum_{1\leqslant i<j\leqslant n} \Pi_i^{-1}\gamma_i \Pi_j^{-1}\gamma_j \mathbb{C}ov\left(Z_i(x), Z_j(x)\right) \\
&=: I_1 + I_2,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \mathbb{V}ar\left(Z_k(x)\right), \\
I_2 &= 2\Pi_n^2 \sum_{1\leqslant i<j\leqslant n} \Pi_i^{-1}\gamma_i \Pi_j^{-1}\gamma_j \mathbb{C}ov\left(Z_i(x), Z_j(x)\right).
\end{aligned}$$

Now, in order to compute $I_1$, we use the following decomposition

$$I_1 = \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left(E\left(Z_k^2(x)\right) - E^2\left(Z_k(x)\right)\right). \tag{3.6.4}$$

An integration by parts ensures that

$$\begin{aligned}
E\left(Z_k^2(x)\right) &= \int_{\mathbb{R}} \mathcal{K}^2\left(\frac{x-t}{h_k}\right) f(t)dt \\
&= 2\int_{\mathbb{R}} \mathcal{K}\left(-z\right) K(z) F(x+zh_k)dz \\
&= v_k(x) + F(x) - h_k f(x)\phi(K)
\end{aligned} \tag{3.6.5}$$

with

$$v_k(x) = 2\int_{\mathbb{R}} K(z)\mathcal{K}(-z)\left[F(x+zh_k) - F(x) - zh_k f(x)\right]dz.$$

Moreover, it follows from (3.6.1), that

$$E\left(Z_k(x)\right) = F(x) + \widetilde{v}_k(x), \tag{3.6.6}$$

with

$$\widetilde{v}_k(x) = \int_{\mathbb{R}} K(z)\left[F(x+zh_k) - F(x)\right]dz.$$

Then, the combination of (3.6.4), (3.6.5) and (3.6.6) gives

$$I_1 = \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left[F(x) - h_k f(x)\phi(K) + v_k(x) - F^2(x) - \widetilde{v}_k^2(x) - 2F(x)\widetilde{v}_k(x)\right]$$

$$
\begin{aligned}
&= F(x)\left(1-F(x)\right)\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 - f(x)\phi(K)\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 h_k \\
&\quad + \left(v_k(x)-\widetilde{v}_k^2(x)-2F(x)\widetilde{v}_k(x)\right)\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2. \qquad (3.6.7)
\end{aligned}
$$

Since $f$, $f'$ are continuous and bounded, we have $\lim_{k\to\infty} v_k(x)=0$ and $\lim_{k\to\infty}\widetilde{v_k}(x)=0$.
In the case $a\geqslant \alpha/3$, we have $\lim_{n\to\infty} n\gamma_n > (a+\alpha)/2$, and the application of Lemma 4.5.1 gives

$$
I_1 = \frac{\gamma_n}{2-\alpha\xi}F(x)(1-F(x)) - \frac{\gamma_n h_n}{2-(a+\alpha)\xi}f(x)\phi(K)+o(\gamma_n h_n). \qquad (3.6.8)
$$

Moreover, in the case when $a\in[\alpha/4,\alpha/3)$, we have $\gamma_n h_n = o(h_n^4)$, and $\lim_{n\to\infty} n\gamma_n > \alpha/2$, Lemma 4.5.1 then ensures that

$$
\begin{aligned}
I_1 &= \frac{\gamma_n}{2-\alpha\xi}F(x)(1-F(x)) + \Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k o(h_k^4) \\
&= \frac{\gamma_n}{2-\alpha\xi}F(x)(1-F(x)) + o(\gamma_n). \qquad (3.6.9)
\end{aligned}
$$

Now, in the case when $a\in(0,\alpha/4)$, we have $\gamma_n = o(h_n^4)$, and $\lim_{n\to\infty} n\gamma_n > 2a$, then the application of Lemma 4.5.1 gives

$$
I_1 = \Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k o(h_k^4) = o(h_n^4). \qquad (3.6.10)
$$

Then, (3.2.4) follows from the combination of (3.6.8), (3.6.9) and (3.6.10). Let us now compute $I_2$, we have

$$
\begin{aligned}
I_2 &= 2\Pi_n^2\sum_{1\leqslant i<j\leqslant n}\Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j\mathbb{C}ov\left(Z_i(x),Z_j(x)\right) \\
&= 2\Pi_n^2\sum_{1\leqslant i<j\leqslant n} B_{i,j},s
\end{aligned}
$$

where $\quad B_{i,j}=\Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j\mathbb{C}ov\left(\mathcal{K}\left(\frac{x-X_i}{h_i}\right),\mathcal{K}\left(\frac{x-X_j}{h_j}\right)\right).$

$$
\begin{aligned}
|B_{i,j}| &= \Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j\left|\mathbb{C}ov\left(\mathcal{K}\left(\frac{x-X_i}{h_i}\right),\mathcal{K}\left(\frac{x-X_j}{h_j}\right)\right)\right| \\
&= \Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j\left|\int_{\mathbb{R}^2}\mathcal{K}\left(\frac{x-t_1}{h_i}\right)\mathcal{K}\left(\frac{x-t_2}{h_j}\right)f_{X_i,X_j}(t_1,t_2)dt_1dt_2\right. \\
&\quad \left. - \int_{\mathbb{R}}\mathcal{K}\left(\frac{x-t_1}{h_i}\right)f(t_1)dt_1\int_{\mathbb{R}}\mathcal{K}\left(\frac{x-t_2}{h_j}\right)f(t_2)dt_2\right|.
\end{aligned}
$$

Assumption **(H4)**, a simple change of variables and an integration by parts imply

$$
|B_{i,j}| \leqslant M\Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j\int_{\mathbb{R}^2}K\left(s_1\right)K\left(s_2\right)ds_1ds_2
$$

$$= \quad O(\Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j). \qquad (3.6.11)$$

Next, to evaluate the asymptotic behavior of $I_2$, we define the sets

$$F_1 = \{(i,j) \; such \; that \; 1 \leqslant |i-j| \leqslant \sigma_n\},$$

and

$$F_2 = \{(i,j) \; such \; that \; \sigma_n + 1 \leqslant |i-j| \leqslant n-1\}$$

where $\sigma_n = o(n)$. Let

$$J_{1,n} = 2\Pi_n^2 \sum_{(i,j)\in F_1} B_{i,j} \;\text{ and }\; J_{2,n} = 2\Pi_n^2 \sum_{(i,j)\in F_2} B_{i,j}.$$

Then it follows from (3.6.11) and **(H4)**, that

$$|J_{1,n}| \quad \leqslant \quad 2M\Pi_n^2 \sum_{(i,j)\in F_1} \Pi_i^{-1}\gamma_i\Pi_j^{-1}\gamma_j$$

and applying Lemma 4.5.1, we infer that

$$\begin{aligned} |J_{1,n}| \quad &\leqslant \quad 2M'\sigma_n\gamma_n\frac{1}{2-\alpha\xi} \\ &= \quad O\left(\frac{\sigma_n\gamma_n}{2-\alpha\xi}\right). \end{aligned}$$

In order to compute $J_{2,n}$, we use the Davydov inequality for mixing processes (see Rio (2000), p. 10, Formula 1.12a). This leads, for all $i \neq j$, to

$$\left|\mathbb{C}ov\left(\Pi_i^{-1}\gamma_i Z_i(x), \Pi_j^{-1}\gamma_j Z_j(x)\right)\right| \leqslant c\alpha\left(|i-j|\right).$$

Therefore, using **(H5)**, we have

$$\begin{aligned} J_{2,n} \quad &\leqslant \quad 2c\Pi_n^2 \sum_{j=1}^{n} \sum_{\sigma_n+1<k\leqslant n-1} \alpha\left(k\right) \\ &< \quad 2cn\Pi_n^2 \int_{\sigma_n+1}^{n-1} k^{-v}dk \\ &= \quad O\left(n\Pi_n^2\sigma_n^{1-v}\right). \end{aligned}$$

By choosing $\sigma_n = n^{1/v}\gamma_n^{-1/v}\Pi_n^{2/v}\left(\frac{1}{2-\alpha\xi}\right)^{-1/v}$, the assumption **(H5)**, ensures that

$$I_2 = J_{1,n} + J_{2,n} = O\left(\frac{n^{1/v}\gamma_n^{(v-1)/v}\Pi_n^{2/v}}{(2-\alpha\xi)^{(v-1)/v}}\right) = o(1),$$

which conclude the proof.

### 3.6.3 Proof of Theorem 3.3.1

We have

$$\begin{aligned}\sqrt{\gamma_n^{-1}}\left(F_n(x)-F(x)\right) &= \sqrt{\gamma_n^{-1}}\left(F_n(x)-\mathbb{E}\left(F_n(x)\right)\right)+\sqrt{\gamma_n^{-1}}\left(\mathbb{E}\left(F_n(x)\right)-F(x)\right)\\ &= \Upsilon_{1,n}+\Upsilon_{2,n}. \end{aligned} \tag{3.6.12}$$

First, we determine that $\Upsilon_{2,n}$ are negligible, whereas $\Upsilon_{1,n}$ is asymptotically normal.

