HAL
archives-ouvertes.fr

Non Parametric Estimation using Lagrange and Bernstein Polynomials

Salima Helali

▶ To cite this version:

Salima Helali. Non Parametric Estimation using Lagrange and Bernstein Polynomials. Statistics [math.ST]. Université de Sfax (Tunisie), 2021. English. tel-03286356

HAL Id: tel-03286356
https://hal.archives-ouvertes.fr/tel-03286356

Submitted on 14 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

SFAX UNIVERSITY
FACULTY OF SCIENCES OF SFAX
DEPARTMENT OF MATHEMATICS

DOCTORATE THESIS IN MATHEMATICS

Presented by

Salima HELALI

Laboratory of Probability and Statistics

Non Parametric Estimation using Lagrange and Bernstein Polynomials

---

Defended on June 10, 2021 in front of the committee:

Pr. **Hédi NABLI**	Sfax University	President
Pr. **Sophie DABO NIANG**	Lille University	Referee
Pr. **Ali LAKSACI**	King Khalid University	Referee
Pr. **Raoudha ZINE**	Sfax University	Member
Pr. **Afif MASMOUDI**	Sfax University	Supervisor
Pr. **Yousri SLAOUI**	Poitiers University	Supervisor
Pr. **Jean François DUPUY**	Rennes University	Invited Member

Abstract

The objective of this thesis is to develop new nonparametric estimation techniques to deal with the problem of edge effect. First, we consider a shrinkage method using Bernstein polynomials and a finite Gaussian mixture model to construct a semi-parametric density estimator. Efficiency and feasibility were confirmed by theoretical results and then by applications on simulated real data sets. Then, we adopted a method based on two-time scale stochastic approximation algorithms and Bernstein polynomials to improve the Tenbush's regression estimators. Some of its asymptotic properties were also discussed. Experimental studies were conducted to compare this new approach with the Bernstein estimator and the classical kernel estimator. Finally, we introduce a new method based on the Lagrange polynomial and Chebytchev-Gauss points to construct non-parametric estimators. We compare the new estimators with approaches based on Bernstein polynomials as well as with approaches based on kernels. Applications based on real data are envisaged to show the efficiency of these new methods as well as to compare them with other estimators.

Keywords: Asymptotic properties, Bernstein polynomial, Distribution estimator, EM algorithm, Lagrange polynomials, Gaussian mixture model, Kernel estimator, Shrinkage estimator, Regression estimation, Tchebychev-Gauss points, Two-time-scale stochastic approximation algorithms.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to my supervisors Prof. Afif MASMOUDI and Prof. Yousri SLAOUI for their full support, guidance and kindness during my thesis research. I am greatly indebted to them. It is an invaluable privilege to have had the opportunity to work with them.

I would also like to extend my gratitude to the jury members. I sincerely thank Prof. Sophie DABO NIANG and Prof. Ali LAKSACI, for the careful reading of the manuscript and their valuable comments. I would like to extend my sincere thanks to Prof. Raoudha ZINE, Prof. Jean François DUPUY and Prof. Hédi NABLI for having accepted to join the jury of my thesis defense.

I am very grateful to my sister Wafa HELALI for all the constant support, encouragements and valuable advices, she gave me. Many thinks also go to all the members of LPS SFAX. I have learned a lot from them. They inspired my love for statistics and their teaching helped me on my path to become a researcher in this field.

Special thanks go to Fatma, Abir, Rihem, Raja, Rahma, Dorsaf, Seloua, Itidel, Nadia, Meryam and Carlos, for all their support and friendship.

My greatest gratitude is to my dear family, specially to my parents, my brother, my two sisters and my little nephews and nieces: Zeineb, Yasmine, Omar and Idris, for all the sacrifices and the confidence they have given me. I dedicate this thesis to them.

Chapter 1

Introduction

Estimation is an intrinsic topic of statistics. For example, the estimation of regression function is undoubtedly a useful tool of data analysis to explore the relationship between a response variable and an explanatory variable. In addition, it has much to offer in terms of such applications as forecasting future opportunities and risks in business, predicting the causal relationships between parameters in biological systems and various other situations. It makes it possible to tackle observed phenomena, in a variety of fields such as geology, oceanography, economics, epidemiology, environmental sciences. These phenomena are often characterized by a real random variable X. Hence, the need to estimate the probability density function f seems to be quite crucial. Basically, existing estimation algorithms can be categorized into either parametric or non parametric approaches.

- ⋆ Within the parametric framework, in the case of density estimation, we suppose that the model f_θ is known, where θ is a vector parameter whose exact values are unknown. We, thus, reduce the problem of estimating f_θ to the problem of estimating the parameter θ, which stands for the main objective of the parametric approach. Several authors have developed statistical parametric models. Among them, we state the Expectation Maximization (EM) algorithm proposed by Dempster et al (1977).

- ⋆ The non-parametric approach supposes no particular shape for the probability density f. In this situation, it is natural to estimate one of the unknown functions characterizing the model (density function, regression...): this is the objective of functional estimation. In this context, several methods have been set forward for smooth estimation of density and distribution functions. The most popular one, called kernel method, was elaborated by Rosenblatt (1956). The advances were carried out by Parzen (1962) so as to estimate a density function. These estimators

have been further developed using stochastic approximation methods to make it updated from a sample of size n to one of size $n+1$ (see Mokkadem et al (2009) and Slaoui (2014)). However, kernel methods display estimation problems at the edges, when we have a random variable X with density function f supported on a compact interval. Several solutions to this problem have been identified in the literature. See, for instance, the original work of Vitale (1975) and extensions given by Tenbusch (1994), Babu et al. (2002), Kakizawa (2004) and Babu and Chaubey (2006).

The contribution of this thesis lies in establishing the interface between these large categories of estimations. The basic target of this research work resides in crystallizing new techniques of estimation using Bernstein polynomial and Lagrange polynomial in order to resolve the boundaries effects of kernel estimators. It rests upon the following parts:

- Non parametric estimation using Bernstein polynomials.
- Non parametric estimation using Lagrange polynomials.

This manuscript is structured as follows.

Part I First, our central focus is on considering the Bernstein polynomial in order to construct non parametric recursive estimators of regression function and semi parametric density approach. This is particularly appealing since Bernstein polynomials are known to yield very smooth estimates that typically have acceptable behaviour at the boundaries.

Contribution of Chapter 2: In this chapter, we introduce a brief review of the scientific background and display the terminology notations required to discuss the parametric and non parametric approaches. The opening was marked by recalling some basic definitions of parametric models. Afterwards, we explored the commonly used methods for non parametric estimation. In a nut shell, this chapter corresponds to the cornerstone for the subsequent chapters of this thesis.

Contribution of Chapter 3: This chapter revolves around the alleviation of the boundary problem when the probability density function has a bounded support. To this extent, we set forward a shrinkage method using the Bernstein polynomial and a finite Gaussian mixture model to construct a semi parametric density estimator, which improves the approximation at the edges.

Some asymptotic properties of the proposed approach are addressed, such as its probability convergence and its asymptotic normality. In order to assess the performance of the proposed estimator, a simulation study and some real data sets are undertaken.

Contribution of Chapter 4: In this chapter, we introduce a recursive estimators of the regression function grounded upon the two-time-scale stochastic approximation algorithms and the Bernstein polynomials. We investigate the asymptotic properties of these estimators. We compare the proposed estimators to the classic regression estimator using the Bernstein polynomial defined by Tenbusch. Results revealed that our identified recursive estimators can overcome the problem of the edges related kernel regression estimation with a compact support. The proposed recursive two-time-scale estimators are compared to the non recursive estimator elaborated by Tenbusch. The performance of the two estimators is illustrated via simulations as well as two real datasets.

Part II The second goal of this research project is to display what seems to be a new method based on Lagrange polynomial $\mathcal{L}_i$ and Tchebytchev-Gauss points in order to construct non parametric estimators and their recursive versions. The Tchebytchev-Gauss points $(x_i)_{1\leq i\leq m}$ correspond to the optimal choice of grids that give the best convergence $\sum_{i=1}^{m} v(x_i)\mathcal{L}_i(.) \to v(.)$ uniformly, when $m \to \infty$, for any continuous function v of class C^k (for $k \geq 1$) on the interval $[-1, 1]$. This finding was handled by Jackson in the early 20th century (see Jackson (1911), Jackson (1912)), which stands for the main idea of the proposed estimators.

Contribution of Chapter 5: In this chapter, we exhibit a new method to estimate distribution function, based on Lagrange polynomials and Tchebytchev-Gauss points. Some asymptotic properties of the proposed estimator are explored, such as its asymptotic bias, variance, mean squared error and Chung-Smirnov propriety. The asymptotic normality and the uniform convergence of the proposed estimator are equally tackled. To sum up, we would simply assert that the performance of the proposed estimator is explored through a certain simulation examples. This research work was the subject of the following publication: Helali and Slaoui (2020).

Contribution of Chapter 6: In this chapter, we are basically concerned with introducing a recursive distribution estimator using Robbins Monro's al-

gorithm and Lagrange polynomial. Such asymptotic properties of the proposed estimator are examined, such as its asymptotic bias and variance. The asymptotic normality is also enacted. We attempt to demonstrate that the proposed estimator outperforms the non recursive distribution function introduced in the previous chapter, in terms of the asymptotic mean squared error and the computational costs. Generally, the proposed estimator compares favorably with other competitors in theoretical comparisons, as well as in simulation study and in real data sets examples.

Thesis organization Chapters 2, 3, 4, 5 and 6 illustrate the contributions of the thesis as pointed out above. The chapter concludes in a classical way, with a mini chapter to wrap up the whole work and offer some research perspectives, while the closing section highlights the different references invested in the document.

Chapter 2

Mathematical preliminaries

In this chapter, the readers are introduced to some basic concepts of statistical estimation. We first briefly display generalities and outstanding examples of the parametric estimation. Then, we provide general approaches of non parametric estimation and we detail methods allowing to determine optimal choice of smoothing parametric estimation. Subsequently, we introduce the Robbins-Monro stochastic algorithm which creates the recursive estimators. In this chapter, we consider $X, X_1, \ldots, X_n$ $(n \in \mathbb{N}^\star)$, a sequence of independent and identically distributed (*i.i.d.*) random variables having a common unknown density function f and a distribution function F. In addition, departing from the exhibited material, we shall invest the most pertinent concepts as well as prominent results leading us to opt for this specific path of research.

2.1 Parametric estimation

In this section, we first present the basic definitions and properties for the parametric estimation characterizing the unknown density f. Then, we shall provide a description of the finite mixture Gaussian model and the *EM* algorithm.

2.1.1 Definition of parametric estimation

Let's consider first the following definition.

Definition 2.1. *A parametric family of densities is a parametric family of densities on* $\mathbb{R}$*, defined by*

$$\mathcal{F} = \{f_\theta, \theta \in \Theta\},$$

where f_θ *is a distribution on* $\mathbb{R}$ *and* Θ *is a subset of* $\mathbb{R}^p$*.*

The basic idea of Definition (2.1) is that for the parametric estimation approach, the model structure of the distribution function is known beforehand. As matter of fact, there are usually uniquely a few characteristic parameters in the distribution functions that are needed to be estimated using different methods, such as the Maximum Likelihood (*ML*) method, Least Square approach and Moment method. The *ML* estimator is likely to be the most addressed notion theoretically, and the most used in practice, which is illustrated in the following definition.

Definition 2.2. *Assume all the distributions f_θ have a probability density function (pdf) l_θ on $\mathbb{R}$. Then the ML estimator $\widehat{\theta}_{ML}$ of θ is defined by*

$$\widehat{\theta}_{ML} \in \underset{\theta\in\Theta}{\operatorname{argmax}}\, l_\theta(x).$$

Let us display a simple example for the *ML* estimator.

Example 2.1.1. *Consider $\theta \in \mathbb{R}$, and l_θ is the pdf of n observations $x_1 \ldots, x_n$ from the Gaussian distribution $\mathcal{N}(\theta, 1)$ with mean θ and variance 1. In this case, we write*

$$l_\theta(x_1, \ldots, x_n) \quad = \quad (2\pi)^{-1/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right). \tag{2.1.1}$$

Maximizing the equation (2.1.1) yields the ML estimator $\widehat{\theta}_{ML} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$. One can also calculate $\mathbb{E}(\widehat{\theta}_{ML}) = \theta$. Therefore, the ML estimator is unbiased here.

Besides, owing to missing data, we need to find an estimator using an iterative algorithm. The most frequently used algorithm is the *EM* algorithm. It was developed by Dempster et al (1977). It is beneficial in a variety of incomplete-data problems.

2.1.2 Expectation maximization algorithm

The *EM* algorithm is an iterative way to find the *ML* estimator for the parameter θ, when data are incomplete or have unobserved latent variables. Each *EM* iteration alternates with an Expectation (*E-step*) and a maximization (*M-step*):

- The *E-step* holds the conditional expectation of the complete data log-likelihood given the observed data, using the current fit for the parameters.

- The *M-step* computes parameters maximizing the expected log-likelihood found in the E-step.

The computation of the likelihood equation is costly and time consuming as this equation tends to become complex. Therefore, the *EM* algorithm rests on simplifying this equation by using the log-likelihood equation instead of the likelihood equation. The *EM* algorithm stands for a popular tool in statistical estimation issues involving incomplete data or problems which can be posed in a similar form, such as the mixture parameters estimation (see Dempster et al (1977), McLachlan and Peel (2004)). For a better and thorough understanding of the *EM* algorithm, we try to elaborate in the next paragraph, the case of a Gaussian mixture model.

The Gaussian mixture model:

Within the parametric framework, it is noteworthy that the Gaussian mixture model can be used to estimate any density function. Consider $X_1, \ldots, X_n$, a sequence of *i.i.d.* with common Gaussian mixture density defined by

$$g(x|\theta) \quad = \quad \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \sigma_k)(x), \tag{2.1.2}$$

where

$$\theta = (\pi, \mu, \sigma) = (\pi_1 \ldots, \pi_K, \mu_1 \ldots, \mu_K, \sigma_1, \ldots, \sigma_K),$$

satisfies

$$0 \leq \pi_k \leq 1, \sum_{k=1}^{K} \pi_k = 1, \mu_k \in \mathbb{R}, \sigma_k > 0, \text{ for } k = 1 \ldots, K \text{ where } K > 0,$$

and

$$\mathcal{N}(\mu, \sigma)(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Finally, for each observed data point X_i, we associate a component label vector Z_i in order to manage the data clustering. This random vector $Z_i = (Z_{ik})_{1 \leq k \leq K}$ is defined such that $Z_{ik} = 1$ if the considered observation X_i is drawn form the k^{kt} component of the mixture, and $Z_{ik} = 0$ otherwise. Consequently, Z_i is distributed as a multivariate Bernoulli distribution with vector parameters $(\pi_1, \ldots, \pi_K)$ as follows:

$$\mathbb{P}(Z_i = z_i) = \prod_{k=1}^{K} \pi_k^{z_{ik}}.$$

Within the *EM* framework, $(X_1, \ldots, X_n, Z_1, \ldots, Z_n)$ corresponds to the complete data and $(Z_1, \ldots, Z_n)$ stands for the hidden data. Hence, the complete-data log likelihood is

expressed by

$$L(X_1,\ldots,X_n,Z_1,\ldots,Z_n,\theta) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{K} Z_{ij}\left[\log(\pi_j)+\log\left(\mathcal{N}(\mu_j,\sigma_j)(X_i)\right)\right]. \quad (2.1.3)$$

The two steps of the *EM* algorithm, after l iterations, are:

- *E-step*: The conditional expectation of the complete data log-likelihood given the observed data, using the current fit $\theta^{(l)}$, is defined by

$$\varphi\left(\theta|\theta^{(l)}\right) \;=\; \mathbb{E}_{\theta^{(l)}}\left(L(X_1,\ldots,X_n,Z_1,\ldots,Z_n,\theta)|X_1,\ldots,X_n\right). \quad (2.1.4)$$

  The posterior probability that X_i belongs to the j^{th} component of the mixture at the l^{th} iteration, is expressed as

$$\tau_{ij}^{(l)} = \mathbb{E}_{\theta^{(l)}}\left(Z_{ij}|X_1,\ldots,X_n\right) = \frac{\pi_j^{(l)}\mathcal{N}\left(\mu_j^{(l)},(\sigma^2)_j^{(l)}\right)(X_i)}{\sum_{h=1}^{K}\pi_h^{(l)}\mathcal{N}\left(\mu_h^{(l)},\left(\sigma_h^2\right)^{(l)}\right)(X_i)}. \quad (2.1.5)$$

  Finally, we get

$$\varphi\left(\theta|\theta^{(l)}\right) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{K}\tau_{ij}^{(l)}\left[\log(\pi_j)+\log\left(\mathcal{N}(\mu_j,\sigma_j)(X_i)\right)\right]. \quad (2.1.6)$$

- *M-step*: It consists of a global maximization of $\varphi\left(\theta|\theta^{(l)}\right)$ with respect to θ:

$$\theta^{(l+1)} \;=\; \arg\max_{\theta}\varphi\left(\theta|\theta^{(l)}\right). \quad (2.1.7)$$

  The updated estimates verify the following recursive formulas

$$\pi_j^{(l+1)} = \frac{1}{n}\sum_{i=1}^{n}\tau_{ij}^{(l)}, \quad (2.1.8)$$

$$\mu_j^{(l+1)} \;=\; \frac{\sum_{i=1}^{n}\tau_{ij}^{(l)}X_i}{\sum_{i=1}^{n}\tau_{ij}^{(l)}}, \quad (2.1.9)$$

$$(\sigma_j^2)^{(l+1)} \;=\; \frac{\sum_{i=1}^{n}\tau_{ij}^{(l)}\left(X_i-\mu_j^{(l+1)}\right)^2}{\sum_{i=1}^{n}\tau_{ij}^{(l)}}. \quad (2.1.10)$$

We repeat these two steps until $\left|\left|\theta^{(l+1)} - \theta^{(l)}\right|\right| < \epsilon$, where ϵ is a fixed threshold of convergence. The convergence properties of the *EM* algorithm have been investigated by Dempster et al (1977) and by Wu (1983). Relying upon Jensen's inequality, it can be noticed that, as $\varphi\left(\theta|\theta^{(l)}\right)$ is increasing, the log-likelihood function is also increasing (see Dempster et al (1977)). Consequently, the *EM* algorithm converges within a finite iterations number and gives the parameters maximum likelihood estimates. Therefore, under some conditions and according to Dempster et al (1977), we have

$$\lim_{l\to\infty} \pi_j^{(l)} = \widehat{\pi}_j, \ \lim_{l\to\infty} \mu_j^{(l)} = \widehat{\mu}_j \text{ and } \lim_{l\to\infty} (\sigma_j^2)^{(l)} = \widehat{\sigma^2}_j \text{ almost surely (a-s).} \qquad (2.1.11)$$

In what follows, we denote by $\widehat{\theta} = (\widehat{\pi}_1, \ldots, \widehat{\pi}_K, \widehat{\mu}_1, \ldots, \widehat{\mu}_K, \widehat{\sigma}_1, \ldots, \widehat{\sigma}_K)$.