In order to establish the asymptotic normality, dealing with strong mixing random variables (under **(H5)**), we use the well-known sectioning device introduced by (Doob (1955), p. 228-232). We first split the sum in (3.6.15) below into large $p_n$ blocks and small $q_n$ blocks under **(H6)**. For that, observe that $w_n(p_n+q_n)\leqslant n$ and $w_n(p_n+q_n)/n\longrightarrow 1$ as $n\longrightarrow\infty$ and, for $j=1,\cdots,w_n$, partition the set $\{1,2,\cdots,n\}$ into $(2w_n+1)$ subsets with $w=: w_n$ blocks of size $p_n$ and $k_n$ blocks of size $q_n$, as follows: Let

$$y_{mn}=\sum_{i=k_m}^{k_m+p-1}L_i(x),\qquad y'_{mn}=\sum_{j=l_m}^{l_m+q-1}L_j(x),\qquad y''_{wn}=\sum_{k=w(p+q)+1}^{n}L_k(x) \tag{3.6.13}$$

where $k_m=(m-1)(p+q)+1$, $l_m=(m-1)(p+q)+p+1$, $m=1,\cdots,w$. Let us first assume that if $a\geqslant\alpha/3$, we have

$$\sqrt{\gamma_n^{-1}}\left(F_n(x)-F(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\frac{1}{2-\alpha\xi}F(x)(1-F(x))\right). \tag{3.6.14}$$

In order to prove (3.6.14), we set $L_k(x)=\Pi_k^{-1}\gamma_k\left(Z_k(x)-\mathbb{E}(Z_k(x))\right)$, then, it comes from (3.1.1), that

$$\begin{aligned}F_n(x)-\mathbb{E}\left(F_n(x)\right) &= \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\left(Z_k(x)-\mathbb{E}(Z_k(x)\right)\\ &= \Pi_n\sum_{k=1}^{n}L_k(x). \end{aligned} \tag{3.6.15}$$

Moreover, since we have

$$\Upsilon_{1,n}(x)=\Pi_n\sqrt{\gamma_n^{-1}}\sum_{k=1}^{n}L_k(x).$$

We infer that,

$$\begin{aligned}\Upsilon_{1,n}(x) &= \Pi_n\sqrt{\gamma_n^{-1}}\left\{\sum_{m=1}^{w}y_{mn}+\sum_{m=1}^{w}y'_{mn}+y''_{wn}\right\}\\ &:= \Pi_n\sqrt{\gamma_n^{-1}}\left\{T_{n,1}+T_{n,2}+T_{n,3}\right\}. \end{aligned}$$

We let $\varphi_n = \Pi_n \sqrt{\gamma_n^{-1}}$. Let us first show that

$$\varphi_n \left(T_{n,2} + T_{n,3}\right) \xrightarrow{\mathbb{P}} 0, \quad as \quad n \longrightarrow \infty, \tag{3.6.16}$$

and then we prove that

$$\varphi_n T_{n,1} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{2-\alpha\xi} F(x)(1-F(x))\right). \tag{3.6.17}$$

To proof (3.6.16), we use Tchebychev's inequality. Then, we need to show that

$$\varphi_n^2 \mathbb{E}\left(T_{n,2}^2 + T_{n,3}^2\right) \longrightarrow 0, \quad n \longrightarrow \infty.$$

For the first term, we consider the following decomposition

$$\begin{aligned}
\varphi_n^2 \mathbb{E}\left(T_{n,2}^2\right) &= \varphi_n^2 \sum_{m=1}^{w} \sum_{i=l_m}^{l_m+q-1} \mathbb{E}\left(L_i^2(x)\right) + 2\varphi_n^2 \sum_{1 \leqslant i < j \leqslant w} \mathbb{E}\left(y'_{in} y'_{jn}\right) \\
&\quad + 2\varphi_n^2 \sum_{m=1}^{w} \sum_{l_m \leqslant i < j \leqslant l_m+q+1} \mathbb{C}ov\left(L_i(x) L_j(x)\right) \\
&= \Theta_1 + \Theta_2 + \Theta_3,
\end{aligned}$$

where

$$\begin{aligned}
\Theta_1 &= \varphi_n^2 \sum_{m=1}^{w} \sum_{i=l_m}^{l_m+q-1} \mathbb{E}\left(L_i^2(x)\right) \\
\Theta_2 &= 2\varphi_n^2 \sum_{1 \leqslant i < j \leqslant w} \mathbb{E}\left(y'_{in} y'_{jn}\right) \\
\Theta_3 &= 2\varphi_n^2 \sum_{m=1}^{w} \sum_{l_m \leqslant i < j \leqslant l_m+q+1} \mathbb{C}ov\left(L_i(x) L_j(x)\right).
\end{aligned}$$

The combination of (3.6.8), (3.6.11) and (3.6.13), together with **(H6)**, **(H7)**, ensure that

$$\Theta_1 \longrightarrow 0 \quad \text{and} \quad \Theta_3 \longrightarrow 0 \quad as \quad n \longrightarrow \infty.$$

Now, in order to compute $\Theta_2$, we use Rio (2000), p.10, Formula 1.12a, to infer that

$$\mathbb{E}\left(y'_i y'_j\right) \leqslant C\alpha\left(q_n + (j-i-1)(p_n+q_n)\right),$$

we then deduce that

$$\begin{aligned}
\Theta_2 &\leqslant C\gamma_n^{-1} \Pi_n^2 \sum_{i=1}^{w-1} \sum_{j=0}^{w-i-1} \left(q_n + (j-i-1)(p_n+q_n)\right)^{-\rho} \\
&\leqslant C\gamma_n^{-1} \Pi_n^2 (p_n+q_n)^{-\rho} \sum_{i=1}^{w-1} \sum_{l=0}^{\infty} l^{-\rho}
\end{aligned}$$

$$\leqslant \quad O\left(\frac{w\gamma_n^{-1}\Pi_n^2}{(q_n+q_n)^{\rho}}\right)=o(1).$$

Now, in order to proof (3.6.17), we show that

$$\left|\mathbb{E}\left[\exp\left(it\alpha_n T_{n,1}\right)\right]-\prod_{m=1}^{W_n}\mathbb{E}\left[\exp\left(it\alpha_n y_{mn}\right)\right]\right|\longrightarrow 0 \tag{3.6.18}$$

and

$$\alpha_n^2\sum_{m=1}^{W_n}\mathbb{E}\left[y_{mn}^2\mathbb{1}_{\left\{y_m>\xi\alpha_n^{-1}\sigma(x)\right\}}\right]\longrightarrow 0 \quad as \quad n\longrightarrow\infty, \tag{3.6.19}$$

where $\sigma^2(x)=\frac{1}{2-\alpha\xi}F(x)(1-F(x))$. Using Volkonskii and Rozanov (1959) inequality, we obtain

$$\left|\mathbb{E}\left[\exp\left(it\alpha_n\sum_{m=1}^{w_n}y_{mn}\right)\right]-\prod_{m=1}^{W_n}\mathbb{E}\left[\exp\left(it\alpha_n y_{mn}\right)\right]\right|\leqslant 16(w_n-1)\alpha(q_n)\longrightarrow 0 \quad as \quad n\longrightarrow\infty$$

which, under **(H6)**, yields to (3.6.18).
For (3.6.19), the combination of (3.6.13) together with (**H1**), (**H3**) and (**H6**), for $n$ large enough, ensure that the set $\left\{y_{mn}, n>\xi\alpha^{-1}\sigma(x)\right\}$ become empty which completes the proof (see Khardani and Slaoui (2019)).

### 3.6.4 Proof of Theorem 3.3.2

First, using the compactness property of the set $\mathcal{C}$, we use the fact that, for some $(x_k)_{1\leq k\leq\gamma_n}$, $\mathcal{C}\subset\bigcup_{k=1}^{\gamma_n}B\left(x_k,a_n\right)$, where $\gamma_n\sim a_n^{-1}$ with $a_n=h_n^{\frac{1}{\alpha}+1}$.

Moreover, for any $x\in\mathcal{C}$, we set $\widetilde{k}\left(x\right)=\arg\min_k\|x_k-x\|$. We infer that, for any $x\in\mathcal{C}$, we have

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|F_n\left(x\right)-\mathbb{E}\left[F_n\left(x\right)\right]\right| \quad \leq \quad & \sup_{x\in\mathcal{C}}\left|F_n\left(x\right)-F_n\left(x_{\widetilde{k}}\right)\right| \\
& +\sup_{x\in\mathcal{C}}\left|F_n\left(x_{\widetilde{k}}\right)-\mathbb{E}\left[F_n\left(x_{\widetilde{k}}\right)\right]\right| \\
& +\sup_{x\in\mathcal{C}}\left|\mathbb{E}\left[F_n\left(x_{\widetilde{k}}\right)\right]-\mathbb{E}\left[F_n\left(x\right)\right]\right| \\
& =: \mathcal{T}_{1,n}+\mathcal{T}_{2,n}+\mathcal{T}_{3,n}.
\end{aligned} \tag{3.6.20}$$

We let $\alpha$ denote the Hölder order of $\mathcal{K}$ and $\|\mathcal{K}\|_H$ its corresponding Hölder norm. Then, it follows from (3.1.2) that for any $x\in\mathcal{C}$

$$\left|F_n\left(x\right)-F_n\left(x_{\widetilde{k}}\right)\right| \quad \leq \quad \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\left|\mathcal{K}\left(\frac{X_k-x}{h_k}\right)-\mathcal{K}\left(\frac{X_k-x_{\widetilde{k}}}{h_k}\right)\right|$$

$$
\begin{aligned}
&\leq 2\left\|\mathcal{K}\right\|_{H}\Pi_{n}\sum_{k=1}^{n}\Pi_{k}^{-1}\gamma_{k}\left(\frac{\left\|x-x_{\widetilde{k}}\right\|}{h_{k}}\right)^{\alpha}\\
&\leq 2\left\|\mathcal{K}\right\|_{H}\Pi_{n}\sum_{k=1}^{n}\Pi_{k}^{-1}\gamma_{k}h_{k},
\end{aligned}
$$

we then get $\mathcal{T}_{1,n}=o\left(1\right)$ and $\mathcal{T}_{3,n}=o\left(1\right)$. Moreover, we set $\rho>0$ and $M$ such that $\left\|F_{\infty}\right\|\int_{\|z\|>M}\left|K\left(z\right)\right|dz\leq\rho/2$. Since the application of Lemma 4.5.1 ensures that $\Pi_{n}\sum_{k=1}^{n}\Pi_{k}^{-1}\gamma_{k}=1+o\left(1\right)$, then, it follows from (3.6.1)

$$
\begin{aligned}
\left|F_{n}\left(x_{\widetilde{k}}\right)-\mathbb{E}\left[F_{n}\left(x_{\widetilde{k}}\right)\right]\right| &\leq \Pi_{n}\sum_{k=1}^{n}\Pi_{k}^{-1}\gamma_{k}\left|\mathbb{E}\left[\mathcal{K}\left(\frac{X_{k}-x_{\widetilde{k}}}{h_{k}}\right)\right]\right|\\
&\leq \Pi_{n}\sum_{k=1}^{n}\Pi_{k}^{-1}\gamma_{k}\left|\int_{\mathbb{R}}K\left(z\right)F\left(x_{\widetilde{k}}+zh_{k}\right)dz\right|\\
&\leq \frac{\rho}{2}+\int_{\|z\|\leq M}\left|K\left(z\right)\right|\left|F\left(x_{\widetilde{k}}\right)\right|dz\\
&\quad+\Pi_{n}\sum_{k=1}^{n}\Pi_{k}^{-1}\gamma_{k}\int_{\|z\|>M}\left|K\left(z\right)\right|\left|F\left(x_{\widetilde{k}}+zh_{k}\right)-F\left(x_{\widetilde{k}}\right)\right|dz.
\end{aligned}
$$

Then, the uniform continuity of $F$ combined with the dominate convergence and the existence of $\eta>0$ such that $z\longrightarrow\|z\|^{\eta}\left|F\left(x\right)\right|$ is a bounded function ensure that $\mathcal{T}_{2,n}=o\left(1\right)$. Then the combination of Proposition 3.2.1 and (3.6.20) concludes the proof of Theorem 3.3.2.

# Chapter 4

# Strong consistency of the mode of multivariate recursive kernel density estimator under strong mixing hypothesis

***Abstract***: *In this research chapter, we attempt to define a kernel estimator of the mode based on the recursive kernel density estimator developed by Mokkadem et al. (2009a). In addition, we establish its almost sure convergence under strong mixing hypothesis. Finally, we corroborate these theoretical results through numerical simulations.*

***Keywords***: Density estimation, Mode, Non parametric estimation, Stochastic approximation, Strong consistency, Strong mixing.

## 4.1 Introduction

The estimation of mode function stands for a classical problem in statistics which has whetted considerable interest in various fields of applications. Indeed, it is widely used in machine learning applications and, in particular, in clustering methods (see Cheng (1995); Sheikh et al. (2007); Jiang and Kpotufe (2017)), computer vision (see Yin et al. (2003); Tao et al. (2007)), power systems (see Williams et al. (2001); Nezam Sarmadi and Venkatasubramanian (2014)), control (see Hofbaur and Williams (2002)) and bioinformatics (see Hedges and Shah (2003)). Multiple research works related to this topic within the frame work of non parametric estimation have been elaborated. Among the most prominent ones, we mention Parzen (1962), Samanta (1973) and Tsybakov (1990). Recently, there has been a spate of interest in recursive estimation which has drawn the attention of multiple researchers. The basic merit of the recursive estimator lies in the fact that it can not only be updated with each additional new observation especially in large sample sizes but it can also be much better in terms of computational costs. In this work, our central focus is upon a recursive kernel estimator of the mode function defined by stochastic approximation method.

Let $X_1, \cdots, X_n$ be identically distributed $\mathbb{R}^d$-valued random vectors satisfy the $\alpha$-mixing dependency property (see Definition 1.2.1) and let $f$ denote the probability density of $X_i$, $i = 1, \cdots, n$. We consider a compact set $\Omega$ such that $\Omega \subset \mathbb{R}^d$, and we define the mode as follows

$$\theta := \arg\max_{y \in \Omega} f(y).$$

We assume that $\theta$ is unique.

In order to define our estimator of the mode, we first begin by constructing a stochastic algorithm for the estimation of the function $f$ at a point $x$. We present an algorithm to search for the zero of the function $g : y \longmapsto f(x) - y$. Following Robbins-Monro's procedure, this algorithm is defined below as

(i) $f_0(x)$ is an arbitrary choice belonging to $\mathbb{R}$,

(ii) $\forall n \geqslant 1$, we set $f_n(x) = f_{n-1}(x) + \gamma_n W_n(x)$, where the stepsize $(\gamma_n)_n$ is a sequence of positive real numbers that goes to zero and $W_n(x)$ is an observation of the function $g$ at the point $f_{n-1}(x)$.

To construct $W_n(x)$, we follow the approach of Révész (1973, 1977) and Tsybakov (1990) which are based on the classical property of stochastic algorithms ($\mathbb{E}\left[W_n(x)|\mathcal{F}_1^{n-1}\right] = 0$, where $\mathcal{F}_1^{n-1}$ stands for the $\sigma$-field of events generated by $\{X_1, \cdots, X_{n-1}\}$). In addition, we introduce a kernel $K$ (which is a function satisfying $\int_{\mathbb{R}^d} K(z)dz = 1$), and a bandwidth $(h_n)$ (which is a sequence of positive real numbers that goes to zero when $n \longrightarrow \infty$), and we set $W_n(x) = K_{h_n}(x - X_n) - f_{n-1}(x)$, with $K_h(x) := h^{-d} K\left(h^{-1}x\right)$. Therefore, the recursive estimator of the density function $f$ at the point $x$ can be written as

$$f_n(x) = \pi_n f_0 + \pi_n \sum_{k=1}^{n} \pi_k^{-1} \gamma_k h_k^{-d} K\left(\frac{x - X_k}{h_k}\right)$$

with $\pi_n = \prod_{k=1}^{n} (1 - \gamma_k)$. Our estimator of mode $\theta$ is defined as the random variable $\theta_n$ maximizing the recursive estimator $f_n$ of $f$, which is expressed as

$$\theta_n := \arg\max_{t \in \Omega} f_n(t). \tag{4.1.1}$$

The mode estimator has been investigated by several authors. Based on independent and identically distributed (iid) random data, the weak consistency and the asymptotic normality

of the kernel sample mode was addressed by Parzen (1962). This result was extended in several directions by Chernoff (1964), Eddy (1980, 1982) and Vieu (1996). Strong consistency was explored by Nadaraya (1965) and Van Ryzin (1969). Asymptotic normality of kernel estimate of the mode was elaborated by Romano (1988). The multidimensional study of the mode was carried out by Samanta (1973) and Konakov (1974).
Based on dependent random data, some studies have been performed for mode estimation. In $\phi$-mixing condition as well as the conditional case, the strong consistency was enacted by Collombs et al (1987). In alpha mixing case, the strong consistency was established by Ould Saïd (1993) and the asymptotic normality was set forward by Louani and Ould Saïd (1999). Numerous works were conducted, under censored and truncated data, to explore the property of non parametric mode estimators (see Louani (1998), Ould Saïd (2005) and Gannoun and Louani (1996)).

The majority of properties of mode estimators are related to those of density estimators. We need always to handle the density case before that of the mode. This chapter investigates the estimation of the mode, which is based on non parametric recursive kernel density estimator developed by Mokkadem et al. (2009a), under strong mixing conditions. The rest of the chapter is organized as follows. In Section 4.2, the assumptions and main results are displayed. Section 4.3 is devoted to simulation study. Finally, a conclusion is presented in Section 4.4. The details of proofs are exhibited in Section 4.5 along with some auxiliary results.

## 4.2 Assumptions and main results

We consider stepsizes and bandwidths, which belong to the following class of regularly varying sequences.

**Definition 4.2.1.** *Let $\gamma \in \mathbb{R}$ and $(\gamma_n)_{n\geqslant 1}$ be a nonrandom positive sequence. We state that $\gamma_n \in \mathcal{GS}(\gamma)$ if $\lim_{n\to\infty} n[1-\frac{\gamma_{n-1}}{\gamma_n}] = \gamma$.*

The assumptions to which we shall refer are the following:

**(A1)** The kernel function K:$\mathbb{R}^d \longrightarrow \mathbb{R}$ is a bounded probability density, lipschitz and satisfies for all $j \in \{1,\dots,d\}$, $\int_{\mathbb{R}} z_j K(z)dz_j = 0$ and $\int_{\mathbb{R}^d} z_j^2 K(z)dz < \infty$.

**(A2)**

(i) $\gamma_n \in \mathcal{GS}(-\alpha)$ with $\alpha \in ]1/2, 1]$.

(ii) $h_n \in \mathcal{GS}(-a)$ with $a \in ]0, \alpha/d[$.

(iii) $\lim_{n\to\infty} n\gamma_n \in ]\min\{2a, (1-ad)/2\}, \infty]$.

**(A3)** $f$ is bounded, twice differentiable on $\Omega$, and, for all $i, j \in \{1, \cdots, d\}$, $\partial^2 f/\partial x_i \partial x_j$ is bounded.