In general, the parametric estimation will be more powerful if the form of the distribution function is somehow known in advance. In addition, the estimation technique is much simpler. However, sometimes, there is not enough prior information about the model of the distribution. To solve the problem, non-parametric estimation approach is set forward.

2.2 Non parametric estimation

Within the framework of the non parametric estimation, knowing that F is continuous, we consider the estimation of F by using smooth functions instead of the empirical distribution function, which is not continuous. Numerous methods have been set forward for smooth estimation of density and distribution functions. The most popular one, is called kernel method. In this section, we recall some results concerning kernel estimators. However, these methods have estimation problems at the edges, when we have a random variable X with distribution function supported on a compact interval. For this reason, we recall then, the estimators using Bernstein polynomials in order to overcome this problem. Moreover, we detail methods allowing to determine optimal choice of smoothing parametric estimation.

2.2.1 Kernel estimators

Kernel method is introduced by Rosenblatt (1956). The advances were carried out by Parzen (1962) to estimate density function. Let f defined on $\mathbb{R}$. We consider first the definition of a kernel.

Definition 2.3. *A function K with a symmetric support $[a,b]$ is called a kernel, if it is a positive function, such as $\int_a^b |K(x)|dx < \infty$,, $K(-x) = K(x)$, $\int_a^b xK(x)dx = 0$, $\int_a^b x^2K(x)dx < \infty$ and $\int_a^b K^2(x)dx < \infty$.*

All of these conditions make an asymptotically unbiased estimator with variance tending to zero when $n \to \infty$ (see Tsybakov (2009)). Some classical examples of Kernels are the following.

Example 2.2.1.

- *The rectangular Kernel:* $K(u) = \frac{1}{2}\mathbb{1}_{\{|u|\leq 1\}}$.
- *The triangular Kernel:* $K(u) = (1-|u|)\mathbb{1}_{\{|u|\leq 1\}}$.
- *The Epanechnikov Kernel:* $K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{\{|u|\leq 1\}}$.
- *The Gaussian Kernel:* $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$.

Kernel density estimation:

The Kernel estimator (see Rosenblatt (1956) and Parzen (1962)) of the density function is provided by

$$\widehat{f}_n(x) = \frac{1}{nh_n}\sum_{p=1}^{n} K\left(\frac{x-X_p}{h_n}\right), \forall x \in \mathbb{R}, \tag{2.2.1}$$

where (h_n) dependent on the size n, is a bandwidth (smoothing parameter, a sequence of positive real numbers that goes to zero). This approach is asymptotically unbiased with a variance that goes to zero when $n \to \infty$ with an optimal choice of the bandwidth (h_n). Multiple approaches were invested to estimate the smoothing parameter (h_n). The most ones beneficial are the cross-validation criterion and the Plug-in approach, which are further developed in the subsection (2.2.3).

Kernel distribution estimation:

The kernel distribution estimator was introduced by Nadaraya (1964) and was defined as

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{p=1}^{n} \mathcal{K}\left(\frac{x-X_p}{h_n}\right), \forall x \in \mathbb{R}, \tag{2.2.2}$$

where $\mathcal{K}(y) = \int_{-\infty}^{y} K(u)du$. Some theoretical properties of the estimator $\widehat{F}_n$ were explored (see among many others, Nadaraya (1964), Reiss (1981) and Peter (1985)). Reiss

(1981) and Falk (1977) showed that Nadaraya's kernel distribution estimator (5.1.1) has an asymptotically better performance than the empirical distribution function which does not consider the smoothness of F.

Kernel regression estimation:

Regression corresponds to a set of statistical methods that are widely used to analyze the relationship between a target variable and one or more other variables. Let $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ be *i.i.d.* pairs of random variables with joints density function g and let f such that $f(x)=\int_{\mathbb{R}} g(x,t)dt$ denote the density of X. The non parametric regression model is represented as

$$Y = r(X)+\epsilon, \tag{2.2.3}$$

where ϵ and X are independent with $\mathbb{E}(\epsilon)=0$. Then, the regression function is $r(x)=\mathbb{E}(Y|X=x)$, which models the relationship between X and Y. In what follows, the random variable Y is supposed to be integrable. The kernel regression estimator r is based on the concept of the kernel density estimator as

$$\widehat{r}_n(x) = \frac{\int y\widehat{g}_n(x,y)dy}{\int \widehat{g}_n(x,y)dy}\mathbb{1}_{\{\int \widehat{g}_n(x,y)dy\neq 0\}} = \frac{\int y\widehat{g}_n(x,y)dy}{\widehat{f}_n(x)}\mathbb{1}_{\{\widehat{f}_n(x)\neq 0\}}, \tag{2.2.4}$$

where $\widehat{g}_n$ is a kernel estimator of g defined as

$$\widehat{g}_n(x,y) = \frac{1}{nh_n^2}\sum_{p)1}^{n} K\left(\frac{x-X_p}{h_n}\right)K\left(\frac{y-Y_p}{h_n}\right). \tag{2.2.5}$$

Using Equation (2.2.5) leads to

$$\begin{aligned}
&\int_{\mathbb{R}} y\widehat{g}_n(x,y)dy \\
&= \frac{1}{nh_n}\sum_{p=1}^{n} K\left(\frac{x-X_p}{h_n}\right)\left[\int_{\mathbb{R}}\frac{y-Y_p}{h_n}K\left(\frac{y-Y_p}{h_n}\right)dy+\int_{\mathbb{R}}\frac{Y_p}{h_n}K\left(\frac{y-Y_p}{h_n}\right)dy\right] \\
&= \frac{1}{nh_n}\sum_{p=1}^{n} K\left(\frac{x-X_p}{h_n}\right)\left[h_n\int_{\mathbb{R}} uK(u)du+Y_p\int_{\mathbb{R}}K(u)du\right].
\end{aligned}$$

Substituting this result into (2.2.4), according to Equation (2.2.1) and definition of the kernel function (2.3), we get the Nadaraya-Watson estimator of the regression function r

(see Nadaraya (1964) and Watson (1964)) defined as,

$$\widehat{r}_n(x) = \frac{\sum_{p=1}^{n} Y_p K\left(\frac{x-X_p}{h_n}\right)}{\sum_{p=1}^{n} K\left(\frac{x-X_p}{h_n}\right)} \mathbb{1}_{\sum_{p=1}^{n} K\left(\frac{x-X_p}{h_n}\right)\neq 0}. \tag{2.2.6}$$

This estimator exhibits an undesirable behaviour at the boundaries of the support of X (bias is of order h in the boundary region while it is of order h^2 away from the boundaries, see Wan and Jones (1994), Section 5.6.1). In fact, it is well known that kernel estimators have a bias that is larger by a full order of magnitude in the boundary region, when we have a random variable X with density function supported on a compact interval.

Support problem:

We suppose for simplicity that there is a single known boundary to the support of the density function f which we might as well take to be at the origin. Then, we deal with positive data. For convenience, we consider a symmetric kernel (for instance, normal kernels). Away from the boundary, which means that at any $x > h_n$, the usual asymptotic mean and variance expressions are applied. Let us now suppose that f has two continuous derivatives everywhere, and that as $n \to \infty$, $h_n \to 0$, and $nh_n \to 0$. Hence

$$\mathbb{E}(\widehat{f}_n(x)) \simeq f(x) + \frac{1}{2}h_n^2 f''(x) \int x^2 K(x)dx,$$

and

$$Var(\widehat{f}_n(x)) \simeq (nh_n)^{-1} f(x) \int K^2(x)dx.$$

Near from the boundary, the expression of the mean and the variance are different. Let $x = ph_n$, we have

$$\mathbb{E}(\widehat{f}_n(x)) \simeq f(x) \int_{-\infty}^{p} K(x)dx - f'(x) \int_{-\infty}^{p} xK(x)dx + \frac{1}{2}h_n^2 f''(x) \int_{-\infty}^{p} x^2 K(x)dx,$$

and

$$Var(\widehat{f}_n(x)) \simeq (nh_n)^{-1} f(x) \int_{-\infty}^{p} K^2(x)dx.$$

These bias phenomena are named boundary bias. Multiple authors set forward methods for reducing these phenomena such as data reflection (Schuster (1985)), boundary kernels (Müller (1991) and Müller and Wang (1994)), the local linear estimator (Lejeune and Sarda (1992) and Jones (1993)), the use of beta and gamma kernels (Chen (1996, 1999)). For a smooth estimate of a density function with a finite known support, Vitale's method (Vitale (1975)) based on the Bernstein polynomials is illustrated in the next subsection.

2.2.2 Bernstein estimators

In order to overcome the boundaries problem in the case of density estimation and the case of distribution estimation, various methods were addressed such as the Bernstein polynomial estimators (see the works of Vitale (1975), Tenbusch (1994), Babu et al. (2002) and Kakizawa (2004)). The first steps in non parametric estimation methods based on Bernstein polynomials were undertaken in 1912 in order to construct a probabilistic demonstration of the classical Weierstrass theorem which corresponds to what follows:

Theorem 2.2.1. *Any continuous function f on a segment $[a, b]$ is a uniform limit of polynomial functions on this segment. In other words, for everything $\epsilon > 0$, there is a Q polynomial such as:*

$$\forall x \in [a, b],\ |f(x) - Q(x)| < \epsilon.$$

Within this framework, Bernstein (1912) introduced a family of polynomials, defined as follows.

Bernstein polynomials

Definition 2.4. *For $m \in \mathbb{N}$ and $0 \leq k \leq m$, we call Bernstein's polynomial,*

$$b_k(m, x) = C_m^k x^k (1 - x)^{m-k}.$$

These polynomials have several properties, such as

Proposition 2.1. *Bernstein polynomials have the following properties:*

1. *Partition of the unit:*
$$\sum_{k=0}^{m} b_k(m, x) = 1,\ \forall x \in [0, 1].$$
2. *Positive:*
$$b_k(m, x) \geq 0,\ \forall k \in \{0, \ldots, m\}.$$
3. *Symmetry:*
$$b_k(m, x) = b_{m-k}(m, 1 - x),\ \forall k \in \{0, \ldots, m\}.$$
4. *Recurrence formula:*
$$b_k(m, x) = \begin{cases} (1 - x)b_k(m - 1, x) & \text{if } k = 0, \\ (1 - x)b_k(m - 1, x) + xb_{k-1}(m - 1, x) & \text{if } k \in \{1, \ldots m - 1\}, \\ xb_{k-1}(m - 1, x) & \text{if } k = m. \end{cases}$$

From a probabilistic point of view, the polynomial $b_k(m,p)$ is the probability $P(X=p)$, where X is a random variable following a binomial law of parameter (m,p). This is the idea that Bernstein adopted in his demonstration of the Weierstrass theorem, which is otherwise written as follows.

Theorem 2.2.2. *Let $f:[0,1]\to\mathbb{R}$ be a continuous function. We define the Bernstein polynomial associated with f of order $m\in\mathbb{N}$ in terms of:*

$$\forall x\in[0,1],\ B_m(f)(x)=\sum_{k=0}^{m} f\left(\frac{k}{m}\right) b_k(m,x).$$

Thus, we have

$$\lim_{m\to\infty}||f-B_m(f)||=\lim_{m\to\infty}\sup_{x\in[0,1]}|f(x)-B_m(f)(x)|=0.$$

As an excellent reference for properties of Bernstein polynomials, we refer the reader to Lorentz (1986). Vitale (1975) invested the preceding theorem to construct an estimator of the distribution function which will be discussed in the next paragraph. In the following, we assume that $m=m_n$ depends on n.

Distribution estimation

Vitale (1975) defined the estimator of F as follows,

$$\tilde{F}_{n,m}(x) \quad = \quad \sum_{k=0}^{m}\overline{F}_n\left(\frac{k}{m}\right) b_k(m,x), \tag{2.2.7}$$

where $\overline{F}_n(x)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbb{1}_{X_i\leq x}$ is the empirical distribution function. We note that

$$\tilde{F}_{n,m}(0)=0 \text{ and } \tilde{F}_{n,m}(1)=1.$$

The estimator defined in (2.2.7) was tackled by Babu et al. (2002) and Leblanc (2010, 2012a). The bias and the variance of the estimator $\tilde{F}_{n,m}$ are given in the following proposition.

Proposition 2.2 (Leblan 2012)**.** *Let F be a continuous function and let it admit two continuous and bounded derivatives on $[0,1]$. We have for $x\in(0,1)$ that*

$$Bias(\tilde{F}_{n,m}(x))=m^{-1}b(x)+o(m^{-1}),$$

where $b(x) = [x(x-1)f'(x)]/2$. *Besides, we have*

$$Var(\tilde{F}_{n,m}(x)) = n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + o_x\left(m^{-1/2}n^{-1}\right),$$

where $\sigma^2(x) = F(x)[1-F(x)]$ *and* $V(x) = f(x)[2x(1-x)/\pi]^{1/2}$.

We note that if we consider $h = m^{-1}$ as the 'bandwidth' of the Bernstein estimator, the bias of $\tilde{F}_{n,m}$ becomes $O(m^{-1}) = O(h)$, which is bigger than the typically obtained bias using kernel estimators which have basically at least a bias as small as $O(h^2)$ (except possibly near from the boundaries). In addition, the previous proposition implies that,

$$MSE(\tilde{F}_{n,m}) = n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + m^{-2}b^2(x) + o(m^{-2}) + o_x\left(m^{-1/2}n^{-1}\right).$$

The optimal choice of the order m that minimizes MSE is indicated by

$$m_{opt} = n^{2/3}\left[\frac{4b^2(x)}{V(x)}.\right]^{2/3}.$$

it is noteworthy that the plug-in approach elaborated by Altman and Léger (1995) and the cross-validation method of Bowman et al (1998) can also be adapted in the practical choice of m. For further details, please refer back to the section (2.2.3). Starting from the distribution estimator $\tilde{F}_{n,m}$, Vitale constructed the density estimator $f_{n,m}$ which will be handled in the following paragraph.

Density estimation

To obtain the density estimator $f_{n,m}$ using the Bernstein polynomial, we first write the distribution estimator in the Babu et al. (2002) form as follows,

$$\tilde{F}_{n,m}(x) = \sum_{k=0}^{m} f_n\left(\frac{k}{m}\right) B_k(m,x), \tag{2.2.8}$$

where $f_n(0) = 0$, $f_n(k/m) = \widehat{F}_n(k/m) - \widehat{F}_n((k-1)/m)$ for $k = 1 \ldots m$
and $B_k(m,x) = \sum_{j=k}^{m} b_j(m,x)$. With respect to x Equation (2.2.8), we obtain

$$\begin{aligned}\tilde{f}_{n,m}(x) &= \sum_{k=0}^{m} f_n\left(\frac{k}{m}\right)\frac{d}{dx}B_k(m,x) \\ &= m\sum_{k=1}^{m-1}\left[\overline{F}_n\left(\frac{k+1}{m}\right) - \overline{F}_n\left(\frac{k}{m}\right)\right] b_k(m-1,x).\end{aligned} \tag{2.2.9}$$

This estimator was introduced by Vitale (1975) and examined by Babu et al. (2002). In addition, Bouezmarni and Rolin (2007) considered the case of unbounded densities As for, Tenbusch (1994) and Babu and Chaubey (2006), they generalized this estimator to the multidimensional case. For additional details, the reader can consult the work of Leblanc (2012b) and Babu et al. (2002). Results are then exhibited for the bias and the variance of the estimator $\tilde{f}_{n,m}$.

Proposition 2.3 (Leblan 2012)**.** *Let f be a density function. We assume that f is a continuous function with two derivatives bounded on $[0,1]$. For $x \in [0,1]$, we have*

$$Bias(\tilde{f}_{n,m}(x)) = m^{-1}\delta_1(x) + o(m^{-1}),$$

where $\delta_1(x) = \frac{1}{2}[(1-2x)f'(x) + x(1-x)f''(x)]$. On the other side, we have

$$Var(\tilde{f}_{n,m}(x)) = \begin{cases} \frac{m^{1/2}}{n} f(x)\psi(x) + o_n\left(\frac{m^{1/2}}{n}\right) & \text{if } x \in]0,1[, \\ \frac{m}{n} f(x) + o_x\left(\frac{m}{n}\right) & \text{if } x = 0, 1, \end{cases}$$

where $\psi(x) = (4\pi x(1-x))^{-1/2}$.

If we consider that $h = m^{-1}$ the bandwidth of the density estimator $\tilde{f}_{n,m}$, we get a bias of the order of $O(h)$ for the estimator $\tilde{f}_{n,m}$, which is bigger than that obtained by the kernel estimator $\widehat{f}_n$. On the other side, the variance of the estimator $\tilde{f}_{n,m}$ is of the order of $O(h^{-1/2}/n))$ for $x \in]0,1[$, which is small that that of the kernel estimator, which is of the order of $O(h^{-1}/n)$. We also notice that the variance is of the order of $O(m^{1/2}/n)$ within the interval $]0,1[$, but it is of a higher order that is equal to $O(m/n)$ at the boundaries. Based on the density estimator $\tilde{f}_{n,m}$ defined in (2.2.9), Tenbusch (1997) identified a regression estimator using Bernstein polynomials, which is elucidated in the next paragraph.

Regression estimation

Let $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ be *i.i.d.* pairs of random variables with joints density function $g(x,y)$ and let f such that $f(x) = \int_{\mathbb{R}} g(x,t)dt$, denote the density of X which is supported on $[0,1]$. Let $\mathbb{E}(|Y|) < \infty$. The Bernstein estimator of the regression function

r is $x \in [0,1]$

$$\tilde{r}_{n,m}(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} Y_i \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x)}{\sum_{i=1}^{n} \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x)} & \text{if } \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x) \neq 0, \\ 0 & \text{if } \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x) = 0. \end{cases} \tag{2.2.10}$$

Like the smoother kernel, the Bernstein estimator corresponds to an intuitive estimator of a regression function. At each point x_0, the estimator is a weighted mean of the sample Y_i. However, contrarily to the kernel estimator, the Bernstein estimator does not always use the same weight function (kernel). The Bernstein estimator rather adjusts the weight function to the positions of the X_i's in the predictor space. This adjustment of the weight function amounts in better boundary behavior of the Bernstein estimator compared to the Nadaraya-Watson estimator. To get a pertinent estimation by the kernel method or Bernstein polynomials, it is significant to choose the smoothing parameters (h_n) and m. Some methods as far as the choice of these parameters are displayed in the next paragraph.

2.2.3 Choice of smoothing parameter

Within the framework of non parametric estimators, several approaches were used to estimate the smoothing parameter. Among the most famous and useful ones, we mentioned the cross-validation criterion and the Plug-in approach. A detailed comparison of these techniques is stated in Delaigle and Gijbels (2004). We are basically interested in following the principle of cross-validation.

Cross validation method

Step 1: The starting sample $(X_1, \ldots, X_n)$ is cut into V packets $L_1, \ldots, L_V$ of the same size n/V.

Step 2: For $v = 1 \ldots V$,

(a) We proposed an estimator $\widehat{f}_h^v$ of the density f from all observations, except the L_v package.

(b) We proposed an estimator $\widehat{MISE}^{v}(h)$ an estimator of the $MISE(\widehat{f}_h^v)$ from L_v package observations.

Step 3: A global risk estimator is constructed

$$\widehat{MISE}_f^{CV}(h) = \frac{1}{V}\sum_{v=1}^{V}\widehat{MISE}^{v}(h) \text{ and } h_{opt} = arg\,min_{h>0}\widehat{MISE}_f^{CV}(h).$$

In the following section, we will handle recursive estimators using the stochastic algorithm of Robbins and Monro (1951).

2.3 Non parametric recursive estimation

Stochastic algorithms were frequently used in many research applications involving sequential change detection, system identification and transmission systems. The recursive estimators can be updated with each supplementary new observation and grant multiple advantages. They do not need extensive storage of data and they are fast to compute.

2.3.1 Methodology

The general form of stochastic algorithm is:

$$\theta_n = \theta_{n-1} + \gamma_n\phi\left(\theta_{n-1}, W_n\right) + \gamma_n^2\mu_n\left(\theta_{n-1}, W_n\right), \tag{2.3.1}$$

where

- (γ_n) is a positive sequence of real numbers decreasing towards zero.
- (θ_n) the sequence to be recursively updated.
- (W_n) is a sequence of random variables representing the on-line observations.
- $\phi(\theta, W)$ is the function which essentially defines how the parameter θ is updated as a function of new observation.
- $\mu_n\left(\theta_{n-1}, W_n\right)$ is 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 (4.7.10) coincides with the one analysed by Kushner (1977), Ljung (1978) and Ruppert (1982):

$$\theta_n = \theta_{n-1} + \gamma_n\left[\phi(\theta_{n-1}) - W_n + \eta_n\right], \tag{2.3.2}$$

where

- η_n stands for a random variables converges to 0 almost surely.
- ϕ corresponds to a measurable unknown function.

They asserted that (2.3.2) 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 ϕ. 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 σ-algebra of the events occurring up the time $n-1$. Under standard conditions on the function ϕ and on the sequence (γ_n), Kushner and Yin (2003) highlighted that

$$\theta_n \to \theta^\star \text{ almost surely (a.s.).}$$

In the following subsections, two examples of recursive estimators are constructed using the Robbins and Monro algorithm.

2.3.2 Recursive kernel distribution estimator

In order to construct a stochastic algorithm, which approximates the function F at a given point x, Slaoui (2014) defined an algorithm to search 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 (γ_n) is a positive sequence of real numbers decreasing to zero and (Q_n) is a sequence of functions $Q_n : \mathbb{R} \to \mathbb{R}$ defined by $Q_n(x) = \phi(F_{n-1}(x)) - W_n + \eta_n$. Using the fact that $\mathbb{E}(W_n|\mathcal{F}_{n-1}) = 0$, where $\mathcal{F}_{n-1}$ stands for the σ-algebra of the events occurring up the time $n-1$, it comes $\mathbb{E}(Q_n(x)) = F(x) - F_{n-1}(x) + \eta_n$. Following 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 $Q_n(x) = \mathcal{K}(h_n^{-1}(x - X_n)) - F_{n-1}(x)$.

Therefore, the recursive estimator F_n of the distribution function F at the point x can be stated as

$$F_n(x) = (1-\gamma_n)F_{n-1}(x) + \gamma_n \mathcal{K}(h_n^{-1}(x - X_n)). \tag{2.3.3}$$

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 (6.1.3) that F_n can be rewritten as

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k \mathcal{K}\left(\frac{x - X_k}{h_k}\right). \tag{2.3.4}$$

it is inferred that, with an adequate choice of the stepsize (γ_n) and the bandwidth (h_n), the $MWISE$ (Mean Weighted Integrated Squared Error) of the Slaoui'estimator F_n (7) is smaller than that of Nadaraya's estimator $\widehat{F}_n$ (5.1.1). A similar steps were invested by Jmaei et al. (2017) to construct a recursive distribution estimator using Bernstein polynomials in order to resolve the edge effects of kernel estimators.

2.3.3 Recursive kernel regression estimator

Slaoui (2016) developed a stochastic algorithm to construct a semi-recursive estimator of the regression function $r(x) = \mathbb{E}(Y|X = x)$. First we construct a stochastic algorithm for the estimation of the function $a : x \to r(x)f(x)$. Then, we define an algorithm of search for the zero of the function $l : y \to a(x) - y$ as follows:

(i) We set $a_0(x) \in \mathbb{R}$.

(ii) For all $n \geq 1$, we set $a_n(x) = a_{n-1}(x)+\beta_n W_n$, where the stepsize (β_n) is a sequence of positive real numbers that goes to zero and $W_n(x) = h_n^{-1}Y_nK(h_n^{-1}(x-X_n))-a_{n-1}(x)$.

Then, the considered estimator to recursively estimate the function a at the point x can be written as

$$a_n(x) = S_n \sum_{k=1}^{n} S_k^{-1}\beta_k h_k^{-1} Y_k K\left(\frac{x - X_k}{h_k}\right),$$

where $S_n = \prod_{j=1}^{n}(1-\beta_j)$ and $a_0(x) = 0$. Moreover, we use the estimator introduced in Mokkadem et al (2009) to estimate recursively the density f at the point x as follows

$$f_n(x) = (1-\gamma_n)f_{n-1}(x) + \gamma_n h_n^{-1}K(h_n^{-1}[x - X_n]),$$

where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero. Let $f_0(x) = 0$, and $\Pi_n = \prod_{j=1}^{n}(1-\gamma_j)$. Then, it follows that one can estimate f recursively

at the point x by

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

Therefore, we consider the semi-recursive estimator for the regression function r at the point x

$$r_n(x) = \begin{cases} \frac{an(x)}{f_n(x)} & \text{if } f_n(x) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

This semi-recursive estimator is very competitive to the non recursive one (2.2.6) at the level of estimation error and it proves to be much better in terms of computational costs using certain selected bandwidth (h_n) and several special stepsizes (γ_n) and (β_n).

Part I

Non parametric estimation using Bernstein polynomials

Chapter 3

Semi parametric estimator using Bernstein polynomials

Key words and phrases: Asymptotic properties, Bernstein polynomial, EM algorithm, Gaussian mixture model, Kernel estimator, Shrinkage estimator.

Abstract 1: *The central focus of this chapter is upon the alleviation of the boundary problem when the probability density function has a bounded support. Mixtures of beta densities have led to different methods of density estimation for data assumed to have compact support. Among these methods, we mention Bernstein polynomials which leads to an improvement of edge properties for the density function estimator. In this chapter, we set forward a shrinkage method using the Bernstein polynomial and a finite Gaussian mixture model to construct a semi parametric density estimator, which improves the approximation at the edges. Some asymptotic properties of the proposed approach are investigated, such as its probability convergence and its asymptotic normality. In order to evaluate the performance of the proposed estimator, a simulation study and some real data sets are carried out.*

3.1 Introduction

Density estimation is a widely adopted tool for multiple tasks in statistical inference, machine learning, visualization, and exploratory data analysis. Existing density estimation algorithms can be categorized into either parametric, semi parametric, or non parametric approaches. In the non parametric framework, several methods have been set forward for smooth estimation of density and distribution functions. The most popular one, called

kernel method, was introduced by Rosenblatt (1956). The advances were carried out by Parzen (1962) to estimate a density function. The reader is recommended to consult the chapter of Härdle (1991) for an introduction of several kernel smoothing techniques. However, kernel methods display estimation problems at the edges, when we have a random variable X with density function f supported on a compact interval. Moreover, if $X_1, \ldots, X_n$ is a sample with same density f, it is well known in non parametric kernel density estimation that the bias of the standard kernel density estimator

$$\widehat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \tag{3.1.1}$$

is of a larger order near the boundary than that in the interior, where K is a kernel (that is, a positive function satisfying $\int K(x)dx = 1$), and (h_n) is a bandwidth (that is, a sequence of positive real numbers that goes to zero). Let us now suppose that f has two continuous derivatives everywhere, and that as $n \to \infty$, $h = h_n \to 0$ and $nh \to 0$. Let $x = ph$ for $p > 0$. Near the boundary, the expression of the mean and the variance are indicated as

$$\mathbb{E}\left[\widehat{f}_n(x)\right] \simeq f(x)\int_{-\infty}^{p} K(x)dx - f'(x)\int_{-\infty}^{p} xK(x)dx + \frac{1}{2}h^2 f''(x)\int_{-\infty}^{p} x^2K(x)dx,$$

and

$$Var\left[\widehat{f}_n(x)\right] \simeq (nh)^{-1}f(x)\int_{-\infty}^{p} K^2(x)dx.$$

These bias phenomena are called boundary bias. Numerous authors have elaborated methods for reducing these phenomena such as data reflection (Schuster (1985)), boundary kernels (Müller (1991), Müller (1993) and Müller and Wang (1994)), the local linear estimator (Lejeune and Sarda (1992) and Jones (1993)), the use of beta and gamma kernels (Chen (1999) and Chen (2000)), the bias reduction (Leblanc (2010) and Slaoui (2018)). For a smooth estimator of a density function f with finite known support, there have been several methods such as Vitale's method (see Vitale (1975)) which is based on the Bernstein polynomials and expressed as

$$\tilde{f}_{1,n,m}(x) = m\sum_{k=0}^{m-1}\left[\overline{F}_n\left(\frac{k+1}{m}\right) - \overline{F}_n\left(\frac{k}{m}\right)\right] b_k(m-1, x), \tag{3.1.2}$$

where $\overline{F}_n$ is the empirical distribution function and $b_k(m, x)$ is the Bernstein polynomial. This estimator was investigated in literature (Ghosal (2000), Babu et al. (2002), Kakizawa (2004), Rao (2005)) and more recently by Leblanc (2010), Igarashi and Kakizawa (2014) and Slaoui and Jmaei (2019).

Within the parametric framework, it is noteworthy that the Gaussian mixture model, can be used to estimate any density function, without any problem of estimation on the edge. This refers to the fact that the set of all normal mixture densities is dense in the set of all density functions under the L^1 metric (see Li and Barron (2000)). The investigation of mixture models stands for a full field in modern statistics. It is a probabilistic model introduced by Pearson (1894) to illustrate the presence of subpopulations within an overall population. It has been developed so far by various authors like McLachlan and Peel (2004). It is used for data classification and it provides efficient approaches of model-Based clustering. Roeder and Wasserman (1997) demonstrated that when a Gaussian mixture model is used to estimate a density nonparametrically, the density estimator that uses the Bayesian information criterion of Schwarz (1978) to select the number of components in the mixture, is consistent (see Leroux (1992)).

However, we obtain the nonparametric kernel estimate of a density if we fit a mixture of n components in equal proportions $1/n$, where n is the size of the observed sample. As a matter of fact, it can be inferred that mixture models occupy an interesting niche between parametric and nonparametric approaches to statistical estimation.

The problem at the edge does not arise for the parametric model. For this reason, the basic idea of this work is to consider a shrinkage method using the Bernstein (Vitale's estimator) and the Gaussian mixture estimators, to construct a shrinkage density estimator, in order to improve the approximation at the edge. A shrinkage estimator is a convex combination between estimators (see James and Stein (1992)). Basically, this implies that a naive or raw estimate is improved by combining it with other information.

The remainder of this chapter is organized as follows. In the next section, we introduce a new semi parametric estimation approach based on the shrinkage method using the Bernstein polynomials and the Gaussian mixture densities. In Section 6.3, the consistency of the proposed estimator is exhibited as well as its asymptotic normality. Section 6.4 highlights a simulation study that compares the performance of the proposed approach with the Bernstein estimator and with the standard Gaussian kernel estimator. The closing Section 6.5 crowns the whole work, wraps the conclusion and provides new perspectives for future work.

3.2 Proposed approach

The proposed semi parametric approach rests upon the *shrinkage* combination between the Gaussian mixture model and the Bernstein density estimators using the *EM* algo-

rithm for the parameters estimations. The literature on shrinkage estimation is enormous. From this perspective, it is noteworthy to mention the most relevant contributions. James and Stein (1992) were the first to introduce the classic shrinkage estimator. Stein (1981) provided theory for the analysis of risk. Oman (Oman (1982a,b)) developed estimators which shrink Gaussian density estimators towards linear subspaces. An in-depth investigation of shrinkage theory is displayed in Chapter 5 of Lehmann and Casella (1992).

The proposed semi parametric approach based upon estimating the density function f, relies on the same principle of Stein's works, and there are two aspects along this line. The first setting is non-parametric in the sense that we do not assume any parametric form of the density. The non-parametric setting is very important as it allows us to perform statistical inference without making any assumption on the parametric form of the true density f. The second setting is to consider the Gaussian mixture model, as a parametric estimator, of the unknown density f.

In what follows, we consider $X_1, \ldots, X_n$ a sequence of *i.i.d.* random variables having a common unknown density function f supported on $[0, 1]$. We develop here a shrinkage method to estimate the density function, which is divided into three steps:

Step 1 We consider the Bernstein estimator of the density function f which is defined as

$$\tilde{f}_{1,n,m}(x) = m \sum_{i=0}^{m-1} \left[\overline{F}_n \left(\frac{i+1}{m} \right) - \overline{F}_n \left(\frac{i}{m} \right) \right] b_i(m-1, x) \tag{3.2.1}$$

Step 2 We consider the Mixture Gaussian density as an estimator of the density function f, given by

$$\tilde{f}_{2,n}(x) = \sum_{k=1}^{K} \widehat{\pi}_k \mathcal{N}(\widehat{\mu}_k, \widehat{\sigma}_k)(x), \tag{3.2.2}$$

where $\widehat{\mu}_k$, $\widehat{\sigma}_k$ and $\widehat{\pi}_k$ are estimated by the *EM* algorithm defined in (2.1.11).

Step 3 We consider the *shrinkage* density estimator $\widehat{f}_{n,m}$ form defined by

$$f_{n,m}^{Sh}(x) = \lambda \tilde{f}_{1,n,m}(x) + (1-\lambda)\tilde{f}_{2,n}(x),$$

and we use the *EM* algorithm to estimate the parameter $\lambda \in [0, 1]$ of the proposed model.