**(A4)** The joint density $f_{(i,j)}$ of $(X_i, X_j)$ exists for all $(i,j)$, and there exists a constant $M > 0$ such that

$$\sup_{|i-j|\geqslant 1}\sup_{t_1,t_2\in\Omega}\left|f_{(i,j)}(t_1,t_2)-f(t_1)f(t_2)\right| < M.$$

**(A5)** The mixing coefficient of the $X_i$'s satisfies $\alpha(n) = O(n^{-\nu})$ for some $\nu \geqslant 3$.
**(A6)** The mode $\theta$ satisfies the following property: for any $\varepsilon > 0$ and $x$, there exists $\eta \neq 0$ such that $|\theta - x| > \varepsilon$ implies that $|f(\theta) - f(x)| > \eta$.
**(A7)**

(i) $n^{1/\nu}\gamma_n^{1-1/\nu} \underset{n\longrightarrow\infty}{\longrightarrow} 0.$

(ii)

$$\begin{cases} a(d\nu-2)-\alpha(d+2) > 6 & \text{if } a \geq \alpha/(d+4) \\ a(d-2\nu-6)-\alpha > 6 & \text{if } a < \alpha/(d+4). \end{cases}$$

**Remark 4.2.1.** Assumption **(A1)** on the kernel is widely used in the recursive and non recursive framework for the functional estimation. Assumptions **(A2)** on the stepsize and the bandwidth are used in the recursive framework for the estimation of the density function (Mokkadem et al. (2009a); Slaoui (2013, 2014a, 2018b)). Hypothesis **(A2)**(i) and **(A2)**(ii) ensure that the bandwidth $(h_n)$ and the stepsize $(\gamma_n)$ go to zero as $n$ goes to infinity. Moreover, the stepsize $(\gamma_n)$ goes to zero more rapidly than the bandwidth $(h_n)$. Assumption **(A2)**(iii) on the limit as $n$ goes to infinity of $(n\gamma_n)$ is usual in the framework of stochastic approximation algorithms. It implies that the limit of $(n\gamma_n)^{-1}$ is finite. Assumption **(A3)** on the function $f$ allows us to calculate the properties of our estimator. Condition **(A4)** is needed to calculate the covariance. **(A5)** states a condition on the mixing coefficient. Assumption **(A6)** is classical in mode estimation. Finally, hypothesis **(A7)** provides a condition for the bandwidth allowing the estimation of the covariance term.

Throughout this chapter, we shall use the following notation:

$$\begin{aligned}
\varepsilon &= \lim_{n\to\infty}(n\gamma_n)^{-1}, && (4.2.1)\\
\mu_j^2 &= \int_{\mathbb{R}^d} z_j^2 K(z)dz, \ \ \forall j \in \{1,\cdots,d\}, && (4.2.2)\\
f_{ij}^{(2)}(x) &= \frac{\partial^2 f(x)}{\partial x_i \partial x_j},\\
Z_n(x) &= \frac{1}{h_n^d}K\left(\frac{x-X_n}{h_n}\right).
\end{aligned}$$

The almost sure convergence is denoted by $a.s.$.

Now, we shall prove the consistency of our estimator (4.1.1) and give the rate of convergence.

**Proposition 4.2.1.** *Let Assumptions **(A1)**-**(A7)** hold.*

$$\sup_{t\in\Omega}|f_n(t)-f(t)| = \begin{cases} O\left(\sqrt{\gamma_n h_n^{-d}\log n}\right) & \text{if } a > \alpha/(d+4) \\ O\left(\max\left(\sqrt{\gamma_n h_n^{-d}\log n}, h_n^2\right)\right) & \text{if } a = \alpha/(d+4) \\ O\left(h_n^2\sqrt{\log n}\right) & \text{if } a < \alpha/(d+4) \end{cases}$$

*a.s. as* $n \longrightarrow \infty$.

**Proposition 4.2.2.** *Under the assumption of Proposition* 4.2.1*, we have*

$$\theta_n - \theta = \begin{cases} O\left(\left(\gamma_n h_n^{-d}\log n\right)^{1/4}\right) & \text{if } a > \alpha/(d+4)) \\ O\left(\max\left(\left(\gamma_n h_n^{-d}\log n\right)^{1/4}, h_n\right)\right) & \text{if } a = \alpha/(d+4) \\ O\left(h_n\left(\log n\right)^{1/4}\right) & \text{if } a < \alpha/(d+4) \end{cases}$$

*a.s. as* $n \longrightarrow \infty$.

## 4.3 Simulation study

In this section, we aim to compare our proposed recursive kernel estimator of mode, defined by (4.1.1), with the mode estimator based on the well-known non recursive kernel density estimator introduced by Rosenblatt (1956b),

$$\tilde{\theta}_n := \arg\max_{t\in\Omega}\tilde{f}_n(t), \tag{4.3.1}$$

where $\tilde{f}_n(t) = \frac{1}{nh_n^d}\sum_{k=1}^{n} K\left(\frac{t-X_k}{h_n}\right)$.

### 4.3.1 The study design

Let us consider the following simulation design, we simulate $N = 500$ samples of sizes, $n = 50$, $n = 100$, $n = 150$ and a sequence of m-dependent variables

$$X_i = \sum_{j}^{i+m}\sqrt{|Y_j|},$$

where $(Y_j)_j$ are generated from the following mixture distributions:

- $Y \sim \frac{1}{2}\mathcal{N}(2.5, 6) + \frac{1}{2}\mathcal{N}(9, 1)$.
- $Y \sim \frac{1}{2}\mathcal{N}(2, 6) + \frac{1}{2}\mathcal{N}(8, 1)$.

Next, we calculate the $\overline{ISE}$ (Integrated Squared Error) and the $\overline{IAE}$ (Integrated Absolute Error) of the two estimators;

$$\overline{ISE} = \frac{1}{N}\sum_{i=1}^{N}\left(\theta_n^{[i]} - \theta\right)^2 \quad \text{and} \quad \overline{IAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\theta_n^{[i]} - \theta\right|,$$

where $\theta_n^{[i]}$ corresponds to the mode estimator computed from the ith sample. In order to calculate the $\overline{ISE}$ and the $\overline{IAE}$ of the two mode estimators, we need to use the following quantities:

- The normal kernel function $K$.
- The stepsize $(\gamma_n)_n = (n^{-1})_n$.
- The bandwidth $(h_n)_n$ is chosen with plug-in method, given in Slaoui (2014a).

### 4.3.2 Results

For each configuration of the simulation design parameters, we calculate the $\overline{ISE}$ and the $\overline{IAE}$ of the non recursive estimator (4.3.1) and the recursive estimator (4.1.1). From Table 1, Table 2, Table 3 and Table 4, it is clear that, the proposed recursive estimator (4.1.1) outperformed the non recursive estimator (4.3.1) in all the considered situations. We can observe that the $\overline{ISE}$ decreases as $m$ increases. We can observe also that the $\overline{ISE}$ decreases as the sample size $n$ increases. This simulation study shows the good performance of the recursive estimator with an appropriate choice of stepsize and bandwidth parameters.

## 4.4 Conclusion

In this chapter, we attempted to elaborate a recursive kernel mode estimator based on stochastic approximation algorithm. We established the strong consistency of this estimator under $\alpha$-mixing condition. Investing the same selected parameters in Mokkadem et al. (2009a), which minimize the mean squared error of recursive density estimator, the proposed recursive mode estimator maintains the same convergence rate with non recursive mode estimator defined by (4.3.1). The two previous estimators are asymptotically equivalent. In addition, the main merit of our estimator resides in its update, when a new sample information becomes available. Tackling this area is extremely interesting as it offers new perspectives for future works to consider multiple directions within this framework. This involves the elaboration of recursive mode estimation for dependent strong mixing functional data like in Slaoui (2020). Furthermore, our proposed

| | | $Y \sim \frac{1}{2}\mathcal{N}(2.5, 6) + \frac{1}{2}\mathcal{N}(9, 1)$ | | |
|---|---|---|---|---|
| | | $n = 50$ | $n = 100$ | $n = 150$ |
| | | $\overline{ISE}$ | $\overline{ISE}$ | $\overline{ISE}$ |
| $m = 2$ | | | | |
| | Non recursive | 0.315226 | 0.148972 | 0.103131 |
| | Recursive | **0.158550** | **0.029883** | **0.021263** |
| $m = 4$ | | | | |
| | Non recursive | 0.194686 | 0.133191 | 0.100986 |
| | Recursive | **0.113563** | **0.125054** | **0.020975** |

Table 4.1: $\overline{ISE}$ for $N = 500$ trials of the non recursive estimator (4.3.1) and the recursive estimator (4.1.1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{ISE}$.

| | | $Y \sim \frac{1}{2}\mathcal{N}(2.5, 6) + \frac{1}{2}\mathcal{N}(9, 1)$ | | |
|---|---|---|---|---|
| | | $n = 50$ | $n = 100$ | $n = 150$ |
| | | $\overline{IAE}$ | $\overline{IAE}$ | $\overline{IAE}$ |
| $m = 2$ | | | | |
| | Non recursive | 0.561450 | 0.385970 | 0.321140 |
| | Recursive | **0.398184** | **0.172867** | **0.145820** |
| $m = 4$ | | | | |
| | Non recursive | 0.441232 | 0.275033 | 0.115408 |
| | Recursive | **0.336991** | **0.155993** | **0.111827** |

Table 4.2: $\overline{IAE}$ for $N = 500$ trials of the non recursive estimator (4.3.1) and the recursive estimator (4.1.1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{IAE}$.

| | | $Y \sim \frac{1}{2}\mathcal{N}(2,6) + \frac{1}{2}\mathcal{N}(8,1)$ | | |
|---|---|---|---|---|
| | | $n = 50$ | $n = 100$ | $n = 150$ |
| | | $\overline{ISE}$ | $\overline{ISE}$ | $\overline{ISE}$ |
| $m = 3$ | | | | |
| | Non recursive | 0.337509 | 0.136912 | 0.079324 |
| | Recursive | **0.305999** | **0.115895** | **0.058715** |
| $m = 5$ | | | | |
| | Non recursive | 0.154778 | 0.142782 | 0.071498 |
| | Recursive | **0.303375** | **0.99815** | **0.045986** |

Table 4.3: $\overline{ISE}$ for $N = 500$ trials of the non recursive estimator (4.3.1) and the recursive estimator (4.1.1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{ISE}$.

| | | $Y \sim \frac{1}{2}\mathcal{N}(2,6) + \frac{1}{2}\mathcal{N}(8,1)$ | | |
|---|---|---|---|---|
| | | $n = 50$ | $n = 100$ | $n = 150$ |
| | | $\overline{IAE}$ | $\overline{IAE}$ | $\overline{IAE}$ |
| $m = 3$ | | | | |
| | Non recursive | 0.580955 | 0.281645 | 0.192126 |
| | Recursive | **0.553173** | **0.242312** | **0.126076** |
| $m = 5$ | | | | |
| | Non recursive | 0.393418 | 0.224771 | 0.99556 |
| | Recursive | **0.550795** | **0.224771** | **0.106438** |

Table 4.4: $\overline{IAE}$ for $N = 500$ trials of the non recursive estimator (4.3.1) and the recursive estimator (4.1.1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{IAE}$.

recursive kernel mode estimator is promising and can be extended in such a way as addressing recursive non parametric estimation in the Bayesian work (see Boukabour and Masmoudi (2020)).