By the same way as considered in Section 2.1.2 of chapter 2, the two steps of the *EM* algorithm, after t iterations, are denoted in terms of:

1. *E-step*: The conditional expectation of the complete data log-likelihood given the observed data, using the current $\lambda^{(t)}$, is provided by

$$Q(\lambda|\lambda^{(t)}) = \sum_{i=1}^{n} \mathbb{E}_{\lambda^{(t)}}(W_{i1} \mid X_i) \log \tilde{f}_{1,n,m}(X_i) + \mathbb{E}_{\lambda^{(t)}}(W_{i2} \mid X_i) \log \tilde{f}_{2,n}(X_i),$$

where $W_i = (W_{i1}, W_{i2})$ is a discrete random vector, following a multivariate Bernoulli distribution with vector parameters $(\lambda, 1-\lambda)$. Using Bays's formula, we obtain the posterior probability in the t^{th} iteration denoted by

$$\overline{\tau}_{i1}^{(t)} = \frac{\tilde{f}_{1,n,m}(X_i)\lambda^{(t)}}{\lambda^{(t)}\tilde{f}_{1,n,m}(X_i) + (1-\lambda^{(t)})\tilde{f}_{2,n}(X_i)},$$

and

$$\overline{\tau}_{i2}^{(t)} = \frac{\tilde{f}_{2,n}(X_i)\lambda^{(t)}}{\lambda^{(t)}\tilde{f}_{1,n,m}(X_i) + (1-\lambda^{(t)})\tilde{f}_{2,n}(X_i)} = 1 - \overline{\tau}_{i1}^{(t)}.$$

2. *M-step*: It consists of a global maximization of $Q(\lambda|\lambda^{(t)})$ with respect to λ:

$$\lambda^{(t+1)} = \underset{\lambda}{\operatorname{argmax}}\, Q(\lambda \mid \lambda^{(t)}).$$

The updated estimate of λ is indicated by:

$$\lambda^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} \overline{\tau}_{i1}^{(t)}.$$

The estimation of λ is obtained from by iterating the EM algorithm until convergence:

$$\lim_{t\to\infty} \lambda^{(t)} = \widehat{\lambda}. \tag{3.2.3}$$

Therefore, the proposed estimator of the density function f is defined by

$$f_{n,m}^{Sh}(x) = \widehat{\lambda}\tilde{f}_{1,n,m}(x) + (1-\widehat{\lambda})\tilde{f}_{2,n}(x). \tag{3.2.4}$$

Basically, it is a *shrinkage* estimator that shrinks the Bernstein estimator towards the Gaussian Mixture density by a specified amount of λ. If $\lambda = 1$, the estimator $f_{n,m}^{Sh}$ reduces to the Bernstein estimator $\tilde{f}_{1,n,n}$.

3.3 Convergence

In this section, we derive some asymptotic properties of the proposed estimator $f_{n,m}^{Sh}$ when the sample size tends to infinity. First, we assume that λ and K are fixed. The following proposition gives the probability convergence of the proposed estimator $f_{n,m}^{Sh}$.

Proposition 3.1 (Probability convergence)**.** *If $m = o(n/\log(n))$, then for $x \in [0,1]$ we have*

$$f_{n,m}^{Sh}(x) \underset{n,m\to+\infty}{\xrightarrow{P}} \lambda f(x) + (1-\lambda) f_2(x),$$

where $f_2(x) = \sum_{j=1}^{K} \pi_j \mathcal{N}(\mu_j, \sigma_j^2)(x)$, $\pi_j = \mathbb{E}(Z_{1j})$, $\mu_j = \mathbb{E}(X_1 \mid Z_{1j} = 1)$, $\sigma_j^2 = Var(X_1 \mid Z_{1j} = 1)$ for $j = 1, \ldots, K$ and $\xrightarrow{P}$ denotes the convergence in probability.

The proof of Proposition 3.1 necessitates the following technical Lemma.

Lemma 3.1. *Let $(S_n)_{n\geq 1}$ be a sequence of i.i.d. random variables in the space of square integral functions L^2 with a common mean μ and let $(T_n)_{n\geq 1}$ be a sequence of random variables. Hence,*

$$\mathbb{E}(\overline{S}_n | T_n) \underset{n\to+\infty}{\xrightarrow{L^2}} \mu, \text{ where } \overline{S}_n = \frac{1}{n}\sum_{i=1}^{n} S_i,$$

where L^2 denotes the mean quadratic convergence L^2.

The proof of this lemma is reported in Zitouni et al (2018).

Proof: [Proof of Proposition 3.1] First, using Lemma 3.1 and following the same steps as the proof of Theorem 4.4 in Zitouni et al (2018), we prove that $\widehat{\pi}_j \underset{n\to+\infty}{\xrightarrow{P}} \pi_j$, $\lim_{n\to\infty} \widehat{\mu}_j \underset{n\to+\infty}{\xrightarrow{P}} \mu_j$, and $\widehat{\sigma^2}_j \underset{n\to+\infty}{\xrightarrow{P}} \sigma_j^2$. Then according to Slutsky's Theorem, we get

$$\sum_{k=1}^{K} \widehat{\pi}_k \mathcal{N}(\widehat{\mu}_k, \widehat{\sigma}_k)(x) \underset{n\to+\infty}{\xrightarrow{P}} \sum_{j=1}^{K} \pi_j \mathcal{N}(\mu_j, \sigma_j^2)(x). \tag{3.3.1}$$

Second, based on Theorem 3.1 in Babu et al. (2002), we obtain that

$$\tilde{f}_{1,n,m}(x) \underset{n\to+\infty}{\xrightarrow{P}} f(x) \text{ for } x \in [0,1]. \tag{3.3.2}$$

In addition, referring to (3.3.1) and (3.3.2) and grounded on the application of Slutsky's Theorem, we conclude the proof.

To study the asymptotic normality of the estimator $f_{n,m}^{Sh}$ given by (3.2.4), we set forward the following assumptions in Redner and Walker (1984)

(A1) For almost $x \in [0,1]$, and for all $i,j,h = 1\ldots,K$, the partial derivatives $\partial g/\partial\xi_i$, $\partial^2 g/\partial\xi_i\partial\xi_j$ and $\partial^3 g/\partial\xi_i\partial\xi_j\partial\xi_h$ of the density g exist and satisfy that $\left|\frac{\partial g(x|\theta)}{\partial\xi_i}\right|$, $\left|\frac{\partial^2 g(x|\theta)}{\partial\xi_i\xi_j}\right|$ and $\left|\frac{\partial^3 g(x|\theta)}{\partial\xi_i\xi_j\xi_h}\right|$ are bounded respectively by J_i, J_{ij} J_{ijh}, where J_i and J_{ij} are integrable, and J_{ijh}, satisfies

$$\int_0^1 J_{ijh}(x)g(x|\widehat{\theta})dx < \infty,$$

and $(\xi_1,\ldots,\xi_\nu) = (\pi_1,\ldots,\pi_K,\mu_1\ldots,\mu_K,\sigma_1,\ldots,\sigma_K)$.

(A2) The Fisher information matrix $I(\theta)$ is positively defined at $\widehat{\theta}$.

Proposition 3.2 (Normality asymptotic)**.** *Under the regularity conditions* ***(A1)-(A2)****, if $f(x) > 0$ for all $x \in [0,1]$, $2 \le m \le (n/\log n)$ and $\lim_{n,m\to\infty} n^{2/3}/m = 0$, then we get*

$$n^{1/2}m^{-1/4}\left[f_{n,m}^{Sh}(x) - \lambda f(x) - (1-\lambda)f_2(x)\right] \xrightarrow[n,m\to+\infty]{\mathcal{D}} \mathcal{N}\left(0,\lambda^2\gamma(x)\right),$$

where $\gamma(x) = f(x)(4\pi x(1-x))^{-1/2}$, for $x \in]0,1[$ and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.

Proof: [Proof of Proposition 3.2] Using Theorem 3.2 in Babu et al. (2002), we obtain

$$n^{1/2}m^{-1/4}(\tilde{f}_{1,n,m}(x) - f(x)) \xrightarrow[n,m\to+\infty]{\mathcal{D}} \mathcal{N}(0,\gamma(x)).$$

Thus

$$n^{1/2}m^{-1/4}(\lambda\tilde{f}_{1,n,m}(x) - \lambda f(x)) \xrightarrow[n,m\to+\infty]{\mathcal{D}} \mathcal{N}(0,\lambda^2\gamma(x)).$$

According to Theorem 3.1 in Redner and Walker (1984), we obtain $\sqrt{n}(\widehat{\theta}-\theta) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,I(\theta)^{-1})$. Using delta method, we get

$$\sqrt{n}(\tilde{f}_{2,n}(x|\widehat{\theta}) - f_2(x|\theta)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, Df_2(x|\theta)I(\theta)^{-1}Df_2(x|\theta)^T\right),$$

with $Df_2(x|\theta)$ is the Jacobian matrix of $f_2(x|\theta) = f_2(x)$ and $\tilde{f}_{2,n}(x|\widehat{\theta}) = \tilde{f}_{2,n}(x)$. Since $m^{-1/4} \to 0$ if $m \to \infty$, then using Slutsky's Theorem, we conclude the proof.

The following corollary is a consequence of the previous proposition which gives an asymptotic confidence interval of the density f, for a risk $\alpha \in]0,1[$.

Corollary 3.3.1. *The* $100(1-\alpha)\%$ *asymptotic confidence interval of* $f(x)$ *is given by*

$$\left(f_{n,m}^{Sh} \pm \frac{z_{1-\frac{\alpha}{2}} \widehat{\lambda} \sqrt{\gamma(x)}}{\sqrt{n} m^{-1/4}} \right),$$

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

In the next section, we study the performance of the proposed estimator in estimating different distributions by comparing it to the performances of the Bernstein estimator and of the Gaussian kernel estimator.

3.4 Numerical studies

3.4.1 Comparison study

In this section, we investigate the performance of the proposed estimator given in (3.2.4), through estimating different densities by comparing it to the performances of Bernstein density estimator defined by $\tilde{f}_{n,m}(x) = m \sum_{k=1}^{m-1} \left[\overline{F}_n \left(\frac{k+1}{m} \right) - \overline{F}_n \left(\frac{k}{m} \right) \right] b_k(m-1, x)$ and of the standard Gaussian kernel estimator defined by $\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K \left(\frac{x - X_i}{h_n} \right)$. We apply the Bernstein estimator when the sample is concentrated on the interval $[0,1]$. For this purpose, we need to make some suitable transformations in different cases that are listed below:

1. Suppose that X is concentrated on a finite support $[a,b]$, then we work with the sample values $Y_1, \ldots, Y_n$ where $Y_i = (X_i - a)/(b-a)$.

2. For the density functions concentrated on $\mathbb{R}$, we can use the transformed sample $Y_i = 1/2 + \pi^{-1} \arctan(X_i)$ which transforms the range to the interval $[0,1]$.

3. For the support $\mathbb{R}_+$, we can use the transformed sample $Y_i = X_i/(1+X_i)$, which transforms the range to the interval $[0,1]$.

In the simulation study, three sample sizes are considered, $n = 50$, $n = 100$, and $n = 200$, as well as the following density functions:

(a) the beta mixture density $0.5\mathcal{B}(3,9) + 0.5\mathcal{B}(9,3)$.

(b) the beta mixture density $0.5\mathcal{B}(3,1) + 0.5\mathcal{B}(10,10)$.

(c) the normal mixture density $1/4\mathcal{N}(2,1) + 3/4\mathcal{N}(-3,1)$.

(d) the Chi-square $\chi_n(2)$ density.

(e) the gamma mixture density $0.5\mathcal{G}(1,6) + 0.5\mathcal{G}(6,1)$.

(f) the gamma mixture density $0.5\mathcal{G}(1,2) + 0.5\mathcal{G}(4,2)$.

Our sample will be decomposed into a learning sample of size $2/3$ of the considered sample on which the various statistical methods are constructed and a second sample of size $1/3$ of the considered sample on which the predictive performances of the three methods are tested. For each density function f and sample size n, we compute the Integrated Squared Error (ISE), the integrated absolute error (IAE) and the Kullback-Leibler divergence (KL) of the estimator $f_{n,m}^{Sh}$ over $N = 500$ trials,

$$\widehat{ISE} = \frac{1}{N}\sum_{k=1}^{N} ISE(\widehat{f}_k),\ \widehat{IAE} = \frac{1}{N}\sum_{k=1}^{N} IAE(\widehat{f}_k) \text{ and } \widehat{KL} = \frac{1}{N}\sum_{k=1}^{N} KL(\widehat{f}_k),$$

where $\widehat{f}_k$ is the estimator computed from the k^{th} sample, and

$$ISE[\widehat{f}_k] = \int_0^1 \left(\widehat{f}_k(x) - f(x)\right)^2 dx,\ IAE(\widehat{f}_k) = \int_0^1 |\widehat{f}_k(x) - f(x)|dx,$$

$$KL(\widehat{f}_k|f) = \int_0^1 \widehat{f}_k(x) \log \frac{\widehat{f}_k(x)}{f(x)} dx.$$

To select the optimal parameter K, we use the Gap Statistics algorithm. In each case, we approximated the ISE of the Bernstein estimator (3.2.1), the proposed estimator (3.2.4) (for integers $1 \leq m \leq 300$) and the kernel estimator (3.1.1) (for $h = i/1000$ with $1 \leq i \leq 300$) using $N = 500$ random samples of sizes $n = 50$, $n = 100$ and $n = 200$.

Table 3.1: ISE for $N = 500$ trials of $\tilde{f}_{n,m}$, $\widehat{f}_n$ and $f_{n,m}^{Sh}$, for $n = 50, 100, 200$.
[ISE for $N = 500$ trials of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $f_{n,m}^{Sh}$, for $n = 50$, $n = 100$ and $n = 200$. The bold values indicate the smallest values of ISE.]

	n	m_{opt}	Proposed estimator	Bernstein estimator	kernel estimator
$0.5\mathcal{B}(3,9) + 0.5\mathcal{B}(9,3)$	50	34	**0.086288**	0.101725	0.197497
	100	63	**0.157977**	0.158098	0.174251
	200	97	0.141708	**0.136258**	0.148143
$0.5\mathcal{B}(3,1) + 0.5\mathcal{B}(10,10)$	50	35	1.822368	1.824329	**0.482152**
	100	61	**0.274204**	0.626057	0.530446
	200	100	**0.373920**	0.623460	0.474805
$1/4\mathcal{N}(2,1) + 3/4\mathcal{N}(-3,1)$	50	5	**0.356460**	1.521631	1.077352
	100	2	**0.254651**	1.084614	1.641689
	200	4	**0.262561**	1.117986	2.222369
$\chi_n(2)$	50	2	**0.162917**	0.898103	1.154646
	100	3	**0.492752**	2.483141	2.331765
	200	3	**0.525192**	2.812448	4.936701
$0.5\mathcal{G}(1,6) + 0.5\mathcal{G}(6,1)$	50	21	2.323582	2.599131	**2.295932**
	100	22	**0.925537**	2.718903	2.424656
	200	39	**0.846773**	2.137174	2.053453
$0.5\mathcal{G}(1,2) + 0.5\mathcal{G}(4,2)$	50	4	**0.313388**	0.656400	0.417980
	100	5	**0.186290**	0.577408	0.762742
	200	8	**0.253988**	0.896995	1.397111

Departing from Tables 6.1, 6.2 and 6.3, we deduce that,

- For the case (b) of the beta mixture, the average KL of the kernel density estimator. is smaller than that obtained by the proposed density estimator (3.2.4) and the Bernstein estimator. However, in all the other cases, using an appropriate choice of the degree m, the average KL of the density estimator (3.2.4) is smaller than that achieved by the kernel estimator estimator and the Bernstein estimator even when the sample size is large for same cases.

- Almost in all considered cases, the averages ISE and IAE of the density estimator (3.2.4) is smaller than those obtained by the Bernstein estimator and those of the kernel estimator.

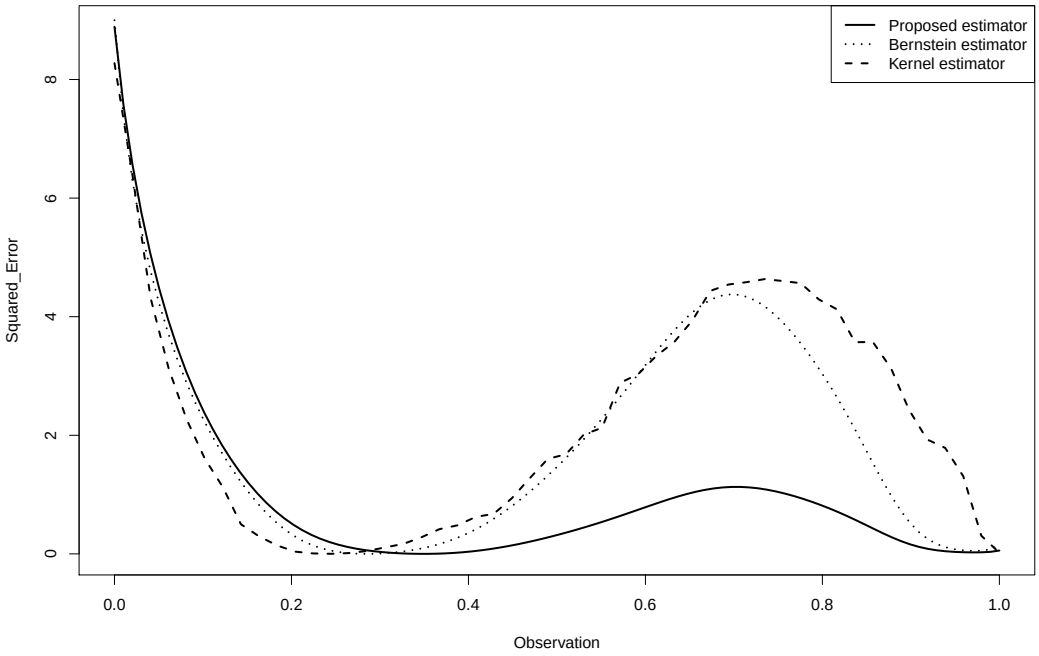

Figure 3.1: ISE for $\widehat{f}_n$, $\tilde{f}_{n,m}$ and $f_{n,m}^{SH}$

[ISE of the kernel estimator $\widehat{f}_n$, of the Bernstein estimator $\tilde{f}_{n,m}$ and of the proposed density estimator $f_{n,m}^{SH}$ of $0.5\mathcal{G}(6,1) + 0.5\mathcal{G}(1,6)$ for $n = 200$]

Table 3.2: IAE for $N = 500$ trials of $\tilde{f}_{n,m}$, $\widehat{f}_n$ and $f_{n,m}^{Sh}$, for $n = 50, 100, 200$.
[IAE for $N = 500$ trials of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $f_{n,m}^{Sh}$, for $n = 50$, $n = 100$ and $n = 200$. The bold values indicate the smallest values of IAE.]

	n	Proposed estimator	Bernstein estimator	kernel estimator
$0.5\mathcal{B}(3,9) + 0.5\mathcal{B}(9,3)$	50	**0.180945**	0.216329	0.391072
	100	**0.327796**	0.327927	0.367361
	200	0.318701	**0.299018**	0.348499
$0.5\mathcal{B}(3,1) + 0.5\mathcal{B}(10,10)$	50	0.884591	0.884942	**0.621137**
	100	**0.435256**	0.713915	0.669027
	200	**0.504220**	0.668169	0.583057
$1/4\mathcal{N}(2,1) + 3/4\mathcal{N}(-3,1)$	50	**0.575782**	1.190177	0.953675
	100	**0.487738**	1.006137	1.157838
	200	**0.498120**	1.028269	1.238839
$\chi_n(2)$	50	**0.351205**	0.807667	0.881866
	100	**0.657815**	1.486974	1.415036
	200	**0.669269**	1.573886	1.942776
$0.5\mathcal{G}(1,6) + 0.5\mathcal{G}(6,1)$	50	**1.326117**	1.481910	1.374609
	100	**0.862592**	1.599692	1.499828
	200	**0.718135**	1.249224	1.225502
$0.5\mathcal{G}(1,2) + 0.5\mathcal{G}(4,2)$	50	**0.414676**	0.640596	0.516527
	100	**0.383248**	0.714844	0.782939
	200	**0.453948**	0.883320	1.068483

Referring to Figure (3.4.1), we infer that all the estimators for the gamma mixture density $0.5\mathcal{G}(6,1) + 0.5\mathcal{G}(1,6)$ do not have good approximations near $x = 0$. However, the ISE of the proposed estimator is closer to zero compared to the Bernstein estimator and the kernel estimator, specially near the edge $x = 1$.

In addition, the performed simulations reveal that, on average, the proposed approach can lead to satisfactory estimates near the boundaries, better than the classical Bernstein estimator.

Table 3.3: KL for $N = 500$ trials of $\tilde{f}_{n,m}$, $\widehat{f}_n$ and $f_{n,m}^{Sh}$, for $n = 50, 100, 200$.
[KL for $N = 500$ trials of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $f_{n,m}^{Sh}$, for $n = 50$, $n = 100$ and $n = 200$. The bold values indicate the smallest values of KL.]

	n	Proposed estimator	Bernstein estimator	kernel estimator
$0.5\mathcal{B}(3,9) + 0.5\mathcal{B}(9,3)$	50	**0.111333**	0.107221	0.289818
	100	0.114439	0.115192	**0.081830**
	200	**0.018268**	0.030360	0.060468
$0.5\mathcal{B}(3,1) + 0.5\mathcal{B}(10,10)$	50	1.350244	1.351966	**0.150096**
	100	0.515686	0.555641	**0.381406**
	200	0.837851	1.097354	**0.575298**
$1/4\mathcal{N}(2,1) + 3/4\mathcal{N}(-3,1)$	50	**2.662741**	3.352308	3.169852
	100	**2.655629**	3.341980	4.196295
	200	**2.689781**	3.377007	4.732324
$\chi_n(2)$	50	**0.281862**	0.870222	1.537355
	100	**0.976450**	1.572251	3.031783
	200	**0.960633**	1.584022	4.702359
$0.5\mathcal{G}(1,6) + 0.5\mathcal{G}(6,1)$	50	0.177888	**0.130495**	1.322299
	100	**2.970688**	3.914360	5.538142
	200	**1.332679**	2.080956	3.498273
$0.5\mathcal{G}(1,2) + 0.5\mathcal{G}(4,2)$	50	**0.062893**	0.589790	0.679154
	100	**0.292017**	0.805962	1.191070
	200	**0.528337**	1.052789	1.952294

3.4.2 Real dataset

Old faithful data

In this subsection, we consider the well known Old Faithful data displayed in Table 2.2 of Silverman (1986). These data concern the eruption lengths (in minutes) of 107 eruptions of the Old Faithful geyser in Yellowstone National Park, U.S.A. These data are such that $\min_i(x_i) = 1.67$ and $\max_i(x_i) = 4.93$. Then, it is convenient to assume that the density of eruption times is defined on the interval $[1.5, 5]$ and transform the data into the interval unit. The Monte Carlo procedure was performed and resulted in $h = 0.1327$ for the standard kernel estimator, $m = 60$ for the Bernstein estimator and the proposed estimator.

These estimators are exhibited in Figure 1 (right panel) along with an histogram of the data. All the estimators are smooth and seem to capture the pattern highlighted by the histogram. We record that the proposed estimator outperforms the other estimators near the boundaries.

Tuna data

The last example concerns the tuna data reported in Chen (1996). The data are derived from an aerial line transect survey of Southern Bluefin Tuna in the Great Australian Bight. An aircraft with two spotters on board flies randomly over allocated line transects. These data correspond to the perpendicular sighting distances (in miles) of 64 detected tuna schools to the transect lines. The survey was conducted in summer when tuna data tend to stay on the surface. The data are such that $\min_i(x_i) = 0.19$ and $\max_i(x_i) = 16.26$. The Monte Carlo procedure was performed and resulted in $h = 0.1079$ for the standard kernel estimator, $m = 13$ for the Bernstein estimator and the proposed estimator. These estimators are illustrated in Figure 3.4.2 (left panal) along with an histogram of the data. All the estimators are smooth and seem to capture the pattern highlighted by the histogram. We assert that the proposed estimator and the Bernstein estimator outperform the standard kernel estimator near the boundaries.

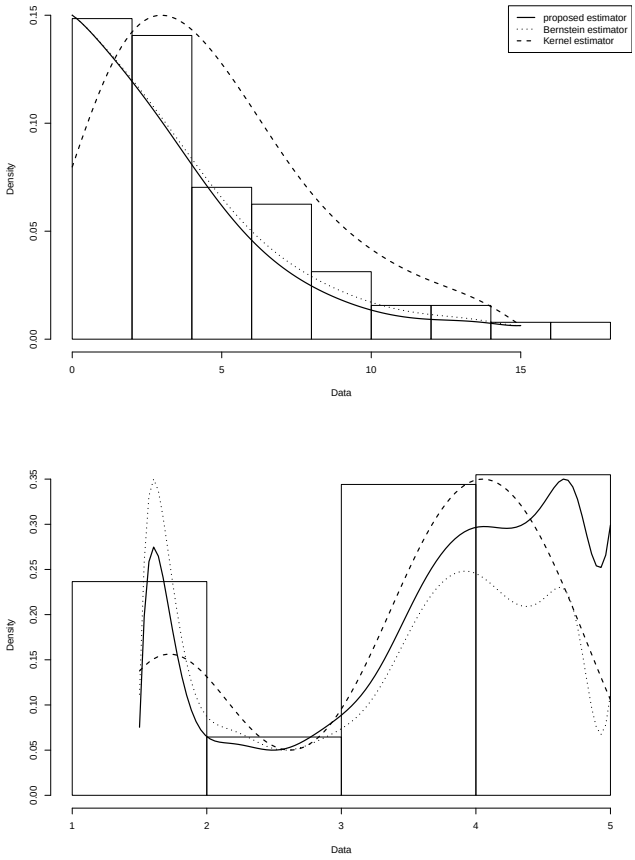

Figure 3.2: Quantitative comparison between the MSE of $\widehat{f}_n$, of $\tilde{f}_{n,m}$ and of $f^{SH}_{n,m}$. [Quantitative comparison between the Mean Squared Error of the kernel estimator $\widehat{f}_n$, of the Bernstein estimator $\tilde{f}_{n,m}$ and of the proposed density estimator $f^{SH}_{n,m}$, of Tuna data (left panel) and of Old Faithful data (right panel).]

3.5 Conclusion

In this chapter, we proposed a *shrinkage* estimator of a density function based on the Bernstein density estimator and using a finite Mixture Gaussian density. This method rests on three steps. The first step consists of considering the Bernstein estimator $\tilde{f}_{1,n,m}$. The second relies upon the Gaussian Mixture density $\tilde{f}_{2,n}$ as an estimator of the unknown density f. The last step consists of considering the shrinkage form $\lambda\tilde{f}_{1,n,m} + (1-\lambda)\tilde{f}_{2,n}$ and EM algorithm in order to estimate the parameter λ. The asymptotic properties of this estimator were established. Afterwards, we demonstrated the effectiveness of the

proposed method using some simulated and real data. We clarified how it can lead to very satisfactory estimates near the boundaries and in terms of ISE, IAE and KL. Eventually, we would simply assert that our research work is a step that may be taken further, extended and built upon as it lays the ground and paves the way for future works to elaborate a semi parametric regression estimator using the shrinkage method. We also plan to work on the case where λ is a random variable.

Chapter 4

Recursive regression estimation based on a stochastic algorithm

Key words and phrases: Bernstein polynomials; Regression estimation; Two-time-scale stochastic approximation algorithms.

Abstract 2: *In this chapter, we propose a recursive estimators of the regression function based on the two-time-scale stochastic approximation algorithms and the Bernstein polynomials. We study the asymptotic properties of this estimators. We compare the proposed estimators with the classic regression estimator using the Bernstein polynomial defined by Tenbusch. Results showed that, our proposed recursive estimators can overcome the problem of the edges associated with kernel regression estimation with a compact support. The proposed recursive two-time-scale estimators are compared to the non recursive estimator introduced by Tenbusch and the performance of the two estimators are illustrated via simulations as well as two real datasets.*

4.1 Introduction

Nonparametric regression estimation methods have attracted a great deal of attention as researchers have realized that parametric regression is not suitable for adequately fitting curves to many datasets that arise in practice. There are many reasons for choosing the nonparametric regression methods, no assumption should be made on the form of the regression function, the complexity of the model will be determined completely by the data, it is applicable for various design situations and it is easy to interpret.

There have been several monograph on the nonparametric regression estimation, see for instance Eubank (1988), Müller (1988), Härdle (1990), Hastie and Tishirani (1990), where it is shown that nonparametric regression techniques have much to offer in applications, such as observational astronomy, forecasting future opportunities and risks in business, causal relationships between parameters in biological systems and various other situations.

The most famous nonparametric estimator of the regression function $r : x \to \mathbb{E}(Y|X = x)$, was proposed by Nadaraya (1964) and Watson (1964). In the current work, we are concerned by the recursive estimation of a regression function. Recursive estimation have become an increasingly important area of research. In many situations, data arrives regularly so that it is impossible to store them in a traditional database. In such a context, building a recursive estimator which does not require to store all the data in memory is of great interest.

The recursive regression estimation was introduced first by Kiefer and Wolfowitz (1952), extended by Révész (1973) using the Robbins-Monro's procedure, generalized by Mokkadem et al (2009) and studied by Slaoui (2015a,b,c). The semi-recursive approach was introduced by Slaoui (2016), while the recursive regression estimation for independent functional data was established by Slaoui (2019, 2020).

However, kernel methods and their recursive improvements have estimation problems in the edges when the regression function has a bounded support. To overcome this problem, their have been many methods such as the approach of regression estimation using Bernstein polynomials proposed by Tenbusch (1997).

Bernstein polynomial was used in several directions to estimate a density of probability and distribution functions. See for instance, the original work of Vitale (1975), extended by Tenbusch (1994), Ghosal (2000), Babu et al. (2002), Kakizawa (2004), Kakizawa (2011), Rao (2005), Leblanc (2010, 2012a,b), Igarashi and Kakizawa (2014), Jmaei et al. (2017) and Slaoui and Jmaei (2019).

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be *i.i.d.* pairs of random variables with joints density function $g(x, y)$ and let f such that $f(x) = \int_{\mathbb{R}} g(x, t)dt$, denote the density of X which is supported on $[0, 1]$. We denote by $a(x) = f(x)r(x)$ and let $\mathbb{E}(|Y|) < \infty$. The

non recursive estimator defined by Tenbusch (1997) such that $f(x) \neq 0$, for $x \in [0,1]$

$$\tilde{r}_{n,m}(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} Y_i \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x)}{\sum_{i=1}^{n} \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x)} & \text{if } \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x) \neq 0, \\ 0 & \text{if } \sum_{k=1}^{m} \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1,x) = 0, \end{cases} \tag{4.1.1}$$

where $b_k(m,x)$ is the Bernstein polynomial of order m.

The aim of this chapter is to introduce two-time-scale stochastic approximation algorithms in order to define a class of recursive estimators of a regression function based on Bernstein polynomials. The two-time-scale stochastic approximation algorithms have been defined by Borkar (1997), Konda and Borkar (1999), Bhatnagar et al. (2001), their convergence rate studied by Konda and Tsitsiklis (1999), Mokkadem and Pelletier (2006a) and Slaoui (2006).

In order to construct a recursive regression estimator defined by the two-time-scale stochastic algorithm, at a point x such as $f(x) \neq 0$, we define an algorithm of search of the common zero of the functions:

$$l_1 : (y,z) \mapsto f(x) - y \text{ and } l_2 : (y,z) \mapsto \frac{r(x)f(x)}{y} - z.$$

We proceed in the following way, for $x \in [0,1]$: *i)* $f_0(x) > 0$ and $r_0(x) \in \mathbb{R}$, *ii)* for $n \geq 1$

$$\begin{cases} f_n(x) = f_{n-1}(x) + \gamma_n W_n^{(1)}(x) \\ r_n(x) = r_{n-1}(x) + \beta_n W_n^{(2)}(x) \end{cases}$$

where $W_n^{(1)}$ and $W_n^{(2)}$ are the observations of the functions l_1 and l_2 at the point $(f_{n-1}(x), r_{n-1}(x))$, and where the stepsizes (γ_n) and (β_n) are two sequences of positive real numbers that go to zero such that $\gamma_n \leq 1$ and $\lim_{n\to\infty} \beta_n \gamma_n^{-1} = 0$. Then, we estimate $f(x)$ and $a(x)$ respectively by

$$m \sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1,x) \text{ and } mY_n \sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1,x).$$

Then, the two-time-scale stochastic approximation algorithm can be rewritten as:

$$\begin{cases} f_n(x) &= (1-\gamma_n) f_{n-1}(x) + \gamma_n m \sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x), \\ r_n(x) &= (1-\beta_n) r_{n-1}(x) + \frac{\beta_n m Y_n}{f_{n-1}(x)} \sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x). \end{cases} \tag{4.1.2}$$

The aim of this chapter is to study the properties of the regression estimators defined in (4.1.2), as a competitor for the non recursive estimator (4.1.1).

The remainder of this chapter is organized as follows. In the next section, we present the assumptions and notations that we need to give the proprieties of our proposed two-time-scale estimators (4.1.2). In Section 6.3, we state our main results. Section 6.4 is devoted to give some numerical comparison between our proposed recursive estimators (4.1.2) and the non recursive estimtaor 6.3 through some simulation studies and then through two real datasets. Some concluding remarks and possible future developments are mentioned in Section 4.6. While, all the mathematical developments are deferred to the Section 6.6.

4.2 Assumptions and notations

We define the following class of regularly varying sequences.

Definition 4.1. *Let $v \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(v)$ if*

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = v.$$

This condition was introduced by Galambos and Seneta (1973) to define regularly varying sequences. To study asymptotic the behaviours of the estimator r_n defined by equations (4.1.2) inside the interval $[0,1]$, we consider the following assumptions for $a \in (0, 1/3)$:

(**A1**) (i) $(m_n) \in \mathcal{GS}(a)$ with $a \in (0,1)$.

(ii) $(\beta_n) \in \mathcal{GS}(-\beta)$ with $\beta \in (3a, 1]$.

(iii) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in (\min\{3a, a+1-\beta\}, \beta]$, such that

$$\lim_{n\to\infty} \beta_n^{-1}\gamma_n \left(\ln \sum_{k=1}^{n} \gamma_k\right)^{-1} = \infty.$$

(iii) $\lim_{n\to\infty} (n\beta_n) > \min\{a, (\beta - a/2)/2\}\mathbb{1}_{\{x\in(0,1)\}} + \min\{a, (\beta-a)/2\}\mathbb{1}_{\{x\in\{0,1\}\}}$.

(**A2**) (*i*) $v \to g(v,w)$ is continuous on $\mathbb{R}$.

(*ii*) For $t>0$, $v \to \int_{\mathbb{R}} |w|^t g(v,w)dw$ is a bounded function.

Discussion of the assumptions

1. Assumptions $(A1)$ and $(A2)$ are regularity conditions which permit us to evaluate the bias term and the variance term of the proposed estimator.

2. The intuition behind the use of such bandwidth (h_n) belonging to $\mathcal{GS}(-a)$ is that the ratio h_{n-1}/h_n is equal to $1+a/n+o(1/n)$, similarly, we use the stepsize (β_n) belonging to $\mathcal{GS}(-\beta)$ then the ratio β_{n-1}/β_n is equal to $1+\beta/n+o(1/n)$, the application of Lemma 6.1 ensures that the bias and the variance will depend only on h_n and β_n and not on $h_1,\ldots,h_n$ and $\beta_1,\ldots,\beta_n$ then the $MISE$ will depend only on h_n and β_n, which will be helpful to deduce an optimal bandwidth and an optimal stepsize.

3. In order to help the readers to follow the main results obtained in this chapter, we underline that the application of Lemma 6.1 under the assumption $(A2)$ ensures, $\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k = 1+o(1)$, $\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k h_k^2 = O(h_n^2)$ and $\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 h_k^{-1} = O\left(\beta_n h_h^{-1}\right)$.

4. Assumption $(A2)(iii)$ on the limit of $(n\beta_n)$ as n goes to infinity is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\beta_n]^{-1}\right)$ is finite.