## 4.5 Proofs

Before setting the outlines of the proofs, we introduce the following technical lemma, which is proved in Mokkadem et al. (2009a), and which will be used throughout the demonstrations.

**Lemma 4.5.1.**
*Let $v_n \in \mathcal{GS}(v^*)$, $\gamma_n \in \mathcal{GS}(-\alpha)$ and $m > 0$ such that $m - v^*\varepsilon > 0$ where $\varepsilon$ is defined in* (4.2.1)*. Then,*

$$\lim_{n\to\infty} v_n \pi_n^m \sum_{k=1}^{n} \pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\varepsilon}.$$

*Moreover, for all positive sequence $(\alpha_n)_n$ such that $\lim_{n\to\infty} \alpha_n = 0$, and all $C \in \mathbb{R}$,*

$$\lim_{n\to\infty} v_n \pi_n^m \left[\sum_{k=1}^{n} \pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + C\right] = 0.$$

Proof of Proposition 4.2.1. The proof rests on the following decomposition

$$|f_n(t) - f(t)| \quad \leq \quad |f_n(t) - \mathbb{E}[f_n(t)]| + |\mathbb{E}[f_n(t)] - f(t)|$$

and is based on the proofs of the following three lemmas.

**Lemma 4.5.2.** *Under Assumptions **(A1)**-**(A3)**, we have*

$$\sup_{t\in\Omega} |\mathbb{E}[f_n(t)] - f(t)| = \begin{cases} O\left(h_n^2\right) & \text{if } a \leq \alpha/(d+4) \\ o\left(\sqrt{\gamma_n h_n^{-d}}\right) & \text{if } a > \alpha/(d+4) \end{cases}$$

*as $n \longrightarrow \infty$.*

The proof of Lemma 4.5.2 is presented in Mokkadem et al. (2009a).

**Lemma 4.5.3.** *(Fuk-Nagaev) Let $(W_i)_{i\in\mathbb{N}}$ be a sequence of centered real random variables, with a strong mixing coefficient $\alpha(n) = O(n^{-\nu})$, $\nu > 1$, such that $\forall n \in \mathbb{N}$, $1 \leqslant i \leqslant n$, $|W_i| < +\infty$. Hence, for all $\varepsilon > 0$ and $r > 1$, there exists a constant $c$ such that*

$$\mathbb{P}\left\{ |\sum_{k=1}^{n} W_i| > \varepsilon \right\} \leqslant c\left(1 + \frac{\varepsilon^2}{16 r S_n^2}\right)^{-r/2} + ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{\nu+1}$$

*where $S_n^2 = \sum_{i,j=1}^{n} |\, \mathbb{C}ov(W_i, W_j)\,|$.*

For more details about previous Lemma 4.5.3, we refer to Rio (2000), p. 87, 6.19b.

**Lemma 4.5.4.** *Under Assumptions **(A1)**-**(A7)**, we have*

$$\sup_{t\in\Omega}|f_n(t)-\mathbb{E}[f_n(t)]| = \begin{cases} O\left(\sqrt{\gamma_n h_n^{-d}\log n}\right) & \text{if } a \geq \alpha/(d+4) \quad (4.5.2a) \\ O\left(h_n^2\sqrt{\log n}\right) & \text{if } a < \alpha/(d+4) \quad (4.5.2b) \end{cases}$$

*a.s. as* $n \longrightarrow \infty$.

Proof of Lemma 4.5.4. The proof relies upon the following assertion: the compact set $\Omega$ can be covered by a finite number $\lambda_n$ of balls $\mathcal{B}_k(t_k^*, b_n)$ centered at $t_k^*$, $1 \leqslant k \leqslant \lambda_n$ where $b_n$ satisfies

$$b_n = \gamma_n^{1/2} h_n^{1+d/2}. \tag{4.5.3}$$

Since $\Omega$ is bounded, one can find $l > 0$ such that $\lambda_n \leq l b_n^{-1}$. For any $t \in \Omega$, there exists $k$ such that

$$|t - t_k^*| \leqslant b_n. \tag{4.5.4}$$

Now, we set for $t \in \Omega$

$$T_i(t) = \pi_i^{-1}\gamma_i h_i^{-d}\left\{K\left(\frac{t-X_i}{h_i}\right) - \mathbb{E}\left(K\left(\frac{t-X_i}{h_i}\right)\right)\right\}. \tag{4.5.5}$$

Evidently, we get

$$\begin{aligned} \pi_n\sum_{i=1}^n T_i(t) &= f_n(t) - \mathbb{E}\left(f_n(t)\right) \\ &= \{(f_n(t) - f_n(t_k^*)) - (\mathbb{E}(f_n(t) - \mathbb{E}(f_n(t_k^*))\} + \{f_n(t_k^*) - \mathbb{E}(f_n(t_k^*)\} \\ &:= \pi_n\sum_{i=1}^n \tilde{T}_i(t) + \pi_n\sum_{i=1}^n T_i(t_k^*) \end{aligned}$$

with

$$\begin{aligned} \tilde{T}_i(t) &= \pi_i^{-1}\gamma_i h_i^{-d}\left\{K\left(\frac{t-X_i}{h_i}\right) - K\left(\frac{t_k^*-X_i}{h_i}\right)\right\} \\ &\quad -\pi_i^{-1}\gamma_i h_i^{-d}\left\{\mathbb{E}\left(K\left(\frac{t-X_i}{h_i}\right)\right) - \mathbb{E}\left(K\left(\frac{t_k^*-X_i}{h_i}\right)\right)\right\}. \end{aligned}$$

As a matter of fact, we have

$$\sup_{t\in\Omega}\left|\pi_n\sum_{i=1}^n T_i(t)\right| \leqslant \max_{k\leqslant\lambda_n}\sup_{t\in\mathcal{B}_k}\left|\pi_n\sum_{i=1}^n \tilde{T}_i(t)\right| + \max_{k\leqslant\lambda_n}\left|\pi_n\sum_{i=1}^n T_i(t_k^*)\right|$$

$$:= \quad U_1 + U_2.$$

In order to investigate $U_1$, we observe that

$$\begin{aligned}
\left|\pi_n \sum_{i=1}^{n} \tilde{T}_i(t)\right| &\leqslant \pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-d} \left| K\left(\frac{t-X_i}{h_i}\right) - K\left(\frac{t_k^*-X_i}{h_i}\right)\right| \\
&\quad + \pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-d} \mathbb{E}\left[\left| K\left(\frac{t-X_i}{h_i}\right) - K\left(\frac{t_k^*-X_i}{h_i}\right)\right|\right] \\
&:= V_1(t) + V_2(t).
\end{aligned}$$

Assumptions **(A1)**, (4.5.3) and (4.5.4) and the application of Lemma 4.5.1 provide

$$\begin{aligned}
V_1(t) &\leqslant c\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-d} \left|\frac{t-t_k^*}{h_i}\right| \\
&\leqslant c\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-(d+1)} |t-t_k^*| \\
&\leqslant c b_n h_n^{-(d+1)} \frac{1}{1+a(d+1)\varepsilon} \\
&\leqslant c \gamma_n^{1/2} h_n^{-d/2} \frac{1}{1+a(d+1)\varepsilon} \\
&= O\left(\sqrt{\gamma_n h_n^{-d}}\right),
\end{aligned}$$

and

$$\begin{aligned}
V_2(t) &\leqslant c\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-(d+1)} \mathbb{E}\left[|t-t_k^*|\right] \\
&= O\left(\sqrt{\gamma_n h_n^{-d}}\right).
\end{aligned}$$

Thus, we get

$$U_1 = O\left(\sqrt{\gamma_n h_n^{-d}}\right) a.s \quad \text{as} \quad n \longrightarrow \infty.$$

Now, in order to study $U_2$, we use Lemma 4.5.3. For that, let

$$W_i = \pi_n T_i(t_k^*) = \pi_n \pi_i^{-1}\gamma_i h_i^{-d} \left\{ K\left(\frac{t_k^*-X_i}{h_n}\right) - \mathbb{E}\left(K\left(\frac{t_k^*-X_i}{h_n}\right)\right)\right\}. \tag{4.5.6}$$