Throughout this chapter, we will use the following notations:

$$
\begin{aligned}
\xi &= \lim_{n\to\infty} (n\beta_n)^{-1},\\
\psi(x) &= (4\pi x(1-x))^{-1/2},\\
\mathcal{W}_n(x) &= m_n Y_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\{\frac{k}{m_n}<X_n\leq\frac{k+1}{m_n}\}} b_k(m_n-1,x),\\
\mathcal{Z}_n(x) &= m_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\{\frac{k}{m_n}<X_n\leq\frac{k+1}{m_n}\}} b_k(m_n-1,x),\\
\Delta_1(x) &= \frac{1}{2}[(1-2x)f'(x)+x(1-x)f^2(x)],\\
\Delta_2(x) &= \frac{1}{2}\left[(1-2x)(r'(x)f(x)-r(x)f'(x))\right.\\
&\quad \left.+x(1-x)\left(r^{(2)}(x)f(x)+r(x)f^{(2)}(x)+2f'(x)r'(x)\right)\right],
\end{aligned}
$$

$$\begin{aligned}\mathcal{B}ias\left(x\right) &= \frac{\left(\Delta_2\left(x\right)-r\left(x\right)\Delta_1\left(x\right)\right)}{f\left(x\right)},\\ \mathcal{V}\left(x\right) &= \frac{\mathrm{Var}\left[Y|X=x\right]\psi\left(x\right)}{f\left(x\right)\left(2-\left(\beta-a/2\right)\xi\right)}\mathbb{1}_{\{x\in(0,1)\}}+\frac{\mathrm{Var}\left[Y|X=x\right]}{f\left(x\right)\left(2-\left(\beta-a\right)\xi\right)}\mathbb{1}_{\{x\in\{0,1\}\}}.\end{aligned}$$

4.3 Main results

Throughout this chapter we consider the two following sequences:

$(v_n)=\left(m_n^{1/2}\mathbb{1}_{\{x\in(0,1)\}}+m_n\mathbb{1}_{\{x\in\{0,1\}\}}\right)$ and $(\mathcal{V}_n)=\left(\beta_n^{-1/2}v_n^{-1}\right)$.

For simplicity, we let

$$\begin{aligned}\mathcal{A}_{a,x}^1 &= \left\{a,x|a\in(0,\frac{\beta}{3}],x\in\{0,1\}\right\},\quad \mathcal{A}_{a,x}^2=\left\{a,x|a\in(0,\frac{2}{5}\beta],x\in(0,1)\right\},\\ \mathcal{A}_{a,x}^3 &= \left\{a,x|a\in(\frac{\beta}{3},1),x\in\{0,1\}\right\},\quad \mathcal{A}_{a,x}^4=\left\{a,x|a\in(\frac{2}{5}\beta,1),x\in(0,1)\right\}.\end{aligned}$$

Our first result is the following proposition, which gives the bias and the variance of the proposed recursive estimator r_n.

Proposition 4.1 (Bias and Variance of r_n).

1. *Let Assumptions* $(\mathbf{A1})-(\mathbf{A2})$ *hold such as* $f(x)>0$. *Then*

$$\begin{aligned}\mathbb{E}\left[r_n(x)\right]-r\left(x\right) &= m_n^{-1}\mathcal{B}ias\left(x\right)\frac{1}{1-a\xi}\mathbb{1}_{\mathcal{A}_{a,x}^1\bigcup\mathcal{A}_{a,x}^2}\left[1+o\left(1\right)\right]\\ &\quad+\mathbb{1}_{\mathcal{A}_{a,x}^4}o\left(\sqrt{\beta_n m_n^{1/2}}\right)+\mathbb{1}_{\mathcal{A}_{a,x}^3}o\left(\sqrt{\beta_n m_n}\right).\end{aligned}\tag{4.3.1}$$

$$\mathrm{Var}\left[r_n(x)\right]\tag{4.3.2}$$

$$\begin{aligned}&= \beta_n m_n^{1/2}\frac{\mathrm{Var}\left[Y|X=x\right]\psi\left(x\right)}{2-\left(\beta-a/2\right)\xi}\mathbb{1}_{\mathcal{A}_{a,x}^4}\left[1+o_x\left(1\right)\right]\\ &\quad+\beta_n m_n\frac{\mathrm{Var}\left[Y|X=x\right]}{2-\left(\beta-a\right)\xi}\mathbb{1}_{\mathcal{A}_{a,x}^3}\left[1+o\left(1\right)\right]+\mathbb{1}_{\mathcal{A}_{a,x}^1\bigcup\mathcal{A}_{a,x}^2}o\left(m_n^{-2}\right).\end{aligned}\tag{4.3.3}$$

The following proposition gives the mean squared error (MSE) of the proposed two-time-scale recursive estimators r_n.

Proposition 4.2 (MSE of r_n). *Let Assumptions* $(\mathbf{A1})-(\mathbf{A2})$ *hold. Then*

$$\mathrm{MSE}\left[r_n(x)\right]\tag{4.3.4}$$

$$
\begin{aligned}
= \; & m_n^{-2}\frac{\mathcal{B}ias^2(x)}{(1-a\xi)^2}\mathbb{1}_{\mathcal{A}^1_{a,x}\bigcup\mathcal{A}^2_{a,x}}\left[1+o(1)\right]+\beta_n m_n^{1/2}\frac{\mathrm{Var}\left[Y|X=x\right]\psi(x)}{2-(\beta-a/2)\xi}\mathbb{1}_{\mathcal{A}^4_{a,x}}\left[1+o_x(1)\right] \\
& +\beta_n m_n\frac{\mathrm{Var}\left[Y|X=x\right]}{2-(\beta-a)\xi}\mathbb{1}_{\mathcal{A}^3_{a,x}}\left[1+o(1)\right]+\mathbb{1}_{\mathcal{A}^1_{a,x}\bigcup\mathcal{A}^2_{a,x}}o\left(m_n^{-2}\right)+\mathbb{1}_{\mathcal{A}^4_{a,x}}o\left(\beta_n m_n^{1/2}\right) \\
& +\mathbb{1}_{\mathcal{A}^3_{a,x}}o\left(\beta_n m_n\right). \qquad (4.3.5)
\end{aligned}
$$

To minimize the MSE of $r_n(x)$ for $x\in(0,1)$ such that $\psi(x)>0$, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$ and (m_n) must be in $\mathcal{GS}(2/5)$. To minimize the MSE of $r_n(x)$ for $x\in\{0,1\}$ such that $\psi(x)>0$, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$ and (m_n) must be in $\mathcal{GS}(1/3)$.

Corollary 4.3.1. *Let Assumptions* (**A1**)-(**A2**) *hold. To minimize the* MSE *of the proposed two-time-scale recursive estimators $r_n(x)$, the stepsize (γ_n) must be chosen equal to (n^{-1}). The optimal order (m_n) should be equal to*

$$
\begin{aligned}
& 2^{6/5}\left(\frac{5}{3}\right)^{2/5}\left(\frac{\mathcal{B}ias^{4/5}(x)}{\psi(x)^{2/5}\,\mathrm{Var}\left[Y|X=x\right]^{2/5}}\right)n^{2/5}\mathbb{1}_{\{x\in(0,1)\}} \\
+ \; & 6^{1/3}\left(\frac{\mathcal{B}ias^{2/3}(x)}{\mathrm{Var}\left[Y|X=x\right]^{1/3}}\right)n^{1/3}\mathbb{1}_{\{x\in\{0,1\}\}},
\end{aligned}
$$

and the corresponding MSE *must be equal to*

$$
\begin{aligned}
\mathrm{MSE}\left[\widehat{r}_n(x)\right] = \; & \frac{5}{4}2^{-2/5}\left(\frac{5}{3}\right)^{6/5}\psi(x)^{4/5}\,\mathrm{Var}\left[Y|X=x\right]^{4/5}\mathcal{B}ias^{2/5}(x)\,n^{-4/5}\left[1+o(1)\right]\mathbb{1}_{\{x\in(0,1)\}} \\
& +\frac{3}{3}2^{-1/3}\left(\frac{3}{2}\right)^{4/3}\mathrm{Var}\left[Y|X=x\right]^{2/3}\mathcal{B}ias^{2/3}(x)\,n^{-2/3}\left[1+o(1)\right]\mathbb{1}_{\{x\in\{0,1\}\}}.
\end{aligned}
$$

Finally, the following proposition shows the asymptotic normality of the recursive estimators r_n.

Theorem 4.3.1 (**Asymptotic normality of r_n**)**.**

1. *In the case when $x\in(0,1)$, if there exists $c\geq 0$ such that $\beta_n^{-1}m_n^{-5/2}\to c$, (resp. the case when $x\in\{0,1\}$, if $\beta_n^{-1}m_n^{-3}\to c$), then*

$$
\mathcal{V}_n\left(r_n(x)-r(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(\frac{\sqrt{c}}{(1-a\xi)}\mathcal{B}ias(x),\mathcal{V}(x)\right).
$$

2. *In the case when $x\in(0,1)$, if $\beta_n^{-1}m_n^{-5/2}\to\infty$, (resp. the case when $x\in\{0,1\}$, if $\beta_n^{-1}m_n^{-3}\to\infty$), then*

$$
m_n\left(r_n(x)-r(x)\right)\xrightarrow{\mathbb{P}}\frac{1}{(1-a\xi)}\mathcal{B}ias(x).
$$

4.4 Numerical studies

The aim of this subsection is to compare our proposed two-time-scale estimators (4.1.2) with the non recursive estimator (4.1.1), through a simulation. We consider the regression model

$$Y = r(X) + \varepsilon,$$

where $X \sim \mathcal{N}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,\sigma^2)$, where $\sigma \in \{0.05, 0.1, 1\}$. When using our proposed two-time-scale algorithm (4.1.2), the stepsizes (γ_n, β_n) are chosen to be equal to $(n^{-0.98}, 0.31n^{-0.99})$. Moreover, in order to select the smoothing parameter (m_n), we consider a Monte Carlo procedure for each point $x \in [0,1]$. We determine the parameter m for $1 \leq m \leq 400$ by minimizing

$$\frac{1}{N}\sum_{i=1}^{N}[r_{n,m}^{i}(x) - r(x)]^2,$$

with $r_{n,m}^{i}$ is the estimator of $r_{n,m}$ computed from the ith sample of size n. We choose $N = 500$ trials for the Monte Carlo simulation. In our simulation study, we consider three sample sizes; $n = 50$, $n = 100$, $n = 200$ and the following regression functions:

(a) $r(x) = \cos(x)$.

(b) $r(x) = (1 + \exp(x))^{-1}$.

(c) $r(x) = 1 + 0.6x$.

For each model and sample size n, we approximate the average integrated squared error $(\overline{ISE})$ and the integrated absolute error $(\overline{IAE})$ of the estimator using $N = 500$ trials of sample size n:

$$\overline{ISE} = \frac{1}{N}\sum_{k=1}^{N} ISE[\overline{r_k}],\ \overline{IAE} = \frac{1}{N}\sum_{k=1}^{N} IAE[\overline{r_k}],$$

where r_k is the estimator computed from the kth sample, and

$$ISE(r_k) = \int_0^1 (r_k(x) - r(x))^2 dx,\ IAE(r_k) = \int_{\mathbb{R}} |r_k(x) - r(x)| dx.$$

Table 6.2 shows that in terms of the average ISE and IAE of the two considered estimators, the proposed two-time-scale recursive estimators (4.1.2) have a smaller $\overline{ISE}$ and smaller $\overline{IAE}$ compared to the non recursive estimator (4.1.1) by using any one of the three regression functions and by considering any one of the three different noise variance.

4.5 Real dataset

An interesting subject of any dataset is to estimate the unknown regression in order to predict the response variable Y when we know the explanatory variable X. In this section, we apply our proposed two-time-scale estimators r_n defined in (4.1.2) and the non recursive estimator $\tilde{r}_{n,m}$ given in (4.1.1) on the following two datasets.

1. Firstly, we consider the `CO2` dataset[1] which contained 60 observations on two variables: `Year` and `CO2` in August. Scientists recorded `CO2` levels, in parts per million (ppm), in the atmosphere for each `Year` from the start of 1958 through 2018. Finally, we used the Monte-Carlo method to obtain $m = 60$ for our recursive estimator r_n and $m = 50$ for the non recursive estimator $\tilde{r}_{n,m}$. We observe that our estimator is more close to the observed data, than the non recursive estimator, especially near the boundaries.

2. Secondly, we consider the `wage1` dataset which appear in `R` package `np` (see Wooldridge (2000)). Cross-section wage data consisting of a random sample taken from the U.S. Current Population Survey for the year 1976. There are 526 observations of the average hourly earnings and the years of education. We used the Monte-Carlo method to obtain $m = 68$ for our recursive estimator r_n and $m = 170$ for the non recursive estimator $\tilde{r}_{n,m}$. We observe that our estimator is more close to the observed data, than the non recursive estimator, especially near the boundaries.

Figures 5.3 and 5.4 show that our proposed two-time-scale recursive estimators (4.1.2) can give better results compared to the non recursive estimator (4.1.1).

[1] `https://www.co2.earth/monthly-co2`

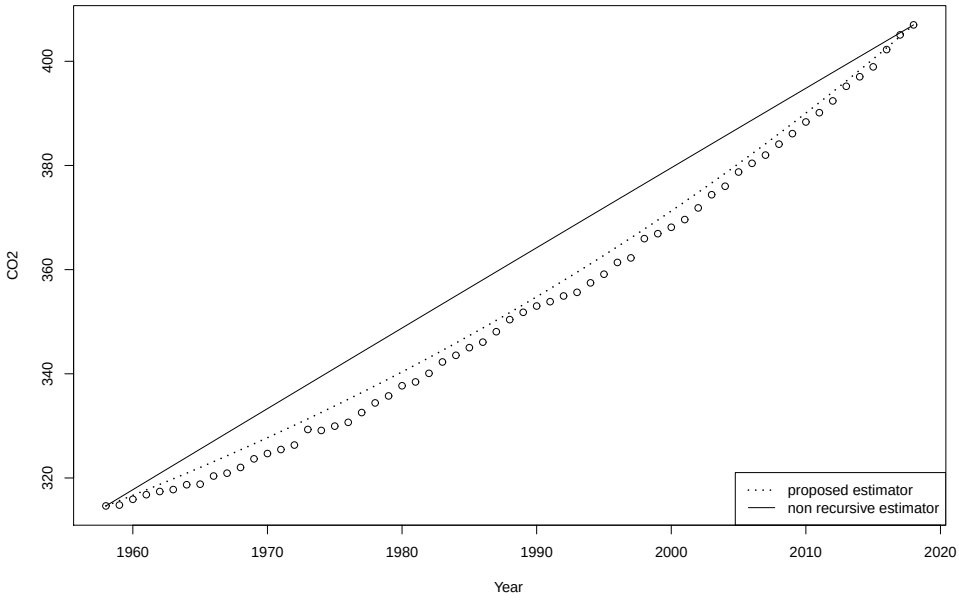

Figure 4.1: Qualitative comparison between r_n and $\tilde{r}_{n,m}$ for `CO2` dataset .

[Qualitative comparison between the two considered estimators. Here we consider the `CO2` dataset using our proposed estimator r_n with the stepsize $(\gamma_n, \beta_n) = (n^{-0.98}, 0.31n^{-0.99})$ and the non recursive estimator $\tilde{r}_{n,m}$]

4.6 Concluding remarks

In the present chapter we investigated a recursive nonparametric regression estimator to overcome the edge estimation problem based on Bernstein polynomials and stochastic algorithm with two-time-scale. The proposed estimator asymptotically follows normal distribution. Moreover, our proposed estimators attained the asymptotic convergence rate $O\left(n^{-4/5}\right)$ within the interval $(0, 1)$ and $O\left(n^{-2/3}\right)$ near the edges $\{0, 1\}$. A future research direction would be to extend our findings to the setting of serially dependent observations, α-mixing framework like in Khardani and Slaoui (2019). Another direction is to investigate the data-driven bandwidth selection procedures (see Slaoui (2015c)), which requires non trivial mathematics, this would go well beyond the scope of the present chapter. We plan also to extend our approach to the case of locally linear two-time-scale recursive regression (see Härdle et al. (2004) in the case of locally linear regression).

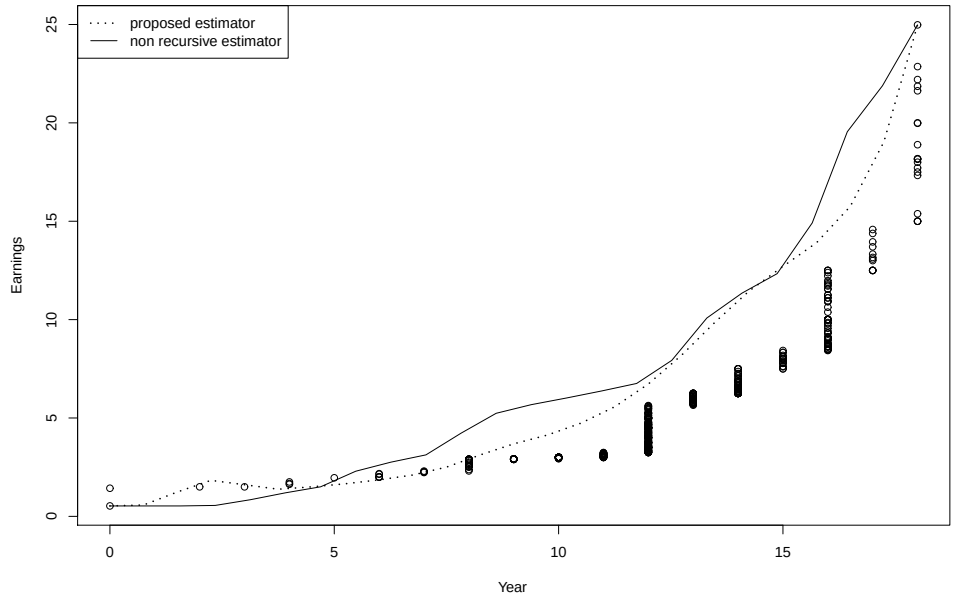

Figure 4.2: Qualitative comparison between $\mathrm{t}r_n$ and $\tilde{r}_{n,m}$ for `wage1` dataset .