Then, we have to calculate

$$\begin{aligned}
S_n^2 &= \sum_{i,j=1}^{n} |\, \mathbb{C}ov\left(W_i, W_j\right) | \\
&= \sum_{i \neq j} |\mathbb{C}ov\left(W_i, W_j\right)| + \sum_{i=1}^{n} \mathbb{V}ar\left(W_i\right)
\end{aligned}$$

$$:= \quad S_n^{2*} + \sum_{i=1}^{n} \mathbb{V}ar\left(W_i\right).$$

On the one hand, under **(A1)**-**(A3)**, we obtain

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{V}ar\left(W_i\right) &= \pi_n^2 \sum_{i=1}^{n} \pi_i^{-2}\gamma_i^2 \mathbb{V}ar\left(Z_i(t_k^*)\right) \\
&= \begin{cases} O\left(\gamma_n h_n^{-d}\right) & \text{if } a \geq \alpha/(d+4) \\ o\left(h_n^4\right) & \text{if } a < \alpha/(d+4), \end{cases}
\end{aligned}$$

see Proposition 1 in Mokkadem et al. (2009a) for more details about computation of the variance. Now, from (4.5.6) as well as under assumptions **(A1)** and **(A4)**, we have

$$\begin{aligned}
\left|\mathbb{C}ov\left(W_i, W_j\right)\right| &= \left|\mathbb{E}\left[\pi_n^2\pi_i^{-1}\pi_j^{-1}\gamma_i\gamma_j h_i^{-d}h_j^{-d} K\left(\frac{t_k^*-X_i}{h_n}\right) K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\right. \\
&\quad \left. - \mathbb{E}\left[\pi_n\pi_i^{-1}\gamma_i h_i^{-d} K\left(\frac{t_k^*-X_i}{h_n}\right)\right] \mathbb{E}\left[\pi_n\pi_j^{-1}\gamma_j h_j^{-d} K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\right| \\
&= \left|\pi_n^2\pi_i^{-1}\pi_j^{-1}\gamma_i\gamma_j h_i^{-d}h_j^{-d} \left(\mathbb{E}\left[K\left(\frac{t_k^*-X_i}{h_n}\right) K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\right.\right. \\
&\quad \left.\left. - \mathbb{E}\left[K\left(\frac{t_k^*-X_i}{h_n}\right)\right] \mathbb{E}\left[K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\right)\right| \\
&= \pi_n^2\pi_i^{-1}\pi_j^{-1}\gamma_i\gamma_j \int_{\mathbb{R}^{2d}} K(t_1)K(t_2) \left| f_{(i,j)}(t_k^*-t_1h_i, t_k^*-t_2h_j) \right. \\
&\qquad \left. - f(t_k^*-t_1h_i) f(t_k^*-t_2h_j) \right| dt_1 dt_2 \\
&\leqslant M\pi_n^2\pi_i^{-1}\gamma_i\pi_j^{1}\gamma_j \\
&= O(\pi_n^2\pi_i^{-1}\gamma_i\pi_j^{1}\gamma_j). \qquad (4.5.7)
\end{aligned}$$

Next, to asses the term $S_n^{2*}$, we use a technique developed by Masry (1986). We define the sets

$$F_1 = \{(i,j) : 1 \leqslant |i-j| \leqslant \beta_n\}$$

and

$$F_2 = \{(i,j) : \beta_n + 1 \leqslant |i-j| \leqslant n-1\}$$

where $\beta_n = o(n)$. Let

$$\mathcal{F}_{1,n} = \sum_{i,j\in F_1} \left|\mathbb{C}ov\left(W_i, W_j\right)\right| \quad and \quad \mathcal{F}_{2,n} = \sum_{i,j\in F_2} \left|\mathbb{C}ov\left(W_i, W_j\right)\right|.$$

Applying the upper bound in (4.5.7), we have

$$\mathcal{F}_{1,n} \quad \leqslant \quad M\pi_n^2 \sum_{i,j\in F_1} \pi_i^{-1}\gamma_i\pi_j^{-1}\gamma_j$$

$$
\begin{aligned}
&\leqslant M\pi_n^2 \sum_{j=1}^{n}\sum_{k=1}^{\beta_n} \pi_{k+j}^{-1}\gamma_{k+j}\pi_j^{-1}\gamma_j \\
&\leqslant M\pi_n^2 \sum_{j=1}^{n}\sum_{k=1}^{\beta_n} \pi_j^{-2}\gamma_j^2 \frac{1}{(1-\gamma_{j+1})\cdots(1-\gamma_{j+k})} \\
&\leqslant M\beta_n\pi_n^2 \sum_{j=1}^{n} \pi_j^{-2}\gamma_j^2,
\end{aligned}
$$

and applying Lemma 4.5.1, we get

$$
\begin{aligned}
\mathcal{F}_{1,n} &\leqslant M\beta_n\gamma_n \frac{1}{2-\alpha\varepsilon} \\
&= O\left(\beta_n\gamma_n\right).
\end{aligned}
$$

For $F_2$, we use the Davydov inequality for mixing processes (see Rio 2000, p. 10, Formula 1.12a). This leads us to get, for all $i \neq j$

$$
\left|\mathbb{C}ov\left(W_i, W_j\right)\right| \leqslant c\alpha\left(|i-j|\right).
$$

Therefore, using **(A5)**, we obtain

$$
\begin{aligned}
\mathcal{F}_{2,n} &\leqslant c\sum_{j=1}^{n} \sum_{\beta_n+1\leqslant k\leqslant n-1} \alpha\left(k\right) \\
&< cn\int_{\beta_n+1}^{n-1} k^{-v}dk \\
&= O\left(n\beta_n^{1-v}\right).
\end{aligned}
$$

Choosing $\beta_n = \left(n\gamma_n^{-1}\right)^{1/\nu}$ and under **(A7)**(i), we obtain

$$
S_n^{2*} = \mathcal{F}_{1,n} + \mathcal{F}_{2,n} = O\left(n^{1/\nu}\gamma_n^{1-1/\nu}\right) = o\left(1\right).
$$

Finally, we get

$$
S_n^2 = \begin{cases} O\left(\gamma_n h_n^{-d}\right) & \text{if } a \geq \alpha/(d+4) \quad (4.5.8a) \\ o\left(h_n^4\right) & \text{if } a < \alpha/(d+4). \quad (4.5.8b) \end{cases}
$$

As a matter of fact, we apply Lemma 4.5.3 in the case $a \geq \alpha/(d+4)$. We obtain, for any $k$

$$
\begin{aligned}
\mathbb{P}\left\{\left|\pi_n\sum_{k=1}^{n} T_i(t_k^*)\right| > \varepsilon\right\} &\leqslant c\left(1+\frac{\varepsilon^2}{16rS_n^2}\right)^{-r/2} + ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{\nu+1} \\
&:= c\left(\Gamma_{1,n}+\Gamma_{2,n}\right).
\end{aligned}
$$

By taking

$$\varepsilon = \varepsilon_0 \left( \sqrt{\gamma_n h_n^{-d} \log n} \right) \quad and \quad r = c \log n (\log_2 n)^{1/\nu} \tag{4.5.9}$$

and using Taylor series expansion of $\log(1+x)$ as well as (4.5.8a)-(4.5.9), we infer

$$\Gamma_{1,n} \leqslant c n^{-\varepsilon_0^2/2}$$

and

$$\Gamma_{2,n} \leqslant c \varepsilon_0^{-(\nu+1)} n \gamma_n^{-(\nu+1)/2} h_n^{d(\nu+1)/2} (\log n)^{(\nu-1)/2} \log_2 n$$

where $\log_2 n = \log(\log n)$ for $n > 2$. Consequently,

$$\begin{aligned}
& \mathbb{P} \quad \left\{ \max_{k=1,\cdots,\lambda_n} \left| \pi_n \sum_{k=1}^{n} T_i(t_k^*) \right| > \varepsilon_0 \left( \sqrt{\gamma_n h_n^{-d} \log n} \right) \right\} \\
\leqslant & \quad \sum_{i=1}^{\lambda_n} \mathbb{P} \left\{ \left| \pi_n \sum_{k=1}^{n} T_i(t_k^*) \right| > \varepsilon_0 \left( \sqrt{\gamma_n h_n^{-d} \log n} \right) \right\} \\
\leqslant & \quad \lambda_n c \left\{ \Gamma_{1,n} + \Gamma_{2,n} \right\} \\
\leqslant & \quad l b_n^{-1} c \left\{ \Gamma_{1,n} + \Gamma_{2,n} \right\} \\
\leqslant & \quad lc \left\{ n^{(\alpha-\varepsilon_0^2)/2} h_n^{-(2+d)/2} + \varepsilon_0^{-(\nu+1)} n^{\alpha(\nu+2)/2+1} h_n^{(d\nu-2)/2} (\log n)^{(\nu-1)/2} \log_2 n \right\} \\
:= & \quad lc \left\{ \tilde{\Gamma}_{1,n} + \tilde{\Gamma}_{2,n} \right\},
\end{aligned}$$

with

$$\tilde{\Gamma}_{1,n} := b_n^{-1} \Gamma_{1,n} \quad and \quad \tilde{\Gamma}_{2,n} := b_n^{-1} \Gamma_{2,n}.$$

Now, referring to **(A7)**(ii), we have

$$h_n^{(d\nu-2)/2} = o\left( n^{-\alpha(\nu+2)/2-2} (\log n)^{-(\nu+1)/2} (\log_2 n)^{-3} \right),$$

which yields

$$\tilde{\Gamma}_{2,n} = o\left( \frac{1}{n \log n (\log_2 n)^2} \right),$$

corresponding to the general term of the convergent Bertrand series. For $\tilde{\Gamma}_{1,n}$, an appropriate choice of $\varepsilon_0$ can be made $O\left(n^{-3/2}\right)$, which corresponds to the general term of convergent series. Hence, $\sum_{n \geqslant 1} \left\{ \tilde{\Gamma}_{1,n} + \tilde{\Gamma}_{2,n} \right\} < \infty$, and therefore (4.5.2a) follows by applying Borel Cantelli Lemma. The same steps shall be used in the second case if $a < \alpha/(d+4)$. The result (4.5.2b) is a consequence of Borel Cantelli Lemma after applying Lemma 4.5.3 and choosing

$$\varepsilon = \varepsilon_0 h_n^2 \sqrt{\log n} \quad and \quad r = c \log n (\log_2 n)^{1/\nu}.$$

Proof of Proposition 4.2.2. Standard argument yields

$$|f(\theta_n) - f(\theta)| \quad \leqslant \quad |f(\theta_n) - f_n(\theta_n)| + |f_n(\theta_n) - f(\theta)|$$

$$\leqslant \sup_{t\in\Omega} |f_n(t) - f(t)| + |f_n(\theta_n) - f(\theta)| \,. \tag{4.5.10}$$

Since

$$|f_n(\theta_n) - f(\theta)| = \left| \sup_{t\in\Omega} f_n(t) - \sup_{t\in\Omega} f(t) \right| \leqslant \sup_{t\in\Omega} |f_n(t) - f(t)| \,,$$

then we have

$$|f(\theta_n) - f(\theta)| \leqslant 2 \sup_{t\in\Omega} |f_n(t) - f(t)| \,. \tag{4.5.11}$$

The a.s. consistency of $\theta_n$ follows then immediately from (4.2.1) and **(A6)**. Now a taylor expansion provides

$$\begin{aligned} f(\theta_n) - f(\theta) &= (\theta_n - \theta) f'(\theta) + \frac{1}{2}(\theta_n - \theta)^2 f^{(2)}(\theta_n^*) \\ &= \frac{1}{2}(\theta_n - \theta)^2 f^{(2)}(\theta_n^*), \end{aligned}$$

where $\theta_n^*$ is between $\theta$ and $\theta_n$. Therefore, based on (4.5.11) and **(A3)**, we get

$$\begin{aligned} |\theta_n - \theta| &\leqslant \sqrt{\frac{2\,|f(\theta_n) - f(\theta)|}{\left|f^{(2)}(\theta_n^*)\right|}} \\ &\leqslant 2\sqrt{\frac{\sup_{t\in\Omega} |f_n(t) - f(t)|}{|\ f^{(2)}(\theta_n^*)\ |}}. \end{aligned}$$

Thus, by (4.2.1) the proof holds.

# Chapter 5

# Conclusion and perspective

For clarity and methodological reasons, the basic concepts and properties used in the subsequent analysis are highlighted in chapter 1.

In chapter 2, we tackled the estimation of the conditional extreme value index $\gamma(x)$ of a heavy-tailed distribution when some random covariate information is available. We elaborated recursive kernel estimator of the extreme value index function based on the stochastic approximation algorithm. We demonstrated that using some particular stepsizes and a specific bandwidth selection through a cross-validation procedure, our recursive estimator could be very competitive to Hill's non recursive version in terms of estimation error and computational costs. We illustrated this performance via simulations and real data.

In chapter 3, we extended the work of Slaoui (2014b) to the case of $\alpha$-mixing data. We established the central limit theorem and the uniform convergence for the proposed estimator under some mild conditions. We confirmed that using a specific plug-in bandwidth selection method and some particularly stepsizes, the proposed recursive estimator yielded better results compared to Nadaraya's non recursive distribution estimator under $\alpha$-mixing condition. However, the basic merit of the recursive method resides essentially in the fact that it is much faster than the classical one. Eventually, these theoretical results were corroborated through a few simulations.

In chapter 4, we elaborated a recursive kernel mode estimator based on stochastic approximation algorithm. We established the strong consistency of this estimator under $\alpha$-mixing condition. Investing the same selected parameters in Mokkadem et al. (2009a), which minimize the mean squared error of recursive density estimator, the proposed recursive mode estimator maintains the same convergence rate with non recursive mode estimator defined by (4.3.1). We shown that two previous estimators are asymptotically equivalent. In addition, the main advantage of our estimator resides in its update, when a new sample information becomes available.

At this stage of synthesis, it is noteworthy that our thesis would be valuable in terms of opening further fruitful lines of investigation and offering promising future perspectives. Indeed, this thesis may be extended in several ways:

- First, we may extend our recursive extreme value index estimator to the case of censored data. We can also propose a new estimator of the conditional extreme quantile using our recursive estimator defined by (2.2.7) and compare it to the classical Weissman estimator.

- Second, we may equally apply Bernstein and Lagrange polynomials to estimate extreme value index and extreme quantile functions (See Slaoui and Jmaei (2019) and Helali and Slaoui (2020)).

- Third we may explore of a recursive mode estimation for dependent strong mixing functional data like in Slaoui (2019, 2020) and for dependent strong mixing spatial data like in Bouzebda and Slaoui (2018, 2020). Furthermore, our proposed recursive kernel mode estimator is promising and can be extended in such a way as addressing recursive non parametric estimation in the Bayesian work (see Boukabour and Masmoudi (2020)).

# Bibliography


Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biom. J.*, **68**, 326-328.

Beirlant, J., Goegebeur, Y. and Teugels, J. (2004). *Statistics of extremes-Theory and applications.* Chichester: Wiley.

Ben Khadher, F. and Slaoui, Y. (2021a). The Stochastic Approximation Method for Estimation of a Distribution Function under $\alpha$-mixing condition. *Int. J. Math. Stat.*, **22**, 1-19.

Ben Khadher, F. and Slaoui, Y. (2021b). Strong consistency of the mode of multivariate recursive kernel density estimator under strong mixing hypothesis. *Theory Stoch. Process.*.

Ben Khadher, F. and Slaoui, Y. (2021c). The stochastic approximation method for recursive kernel estimation of the conditional extreme value index. *J. Stat. Theory Pract.*.

Benveniste, A., Métivier, M. and Priouret, P. (1990). *Adaptive Algorithm and Stochastic Approximations.* Berlin: Springer.

Beran, J., Schell, D. and Stehlík, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Statist. Math.*, **66**, 193-220.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge: Cambridge University Press.

Blum, J.R., Hanson, D. L. and Koopmans, L.H. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrsch. Ver. Geb.*, **2**, 1-11.

Bosq, D. (1999). *Nonparametric statistics for stochastic processes: estimation and prediction.* New York: Springer.

Boukabour, S. and Masmoudi, A. (2020). Semiparametric Bayesian networks for continuous data. *Comm. Statist. Theory Methods* doi.org/10.1080/03610926.2020.1738486.

Bouzebda, S. and Slaoui, Y. (2018). Nonparametric recursive method for kernel-type function estimators for spatial data. *Statist. Probab. Lett.*, **139**, 103-114.

Bouzebda, S. and Slaoui, Y. (2020). Bandwidth selector for nonparametric recursive density estimation for spatial data defined by stochastic approximation method. *Commun. Statist. Theory Methods*, **49**, 2942-2963.

Bowman, A., Hall, P. and Prvan,T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biom. J.*, **85**, 799-808.

Bradley, R. (2007). *Introduction to strong mixing conditions*. Heber City: Kendrick Press.

Brilhante, M. F., Gomes, M. I. and Pestana, D. (2013). A simple generalisation of the Hill estimator. *Comput. Statist. Data Anal.*, **57**, 518-535.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, **17**, 790-799.

Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.*, **16**, 31-41.

Collomb, G., Hardle, W., and Hassani, S. (1987). A note on prediction via estimation of the conditional mode function. *J. Statist. Plann. Inference*, **15**, 227-236.

Daouia, A., Florens, J.P. and Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli*, **16**, 1039-1063.

Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST*, **20**, 311-333.

Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, **19**, 2557-2589.

De Haan, L. and Fereira, A. (2006). *Extreme Value Theory-An Introduction*. New York: Springer.

Dedecker, J., Doukhan, P., Lang, G., Leon, JR., Louhichi, S. and Prieur, C. (2007). *Weak dependence with examples and applications*. Berlin: Springer.

Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833–1855.

Delyon, B. (1996). General results on the convergence of stochastic algorithms. *IEEE Trans. Automat. Control.*, **41**, 1245-1255.

Dempster, A. P ., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **39**, 1—22.

Doob, J.L. (1955). *Stochastic Processes.* NewYork: John Wiley-Sons.

Doukhan, P. (1994). *Mixing: properties and examples.* New York: Springer.

Eddy, W.F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.*, **8**, 870-882.

Eddy, W.F. (1982). The asymptotic distributions of kernel estimators of the mode. *Probab. Theory Related Fields*, **59**, 279-290.

Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events.* New York: Springer.

Gumbel, E. J. (1958). *Statistics of Extremes.* New York: Columbia University Press.

Ferrez, J., Davison, A.C. and Rebetez, M. (2011). Extreme temperature analysis under forest 400 cover compared to an open field . *Agricultural and Forest Meteorology*, **151**, 992-1001.

Fisher, R. and Tippet, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc.*, **24**, 180-190.

Galambos, J. and Seneta, E. (1973). Regularly Varying Sequences. *Proc. Amer. Math. Soc.*, **41**, 110-116.

Gannoun, A., and Louani, D. (1996). Asymptotic properties of kernel estimators of the mode under censoring. *Technical Report no. 96-04, University of Montpellier 11.*

Gannoun, A., Girard, S., Guinot, C. and Saracco, S. (2002). Reference ranges based on nonparametric quantile regression. *Stat. Med.*, **21**, 3119-3135.

Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *J. Multivariate Anal.*, **99**, 2368-2388.

Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *J. Multivariate Anal.*, **101**, 419-433.

Goegebeur, Y., Guillou, A. and Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics*, **48**, 732-755.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.*, **44**, 423-453.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.*, **6**, 109-115.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.*, **20**, 675-694.

Hedges, S.B., and Shah, P. (2003). Comparison of mode estimation methods and application in molecular clock analysis. *BMC bioinformatics*, **4**, 31.

Helali, S. and Slaoui, Y. (2020). Estimation of a distribution function using Lagrange polynomials with Tchebychev–Gauss points. *Stat. Interface*, **13**, 399-410.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163-1174.

Hill, P.D. (1985). Kernel Estimation of a Distribution Function. *Comm. Statist. Theory Methods*, **14**, 605-620.

Hirschfeld, H.O. (1935). A connection between correlation and contingency. *Math. Proceed. of the Cambridge Philosophical Soc*, **31**, 520-524.