[Qualitative comparison between the two considered estimators. Here we consider the The `wage1` dataset of the package `np` using our proposed estimator r_n with the stepsize $(\gamma_n, \beta_n) = (n^{-0.98}, 0.31n^{-0.99})$ and the non recursive estimator $\tilde{r}_{n,m}$]

4.7 Auxiliary results and Mathematical developments

This section is devoted to the detailed proofs of our results. Throughout this chapter we consider the following notations:

$$\xi' = \lim_{n\to\infty} (n\gamma_n)^{-1}, \quad s_n(x) = \sum_{k=1}^{n} \gamma_k,$$

$$\Pi_n = \prod_{j=1}^{n}(1-\beta_j), \quad \mathcal{Q}_n = \prod_{j=1}^{n}(1-\gamma_j), \quad C = \tfrac{1}{\sqrt{2}} + 4\left(1-\sqrt{2/3}\right).$$

Let us first state the following technical lemmas.

Lemma 4.1 (Mokkadem et al (2009)). *Let $(v_n) \in \mathcal{GS}(v)$, $(\gamma_n) \in \mathcal{GS}(-\gamma)$, and let $l > 0$ such that $l - v\xi > 0$. We have*

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

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

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

Let us now use the following decomposition:

$$r_n(x) - r(x) = \frac{1}{f(x)} [\mathcal{T}_n(x) + \mathcal{R}_{n,1}(x) + \mathcal{R}_{n,2}(x) + \mathcal{R}_{n,3}(x) + \mathcal{R}_{n,4}(x)] + \mathcal{R}_{n,5}(x),$$

where

$$\begin{aligned}
\mathcal{T}_n(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k [\mathcal{W}_k(x) - r(x)\mathcal{Z}_k(x)], \qquad (4.7.1)\\
\mathcal{R}_{n,1}(x) &= r(x)\Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k} [f_k(x) - f_{k-1}(x)],\\
\mathcal{R}_{n,2}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k [\mathcal{W}_k(x) - \mathbb{E}(\mathcal{W}_k(x))] \left[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)} \right],\\
\mathcal{R}_{n,3}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k [\mathbb{E}(\mathcal{W}_k(x)) - a(x)] \left[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)} \right],\\
\mathcal{R}_{n,4}(x) &= r(x)\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k \left[\frac{(f(x) - f_{k-1}(x))^2}{f_{k-1}(x)} \right],\\
\mathcal{R}_{n,5}(x) &= \Pi_n (r_0(x) - r(x)).
\end{aligned}$$

Lemma 4.2. *Under assumptions* (**A1**)-(**A2**), *we have*

$$f_n(x) - f(x) = O\left(\max\{ \sqrt{\gamma_n \ln(s_n) v_n}, v_n^{-2} \} \right).$$

Lemma 4.3. *Under assumptions* (**A1**)-(**A2**), *we have*

$$\mathcal{R}_{n,1}(x) = \mathcal{R}_{n,2}(x) = \mathcal{R}_{n,3}(x) = \mathcal{R}_{n,4}(x) = \mathcal{R}_{n,5}(x) = o\left(\max\{ \sqrt{\beta_n v_n}, v_n^{-2} \} \right).$$

Lemma 4.4. *Let* $\mathcal{E}_m(x) = \sum_{k=0}^{m} b_k^2(m,x)$. *We have*

(i) $0 \le \mathcal{E}_m(x) \le 1$, *for* $x \in [0,1]$.

(ii) $\mathcal{E}_m(x) = m^{-1/2}[\psi(x) + o_x((1)]$, *for* $x \in (0,1)$.

(iii) $\mathcal{E}_m(0) = \mathcal{E}_m(1) = 1$.

(*iv*) *Let g be any continuous function on $[0,1]$. Then* $m^{1/2}\int_0^1 g(x)\mathcal{E}_m(x)dx = \int_0^1 g(x)\psi(x)dx + o(1)$.

Lemma 4.5. *For $x \in [0,1]$, we have*

$$\mathbb{E}(\mathcal{Z}_k(x)) = f(x) + \Delta_1(x)m_k^{-1} + o(m_k^{-1}), \tag{4.7.2}$$

$$\mathbb{E}(\mathcal{W}_k(x)) = r(x)f(x) + \Delta_2(x)m_k^{-1} + o(m_k^{-1}), \tag{4.7.3}$$

$$\begin{aligned} Var(\mathcal{W}_k(x)) &= \left(m_k^{1/2}\mathbb{E}(Y^2 \mid X = x)f(x)\psi(x) + o_x\left(m_k^{1/2}\right)\right)\mathbb{1}_{\{x\in(0,1)\}} \\ &\quad + \left(\mathbb{E}[Y^2 \mid X = x]f(x)m_k + o_x(m_k)\right)\mathbb{1}_{\{x\in\{0,1\}\}}, \end{aligned} \tag{4.7.4}$$

$$\begin{aligned} Var(\mathcal{Z}_k(x)) &= \left(f(x)\psi(x)m_k^{1/2} + o_x\left(m_k^{1/2}\right)\right)\mathbb{1}_{\{x\in(0,1)\}} \\ &\quad + \left(f(x)m_k + o(m_k)\right)\mathbb{1}_{\{x\in\{0,1\}\}}, \end{aligned} \tag{4.7.5}$$

and

$$\begin{aligned} Cov(\mathcal{W}_k(x), \mathcal{Z}_k(x)) &= \left(r(x)f(x)\psi(x)m_k^{1/2} + o_x\left(m_k^{1/2}\right)\right)\mathbb{1}_{\{x\in(0,1)\}} \\ &\quad + \left(r(x)f(x)m_k + o_x(m_k)\right)\mathbb{1}_{\{x\in\{0,1\}\}}. \end{aligned} \tag{4.7.6}$$

4.7.1 Proof of lemma 4.2

In order to prove this lemma, we referred to the proposition 3.1 of Slaoui and Jmaei (2019) and we followed the same steps of proof of theorem 2 page 44 of Slaoui (2006).

4.7.2 Proof of lemma 4.3

Let us first state the following notation,

$$(\zeta_n)_n = \begin{cases} \sqrt{\beta_n v_n} & \text{if } \lim\limits_{n\to\infty} \beta_n m_n^2 v_n = \infty, \\ v_n^{-2} & \text{otherwise.} \end{cases}$$

First, we have $\mathcal{R}_{n,5}(x) = o(\zeta_n)$.

Proof of lemma 4.3 for $\mathcal{R}_{n,1}$

We first note that $\mathcal{R}_{n,1}$ can be written as,

$$\begin{aligned}
\mathcal{R}_{n,1}(x) &= \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f_{k-1}(x)] \\
&= \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f(x)] - \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_{k-1}(x) - f(x)] \\
&= \Pi_n \sum_{k=1}^{n-1} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f(x)] + \frac{\beta_n}{\gamma_n}[f_n(x) - f(x)] - \Pi_n \frac{\Pi_1^{-1}\beta_1}{\gamma_1}[f_0(x) - f(x)] \\
&\quad -\Pi_n \sum_{k=1}^{n-1} \frac{\Pi_{k+1}^{-1}\beta_{k+1}}{\gamma_{k+1}}[f_k(x) - f(x)] \\
&= \Pi_n \sum_{k=1}^{n-1} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}\left(1 - \frac{\Pi_k}{\Pi_{k+1}}\frac{\beta_{k+1}}{\beta_k}\frac{\gamma_k}{\gamma_{k+1}}\right)(f_k(x) - f(x)) + \frac{\beta_n}{\gamma_n}(f_n(x) - f(x)) \\
&\quad -\Pi_n \frac{\beta_1}{(1-\beta_1)\gamma_1}(f_0(x) - f(x)). \qquad (4.7.7)
\end{aligned}$$

Moreover, in view of $(A1)$, we infer that

$$1 - \frac{\Pi_k}{\Pi_{k+1}}\frac{\beta_{k+1}}{\beta_k}\frac{\gamma_k}{\gamma_{k+1}} = 1 - \frac{1}{(1-\beta_{k+1})}\frac{\beta_k^{-1}}{\beta_{k+1}^{-1}}\frac{\gamma_k}{\gamma_{k+1}} = O(\beta_k). \qquad (4.7.8)$$

Then, the application of Lemma 4.2 together with (4.7.7) and (4.7.8), ensures that

$$\begin{aligned}
\mathcal{R}_{n,1}(x) &= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\frac{\beta_k^2}{\gamma_k}(f_k(x) - f(x)) + \frac{\beta_n}{\gamma_n}(f_n(x) - f(x)) + \Pi_n\right) \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\frac{\beta_k^2}{\gamma_k}(\gamma_k v_k \ln s_k)^{1/2} + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\frac{\beta_k^2}{\gamma_k}v_k^{-2} + \frac{\beta_n}{\gamma_n}(\gamma_n v_n \ln s_n)^{1/2} + \frac{\beta_n}{\gamma_n}v_n^{-2}\right) \\
&\quad +O(\Pi_n) \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o\left(\sqrt{\beta_k v_k}\right) + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o(v_k^{-2}) + o\left(\sqrt{\beta_n v_n}\right) + o(v_n^{-2}) + \Pi_n\right) \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o(\zeta_n)\right) + o(\zeta_n) + O(\Pi_n) \\
&= o(\zeta_n).
\end{aligned}$$

Proof of lemma 4.3 for $\mathcal{R}_{n,2}$

Throughout this proof we use the following notation:

$$\begin{aligned}
\mathcal{L}_k(x) &= \mathcal{W}_k(x) - \mathbb{E}(\mathcal{W}_k(x)), \\
G_k(x) &= \frac{f(x) - f_k(x)}{f_k(x)}, \\
M_n(x) &= \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \mathcal{L}_k(x) G_{k-1}(x).
\end{aligned}$$

First, we set $\mathcal{F}_k = \sigma((X_1, Y_1), \ldots, (X_k, Y_k))$, and we note that the increasing process of the martingale $(M_n(x)$ satisfies,

$$\begin{aligned}
< M >_n (x) &= \sum_{k=1}^{n} \mathbb{E}\left[\Pi_k^{-2} \beta_k^2 \mathcal{L}_k^2(x) G_{k-1}^2(x) \mid \mathcal{F}_{k-1}\right] \\
&= \sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 G_{k-1}^2(x) \mathbb{E}\left[\mathcal{L}_k^2(x) \mid \mathcal{F}_{k-1}\right] \\
&= \sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 G_{k-1}^2(x) \mathbb{E}(\mathcal{L}_k^2(x)) \\
&= \sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 G_{k-1}^2(x) Var(\mathcal{W}_k(x)).
\end{aligned}$$

In view of (4.7.4), the application of Lemma 4.2 ensures that

$$\begin{aligned}
< M >_n (x) &= O\left(\sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 \left(\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)}\right)^2 v_k\right) \\
&= O\left(\sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 \left(\gamma_k \ln(s_k) v_k + v_k^{-4}\right) v_k\right) \\
&= O\left(\sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 v_k^2 \gamma_k \ln(s_k) + \sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 v_k^{-3}\right). \qquad (4.7.9)
\end{aligned}$$

Let us first consider the case when $\lim_{n\to\infty} n\beta_n = \infty$. The application of Lemma 6.1 gives

$$< M >_n (x) = O\left(\Pi_n^{-2} \beta_n \gamma_n \ln(s_n) v_n^2 + \Pi_n^{-2} \beta_n v_n^{-3}\right).$$

Moreover, we note that for all $\epsilon > 0$ we have

$$\ln(\Pi_n^{-2}) = \sum_{k=1}^{n} \ln(1 - \beta_k)^{-2} = \sum_{k=1}^{n} (2\beta_k + o(\beta_k))$$

$$= O\left(\sum_{k=1}^{n} \beta_k k^{\epsilon}\right).$$

Since we have $\beta_n n^{\epsilon} \in \mathcal{GS}(-(\beta-\epsilon))$ with $(\beta-\epsilon) < 1$. The application of Lemma 6.1 ensures that $\lim_{n\to\infty} \frac{n\beta_n n^{\epsilon}}{\sum_{k=1}^{n} \beta_k k^{\epsilon}} = 1-(\beta-\epsilon)$. It comes that $\ln(\Pi_n^{-2}) = O(n^{1+\epsilon}\beta_n)$. Moreover, the sequences $(\beta_n \gamma_n \ln(s_n) v_n^2)$ and $(\beta_n v_n^{-3})$ tend to zero. It comes that,

$$\ln < M >_n (x) = O\left(\ln(\Pi_n^{-2})\right) = O\left(n^{1+\epsilon}\beta_n\right).$$

Then, the application of Theorem 1.3.15 in Duflo (1997), ensures that for any $\delta > 0$,

$$\begin{aligned} | M_n(x) | &= o\left(< M >_n^{1/2} (x) \left(\ln < M >_n (x)\right)^{\frac{1+\delta}{2}}\right) + O(1) \\ &= o\left(\Pi_n^{-1} \left(\beta_n^{1/2}\gamma_n^{1/2} \ln(s_n)^{1/2} v_n + \beta_n^{1/2} v_n^{-3/2}\right) \left(n^{1+\epsilon}\beta_n\right)^{\frac{1+\delta}{2}}\right) + O(1). \end{aligned}$$

Now, we set $\epsilon > 0$ and $\delta > 0$ such that $\left((\gamma_n v_n^2)^{1/2} \left(n^{1+\epsilon}\beta_n\right)^{\frac{1+\delta}{2}}\right) \in \mathcal{GS}(\mu)$ with $\mu < 0$. For more precision, the existence of ε and δ should be ensured by the condition $\alpha > \left(a\mathbb{1}_{\{x\in\{0,1\}\}} + \frac{a}{2}\mathbb{1}_{\{x\in(0,1)\}}\right) + 1 - \beta$. Moreover, we obtain

$$\begin{aligned} \Pi_n |M_n(x)| &= o\left((\beta_n v_n)^{1/2} (\gamma_n v_n)^{1/2} \ln(s_n)^{1/2} \left(n^{1+\epsilon}\beta_n\right)^{\frac{1+\delta}{2}} + (\beta_n v_n)^{1/2} v_n^{-2} \left(n^{1+\epsilon}\beta_n\right)^{\frac{1+\delta}{2}}\right) \\ &\quad + O(\Pi_n) \\ &= o\left((\beta_n v_n)^{1/2} + v_n^{-2}\right) + o(\zeta_n) \\ &= o(\zeta_n). \end{aligned}$$

Now, in the case when the sequence $(n\beta_n)$ is bounded when n goes to infinity. The application of Lemma 6.1 for all $(\mathcal{C}_n) \in \mathcal{GS}(0)$ together with (4.7.9), ensures that

$$\begin{aligned} < M >_n (x) &= O\left(\sum_{k=1}^{n} \Pi_k^{-2}\beta_k \left[(\beta_k v_k)(\gamma_k v_k \ln(s_k)) + (\beta_k v_k) v_k^{-4}\right]\right) \\ &= O\left(\sum_{k=1}^{n} \Pi_k^{-2}\beta_k \left[o(\beta_k v_k \mathcal{C}_k) + o\left(v_k^{-4}\mathcal{C}_k\right)\right]\right) \\ &= o\left(\sum_{k=1}^{n} \Pi_k^{-2}\beta_k o\left(\zeta_k^2 \mathcal{C}_k\right)\right) \\ &= o\left(\Pi_n^{-2}\zeta_n^2 \mathcal{C}_n\right). \end{aligned}$$

Moreover, since the sequence $(n\beta_n)$ is bounded when n goes to infinity, in this case, we have $\Pi_n^{-1} \in \mathcal{GS}(\xi^{-1})$ and $\ln\left(\Pi_n^{-2}\zeta_n^2\mathcal{L}_n\right) = O(\ln n)$. Then, the application of Theorem 1.3.15 in Duflo (1997), ensures that for all $\delta > 0$,

$$| M_n(x) | = o\left(< M >_n^{1/2} (x) \left(\ln < M >_n (x)\right)^{\frac{1+\delta}{2}}\right) + O(1)$$

$$= \quad o\left(\Pi_n^{-1}\zeta_n\mathcal{C}_n^{1/2}\ln(n)^{\frac{1+\delta}{2}}\right)+O(1).$$

Thus, for $\delta=1$ and $\mathcal{L}_n=(\ln n)^{-2}$, we get $\Pi_n \mid M_n(x)\mid = o(\zeta_n)$, which ensures that

$$\mathcal{R}_{n,2}(x)=o(\zeta_n).$$

Proof of lemma 4.3 for $\mathcal{R}_{n,3}$

The application of Lemma 6.1 together with Lemma 4.2, ensures that

$$\begin{aligned}
\mathcal{R}_{n,3}(x) &= \Pi_n\sum_{k=0}^{n-1}\Pi_{k+1}^{-1}\beta_{k+1}[\mathbb{E}(\mathcal{W}_{k+1}(x))-a(x)]\left[\frac{f(x)-f_k(x)}{f_k(x)}\right]\\
&= \Pi_n\sum_{k=0}^{n-1}\Pi_{k+1}^{-1}\beta_{k+1}[\Delta_2(x)m_{k+1}^{-1}+o(m_{k+1}^{-1})]\left[\frac{f(x)-f_k(x)}{f_k(x)}\right]\\
&= O\left(\Pi_n\sum_{k=0}^{n-1}\Pi_{k+1}^{-1}\beta_{k+1}m_{k+1}^{-1}\mid f(x)-f_k(x)\mid\right)\\
&= o(\zeta_n).
\end{aligned}$$

Proof of lemma 4.3 for $\mathcal{R}_{n,4}$

The application of Lemma 4.2 ensures that

$$\begin{aligned}
\mathcal{R}_{n,4}(x) &= \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k O\left((f(x)-f_{k-1}(x))^2\right)\\
&= O\left(\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k((\gamma_k v_k\ln s_k)+v_k^{-4})\right)\\
&= o(\zeta_n).
\end{aligned}$$

4.7.3 Proof of Proposition 4.1

The application of Lemma 6.1 together with (4.7.1) and (4.7.2), ensures that in the case when $a,x\in\mathcal{A}_{a,x}^1\bigcup\mathcal{A}_{a,x}^2$,

$$\mathbb{E}(\mathcal{T}_n(x))=\frac{1}{1-a\xi}(\Delta_2(x)-r(x)\Delta_1(x)+o(1))m_n^{-1}, \tag{4.7.10}$$

in the case when $a,x\in\mathcal{A}_{a,x}^3$, we have $\mathbb{E}(\mathcal{T}_n(x))=o\left(\sqrt{\beta_n m_n}\right)$, and in the case when $a,x\in\mathcal{A}_{a,x}^4$, we have $\mathbb{E}(\mathcal{T}_n(x))=o\left(\sqrt{\beta_n m_n^{1/2}}\right)$.