Hofbaur, M.W., and Williams, B.C. (2002). Mode estimation of probabilistic hybrid systems. *In International Workshop on Hybrid Systems: Computation and Control (HSCC)*, 253–266.

Ibragimov, I.A. (1962). Some limit theorems from stationary processes. *Theory Proc. Appl.*, **7**, 349-382.

Isogai, E. and Hirose, K. (1994). Nonparametric Recursive Kernel Estimators of a Distribution Function. *Bull. Inform. Cybernet.*, **26**, 87-99.

Jiang, H., and Kpotufe, S. (2017). Modal-set estimation with an application to clustering. *arXiv preprint arXiv:1606.04166.*

Khardani, S. and Slaoui, Y. (2019). Recursive Kernel Density Estimation and Optimal Bandwidth Selection Under $\alpha$: Mixing Data. *J. Stat. Theory Pract.*, **13**, 36.

Kiefer J. and Wolfowitz J. (1952). Stochastic Estimation of the Maximum of a Regression Function. *Ann. Math. Statist.*, **23**, 462-466.

Kolmogorov, A.N. (1931). Uber die analytischen methoden inder wahrschein lichkeitsrechnung. *Math. Ann.*, **104**, 415-458.

Konakov, V.D. (1974). On the asymptotic normality of the mode of multidimensional distributions. *Theory Probab. Appl.*, **19**, 794-799.

Kushner, H. J. (1977). General convergence results for stochastic approximations via weak convergence theory. Journal of mathematical analysis and applications. *J. Math. Anal. Appl.*, **61**, 490-503.

Kushner, H., and Yin, G. G. (2003). Stochastic approximation and recursive algorithms and applications. *Springer Science and Business Media.*

Lekina, A., (2010). Estimation Non-paramétrique des Quantiles Extrêmes Conditionnels. *Ph.D. thesis.*

Ljung L. (1978). Strong Convergence of a Stochastic Approximation Algorithm. *Ann. Statist.*, **6**, 680-696.

Louani, D. (1998). On the asymptotic normality of the kernel estimators of the density function and its derivatives under censoring. *Comm. Statist. Theory Methods*, **27**, 2909-2924.

Louani, D., and Ould Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametr. Stat.*, **11**, 413-442.

Masry, E. (1986). Recursive probability density estimation for weakly dependent process. *IEEE Trans. Inform. Theory*, **32**, 254-267.

McLachlan, G. and Peel, D. (2004). *Finite mixture models.* John Wiley and Sons.

Milet, J., Nuel, G., Watier, L., Courtin, D., Slaoui, Y., Senghor, P., Migot-Nabias, F., Gaye, O. and Garcia, A. (2010). Genome wide linkage study, using a 250K SNP map, of Plasmodium falciparum infection and mild malaria attack in a Senegalese population. *PLoS One*, **5**, e11616.

Mokkadem, A. and Pelletier, M. (2007). A Companion for the Kiefer–Wolfowitz–Blum Stochastic Approximation Algorithm. *Ann. Statist.*, **35**, 1749-1772.

Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference*, **139**, 2459-2478.

Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009b). Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA. Lat. Am. J. Probab. Math. Stat.*, **6**, 63-114.

Nadaraya, E.A. (1964). Some New Estimates for Distribution Functions. *Theory Probab. Appl.*, **9**, 497-500.

Nadaraya, E.N. (1965). On nonparametric estimates of density functions and regression curves. *Theory Probab. Appl.*, **10**, 186-190.

Nezam Sarmadi, S.A., and Venkatasubramanian, V. (2014). Electromechanical mode estimation using recursive adaptive stochastic subspace identification. *IEEE Trans. Power Syst.*, **29**, 349–358.

Ndao, P., Diop, A. and Dupuy, J.F. (2016). Nonparametric estimation of the conditional extreme value index with random covariates and censoring. *J. Statist. Plann. Inference*, **168**, 20-37.

Ould Saïd, E. (1993). Estimation non paramétrique du mode conditionnel application à la prévision. *C. R. Math. Acad. Sci. Paris*, **316**, 943-947.

Ould Saïd, E. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametr. Stat.*, **17**, :797-806.

Ould Saïd, E., and Tatachak, A. (2009). Strong Consistency Rate for the Kernel Mode Estimator Under Strong Mixing Hypothesis and Left Truncation. *Comm. Statist. Theory Methods*, **38**, :1154-1169.

Park, M. U. and Marron, J. S. (1990). Comparison of Data-Driven Bandwidth Selectors. *J. Amer. Statist. Assoc.*, **85**, 66-72.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065-1076.

Paulauskas, V. and Vaiěciulis, M. (2013). On an improvement of Hill and some other estimators. *Lith. Math. J.*, **53**, 336-355.

Paulauskas, V. and Vaiěciulis, M. (2017). A class of new tail index estimators. *Ann. Inst. Statist. Math.*, **69**, 461-487.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119-131.

Pisarenko, V.F. and Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. *Pure and Applied Geophysics*, **160**, 2343-2364.

Quintela-del-Río, A. and Estévez-Pérez, G. (2012). *Nonparametric Kernel Distribution Function Estimation with Kerdist: An R Package for Bandwidth Choice and Applications. J. Statist. Softw.*, **8**, 1-21.

Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions. *Scand. J. Stat.*, **8**, 116-119.

Reiss, R.D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values.* Basel, Switzerland: Birkhauser.

Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.* **8**, 391-398.

Révész, P. (1977). How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics.* **8**, 119-126.

Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants. Mathématiqueset applications.* New York: Springer.

Romano, J.P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.*, **16**, 629-647.

Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *Ann. Math. Statist.*, **22**, 400-407.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA*, **42**, 43-47.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.*, **27**, 832-837.

Ruppert, D. (1982). Almost Sure Approximations to the Robbins–Monro and Kiefer–Wolfowitz Processes with Dependent Noise. *Ann. Probab.*, **10**, 178-187.

Samanta, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.*, **7**, 109-117.

Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plann. Inference*, **35**, 65-75.

Sheikh, Y.A., Khan, E.A., and Kanade, T. (2007). Mode-seeking by medoidshifts. *In Proc. 11th IEEE Internat. Conf. on Computer Vision*, 1-8.

Singh, R. S., Gasser, T., and Prasad, B. (1983). Nonparametric estimates of distribution functions. *Comm. Statist. Theory Methods*, **12(18)**, 2095-2108.

Slaoui, Y. (2013). Large and moderate principles for recursive kernel density estimators defined by stochastic approximation method. *Serdica Math. J.*, **39**, 53-82.

Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *J. Probab. Stat* **2014**, ID 739640, doi:10.1155/2014/739640.

Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.*, **23**, 306-325.

Slaoui, Y. and Nuel, G. (2014c). Parameter estimation in a hierarchical random intercept model with censored response approach using a SEM algorithm and Gibbs sampling, *Sankhya B*, **76**, 210-233.

Slaoui, Y. (2015). Plug-in bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat Neerl*, **69**, 483-509.

Slaoui, Y. (2016a). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat Interface*, **9**, 375-388.

Slaoui, Y. (2016b). On the choice of smoothing parameters for semi-recursive nonparametric hazard estimators. *J Stat Theory Pract*, **10**, 656-672.

Slaoui, Y. (2018). Data-driven bandwidth selection for recursive kernel density estimators under double truncation, *Sankhya B*, **80**, 341-368.

Slaoui, Y. (2018) Bias reduction in kernel density estimation, *J. Nonparametr. Stat.* **30** 505-522.

Slaoui, Y. (2019a). Automatic bandwidth selection for recursive kernel density estimators with length biased data, *Jpn. J. Stat. Data Sci.*, doi: 10.1007/s42081-019-00053-z.

Slaoui, Y. (2019b). Data-driven deconvolution recursive kernel density estimators defined by stochastic approximation method, *Sankhya A*, doi: 10.1007/s13171-019-00182-3.

Slaoui, Y. (2019) Large and moderate deviation principles for recursive kernel distribution estimators defined by stochastic approximation method, *Opuscula Math.* **39**, 733-746.

Slaoui Y. and Jmaei A. (2019). Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials. *Stat. Interface*, **12**, 439-455.

Slaoui Y. (2019). Wild bootstrap bandwidth selection of recursive nonparametric relative regression for independent functional data. *J. Multivariate Anal.*, **173**, 494-511.

Slaoui, Y. (2020). Recursive nonparametric regression estimation for independent functional data. *Statist. Sinica*, **30**, 417-437.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.

Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electron. J. Stat.*, **7**, 2298-2343.

Tao, W., Jin, H. and Zhang, Y. (2007). Colour image segmentation based on mean shift and normalized cuts. *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, **37**, 1382-1389.

Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Probl. Inf. Transm.*, **8**, 119-126.

Tsybakov, A.B. (2009). *Introduction to nonparametric estimation.* New York: Springer Series in Statistics.

Van Ryzin, J. (1969). On strong consistency of density estimates. *South African Statist. J.*, **40**, 1765-1772.

Vieu, P. (1996). A note on density mode function. *Statist. Probab. Lett.*, **26**, 297-307.

Volkonskii, VA. and Rozanov, YuA. (1959). Some limit theorems for random functions. *Theory Probab. Appl*, **2**, 178-197.

Watson, G. S., and Leadbetter, M. R., (1964). Hazard analysis II. *Sankhya A*, **26**, 101-116.

Williams, B., Chung, S., and Gupta, V. (2001). Mode estimation of model-based programs: monitoring systems with complex behavior. *In Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 579-590.

Winter, B. B., (1973). Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.*, **1(1-2)**, 247-253.

Yamato, H., (1973). Uniform convergence of an estimator of a distribution function.

Yao, Q. (1999). Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury.

Yin, P., Tourapis, H.-Y., Tourapis, A., and Boyce, J. (2003). Fast mode decision and motion estimation for JVT/H.264. *In: ICIP. IEEE.*, 853-856.