Let assumptions (**A1**) - (**A′3**) hold such as $f(x) > 0$ for $x \in \{0, 1\}$. The asymptotic behaviour of $r_n - r$ is given by the one of $\mathcal{T}_n$. More precisely, we establish the following lemma.

Lemma 4.6. *1. In the case when $x \in (0, 1)$, if there exists $c \geq 0$ such that $\beta_n^{-1} m_n^{-5/2} \to c$, (resp. the case when $x \in \{0, 1\}$, if $\beta_n^{-1} m_n^{-3} \to c$), then*

$$\mathcal{V}_n \mathcal{T}_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{(1 - a\xi)} \mathcal{B}ias\,(x)\, f\,(x)\,, \mathcal{V}\,(x)\, f^2\,(x)\right).$$

2. In the case when $x \in (0, 1)$, if $\beta_n^{-1} m_n^{-5/2} \to \infty$, (resp. the case when $x \in \{0, 1\}$, if $\beta_n^{-1} m_n^{-3} \to \infty$), then

$$m_n\ \mathcal{T}_n(x) \xrightarrow{\mathbb{P}} \frac{1}{(1 - a\xi)} \mathcal{B}ias\,(x)\, f\,(x)\,.$$

Proof: Let us first note that,

$$\mathcal{T}_n(x) - \mathbb{E}(\mathcal{T}_n(x)) \;=\; \Pi_n \sum_{k=1}^{n} \{v_k(x) - \mathbb{E}(v_k(x))\}\,,$$

where

$$v_k(x) = \Pi_k^{-1} \beta_k (\mathcal{W}_k(x) - r(x) \mathcal{Z}_k(x)).$$

Let us now assume that, when $a, x \in \mathcal{A}_{a,x}^3 \bigcup \mathcal{A}_{a,x}^4$, we have

$$\mathcal{V}_n \left(\mathcal{T}_n\,(x) - \mathbb{E}\left[\mathcal{T}_n\,(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathcal{V}\,(x)\, f^2\,(x)\right). \tag{4.7.11}$$

In the case when $a, x \in \mathcal{A}_{a,x}^1 \bigcup \mathcal{A}_{a,x}^2$, (4.3.2) gives $m_n \left(\mathcal{T}_n\,(x) - \mathbb{E}\left(\mathcal{T}_n\,(x)\right)\right) \xrightarrow{\mathbb{P}} 0$. Let us now prove (4.7.11), for this purpose, we set

$$\mathcal{B}_k\,(x) = \Pi_k^{-1} \beta_k \left\{v_k(x) - \mathbb{E}(v_k(x))\right\}. \tag{4.7.12}$$

Moreover, since $\mathrm{Var}\left(\mathcal{T}_k\,(x)\right) = O\left(m_k^{1/2}\left[1 + o_x\,(1)\right] \mathbb{1}_{\mathcal{A}_{a,x}^4} + m_k\left[1 + o\,(1)\right] \mathbb{1}_{\mathcal{A}_{a,x}^3}\right)$, the application of Lemma 6.1 ensures that

$$\begin{aligned}
\mathcal{Y}_n^2 \;&=\; \sum_{k=1}^{n} \mathrm{Var}\left(\mathcal{B}_k\,(x)\right) = \sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 \mathrm{Var}\left(v_k(x)\right), \\
&=\; \frac{\beta_n m_n^{1/2}}{\Pi_n^2} \mathcal{V}\,(x) + o_x\left(\frac{\beta_n m_n^{1/2}}{\Pi_n^2}\right) \mathbb{1}_{\mathcal{A}_{a,x}^4} + o\left(\frac{\gamma_n m_n}{\Pi_n^2}\right) \mathbb{1}_{\mathcal{A}_{a,x}^3}.
\end{aligned}$$

Further, the application of Lemma 6.1 ensure that for all $p > 0$

$$\begin{aligned}\mathbb{E}\left[\left|\mathcal{B}_k(x)\right|^{2+p}\right] &= m_k^{2+p}\mathbb{E}\left(\left|Y_k - r(x)\right|^{2+p}\left\{\sum_{j=0}^{m_k-1}\mathbb{1}_{\left\{\frac{j}{m_k}<T_k\leq\frac{j+1}{m_k}\right\}}b_j(m_k-1,x)\right\}^{2+p}\right)\\ &= O\left(m_k^{3(2+p)/4}\right)\mathbb{1}_{\mathcal{A}_{a,x}^4} + O\left(m_k^{2+p}\right)\mathbb{1}_{\mathcal{A}_{a,x}^3}. \qquad (4.7.13)\end{aligned}$$

Using the fact that $\lim_{n\to\infty}(n\gamma_n) > \frac{2\alpha-a}{4}\mathbb{1}_{\mathcal{A}_{a,x}^4} + \frac{\alpha-a}{2}\mathbb{1}_{\mathcal{A}_{a,x}^3}$, there exists $p > 0$ such that $\lim_{n\to\infty}(n\gamma_n) > \left(\frac{1+p}{2+p}\alpha - \frac{3}{4}\right)\mathbb{1}_{\mathcal{A}_{a,x}^4} + \left(\frac{1+p}{2+p}\alpha - \frac{3}{2}\right)\mathbb{1}_{\mathcal{A}_{a,x}^3}$. Lemma 6.1, gives

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}\left[\left|\mathcal{B}_k((x)\right|^{2+p}\right] &= O\left(\sum_{k=1}^{n}\Pi_k^{-2-p}\beta_k^{2+p}\mathbb{E}\left[\left|\mathcal{T}_k(x)\right|^{2+p}\right]\right)\\ &= O\left(\gamma_n^{1+p}\Pi_n^{-2-p}m_n^{3(2+p)/4}\right)\mathbb{1}_{\mathcal{A}_{a,x}^4} + O\left(\gamma_n^{1+p}\Pi_n^{-2-p}m_n^{2+p}\right)\mathbb{1}_{\mathcal{A}_{a,x}^3},\end{aligned}$$

from which we deduce that

$$\frac{1}{\mathcal{Y}_n^{2+p}}\sum_{k=1}^{n}\mathbb{E}\left[\left|\mathcal{B}_k((x)\right|^{2+p}\right] = O\left(m_n[\gamma_n m_n]^{p/2}\right) = o(1).$$

The convergence in (4.7.11) then follows from the application of Lyapounov's Theorem.

Table 4.1: Average ISE and IAE for $N = 500$ trials of $\tilde{r}_{n,m}$ and r_n.
[Average ISE and IAE for $N = 500$ trials of $\tilde{r}_{n,m}$ and r_n with the choice $(\gamma_n, \beta_n) = (n^{-0.98}, 0.31n^{-99})$. The bold values indicate the smallest values.]

		ISE		**IAE**	
Model	**n**	**Recursive**	**non recursive**	**Recursive**	**non recursive**
$\sigma = 0.05$					
	50	**0.108152**	0.263466	**0.278830**	0.512572
(a)	100	**0.146173**	0.275628	**0.342211**	0.523425
	200	**0.172898**	0.265569	**0.385824**	0.514904
	50	**0.149485**	0.261717	**0.374100**	0.511469
(b)	100	**0.180104**	0.250942	**0.416693**	0.500793
	200	**0.219903**	0.256299	**0.461882**	0.506157
	50	**0.192147**	0.246308	**0.375262**	0.495552
(c)	100	**0.166492**	0.241962	**0.345006**	0.491480
	200	0.315283	**0.242560**	**0.488220**	0.492026
$\sigma = 0.1$					
	50	**0.237895**	0.279316	**0.459012**	0.527103
(a)	100	**0.142256**	0.273119	**0.339965**	0.521741
	200	**0.205788**	0.261468	**0.429494**	0.510851
	50	**0.240268**	0.262336	**0.478994**	0.512000
(b)	100	**0.186827**	0.263747	**0.422515**	0.513192
	200	**0.168894**	0.260439	**0.403971**	0.510155
	50	0.241451	**0.236557**	**0.432593**	0.485448
(c)	100	**0.169090**	0.2400389	**0.354381**	0.489659
	200	0.669683	**0.237169**	0.714953	**0.486410**
$\sigma = 1$					
	50	**0.153426**	0.245387	**0.256094**	0.495342
(a)	100	**0.089536**	0.231082	**0.242418**	0.480270
	200	**0.081775**	0.216917	**0.234569**	0.465104
	50	**0.076953**	0.238752	**0.257034**	0.487440
(b)	100	**0.069710**	0.214250	**0.231290**	0.460268
	200	**0.086527**	0.240928	**0.263234**	0.488604
	50	**0.044846**	0.259154	**0.113061**	0.508533
(c)	100	**0.030788**	0.236227	**0.074095**	0.483667
	200	**0.026170**	0.227478	**0.078804**	0.474521

Part II

Non parametric estimation using Lagrange polynomials

Chapter 5

Estimation of a distribution function using Lagrange polynomial

Key words and phrases: Asymptotic properties, Distribution estimator, Lagrange polynomials, Tchebychev-Gauss points.

Abstract 3: *The estimation of the distribution function of a real random variable is an intrinsic topic in non parametric estimation. To this end, a distribution estimator based on Lagrange polynomials and Tchebychev-Gauss points, is introduced. Some asymptotic properties of the proposed estimator are investigated, such as its asymptotic bias, variance, mean squared error and Chung-Smirnov propriety. The asymptotic normality and the uniform convergence of the estimator are also established. Lastly, the performance of the proposed estimator is explored through a certain simulation examples.*

5.1 Introduction

Non parametric distribution estimation is undoubtedly a useful tool of data analysis, which is reflected by the multiple literary works addressing the topic. Let $X_1, \ldots, X_n$ be a sequence of *i.i.d.* random variables having a common unknown distribution function F with associated density f supported on a compact interval. Within the framework of the non parametric estimation, since we know that F is continuous, we consider the estimation of F by using smooth functions rather than the empirical distribution function, which is not continuous. Several methods have been set forward for smooth estimation of density and distribution functions. The most popular one, called kernel method, is introduced by

Rosenblatt (1956). The advances were carried out by Parzen (1962) to estimate density function. The kernel distribution estimator was identified by Nadaraya (1964) as

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h_n}\right). \tag{5.1.1}$$

The properties of $\widehat{F}_n$ have been known for a long time, for example its uniform convergence towards F with continuous f (Nadaraya (1964), Winter (1973), Yamato (1973)), then unconditionally on f (Singh et al (1983)) and its asymptotic normality (Watson and Leadletter (1964)). Winter (1979) also demonstrated that $\widehat{F}_n$ checks the Chung-Smirnov property with probability 1.

However, Kernel methods have estimation problems at the edges, when we have a random variable X with distribution function supported on a compact interval. In order to overcome this problem, various methods such as the Bernstein polynomial density and distribution estimators were introduced first by Vitale (1975) and then extended by Tenbusch (1994), Babu et al. (2002) and Kakizawa (2004). In particular, following Babu et al. (2002), the estimator using Bernstein polynomial with order $\nu > 0$ of the distribution F is defined as

$$\tilde{F}_{n,\nu}(x) = \sum_{k=0}^{\nu} \overline{F}_n\left(k/\nu\right) b_k(\nu, x), \tag{5.1.2}$$

with $\overline{F}_n$ is the empirical distribution function and $b_k(\nu, x)$ is the Bernstein polynomial. This estimator is asymptotically unbiased. Babu et al. (2002) found also that $\tilde{F}_{n,\nu}$ to be uniformly strongly consistent. Babu and Chaubey (2006) adapted the Bernstein estimator to the problem of estimating a multivariate distribution function (including the case of dependent observations under α mixing). Leblanc (2009) reported that it has the Chung-Smirnov property, as $n \to \infty$.

In this chapter, we present what appears to be a new method based on Lagrange polynomials and Tchebytchev-Gauss points. When we have a random variable X with distribution F supported on a compact interval $[a, b]$ such as $a < b$, we can transform X into Y, a random variable with support $[-1, 1]$ through the transformation $Y = \dfrac{X - (a+b)/2}{(b-a)/2}$. Transformations such as $Y = 2X/(1+X) - 1$ and $Y = 2\pi^{-1}\arctan(X)$ can be used to cover the cases of random variables X with support $\mathbb{R}_+$ and $\mathbb{R}$ respectively. Once the random variable X is transformed into Y, we can apply Lagrange polynomials with Tchebytchev-Gauss points to approximate the distribution function of Y on the interval $[-1, 1]$. In the theoretical part of this chapter, we consider the case where f is supported

on $[-1,1]$, and we propose an estimator using Lagrange polynomial with order $m > 0$ of the distribution F using Lagrange polynomial expressed as,

$$\mathcal{F}_{n,m}(x) = \sum_{i=1}^{m} \overline{F}_n(x_i)\mathcal{L}_i(x), \tag{5.1.3}$$

where, for all $i = 1 \dots m$, $x_i = \cos\left((2i-1)\pi/2m\right)$ are Tchebytchev-Gauss points,

$$\mathcal{L}_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{m} \frac{x - x_j}{x_i - x_j}$$

is the Lagrange polynomial, and $\overline{F}_n$ denotes the empirical distribution function obtained from a random sample of size n. The points $(x_i)_{1\leq i\leq m}$ are the zeros of the Tchebytchev polynomial $T_m(x) = \cos(m \arccos(x))$. Also, according to Austin (2016), using this choice of points, we have

$$\sup_{x\in[-1,1]} \left| \sum_{i=1}^{m} v(x_i)\mathcal{L}_i(x) - v(x) \right| \leq (\pi/2)^k ||v^{(k)}||_\infty / ((m+1)(m)\dots(m-k+2)),$$

for any continuous function v of class C^k (for $k \geq 1$) on the interval $[-1,1]$. This result was studied by Jackson in the early 20th century (see Jackson (1911), Jackson (1912)). His results can also be found in Cheney (1966), chapter 4, section 6, page 147, which is the main idea of the proposed estimator. As an excellent reference for properties of Lagrange polynomial with Tchebytchev-Gauss points in the deterministic case, we refer the reader to Austin (2016). To the best of our knowledge, the estimator presented here has not been studied so far, which stands for the basic motivation of the chapter.

The main objective of this chapter is to study the properties of the distribution estimator (5.1.3). We consider first the mean squared error for a fixed x, for $-1 < x < 1$, and split it into bias squared and variance terms. Then, we establish the uniform convergence of this estimator, the Chung-Smirnov property and the (pointwise) asymptotic normality of the proposed estimator. Basically, the remainder of the chapter is organized as follows. In the next section, we display the assumptions and notations. In Section 6.3, we exhibit our main results. Section 6.4 highlights a simulation study that compares the performance of the proposed estimator $\mathcal{F}_{n,m}$ with the Bernstein estimator (5.1.2) and with the kernel (standard Gaussian kernel) estimator (5.1.1). Section 6.6 provides the proofs of our theoretical results.

5.2 Lagrange polynomials

Lagrange polynomials, named after Joseph-Louis Lagrange, are used to interpolate a series of points by a polynomial that passes exactly through these points. Let first consider the following theorem.

Theorem 5.2.1. *Let $x_0, \ldots, x_n$ be $n+1$ distinct numbers, and let f be a function defined on a domain $[a, b]$ containing these numbers. Then the polynomial defined by*

$$p_n(x) = \sum_{i=0}^{n} f(x_i)\mathcal{L}_i(x)$$

is the unique polynomial of degree $\leq n$ that satisfies

$$p(x_j) = f(x_j),\ j = 0 \ldots n.$$

The polynomial p_n is called the interpolating polynomial of f and the polynomial $\mathcal{L}_i$ which meets this equality is Lagrange interpolation polynomial.This Lagrange polynomial is given by

$$\mathcal{L}_k(x) = \prod_{i=0, i\neq k} \frac{(x - x_i)}{(x_k - x_i)} \text{ for each } k = 0, \ldots, n.$$

The Lagrange polynomials have several properties, such as

Proposition 5.1. *1. For $i, j = 0 \ldots n$, we have*

$$\mathcal{L}_i(x_j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

2. $\sum_{i=0}^{n} \mathcal{L}_i(x) = 1$.

3. Barycentric form

$$\mathcal{L}_j(x) = \frac{T_n(x)}{T_n'(x_j)(x - x_j)} \text{ for } j = 0 \ldots n,$$

where $T_n(x) = \cos(n \arccos(x))$ is the Tchebytchev polynomial.

5.3 Assumptions and Notations

We consider the following definition.

Definition 5.1. *Let g be a function defined on $[-1,1]$. g is said to be Lipschitz of order $\alpha \in (0,1]$ if there exists a positive constant c such that*

$$\mid g(x) - g(y) \mid \leq c \mid x - y \mid^{\alpha},$$

for all $x, y \in [-1,1]$. For convenience, we write $g \in Lip(\alpha, c)$.

To study the asymptotic behaviours of the estimator (5.1.3) inside the interval $[-1,1]$, the following assumption is considered:

$(\mathcal{A}_1)$ F is of class C^2 on $[-1,1]$.

$(\mathcal{A}_2)$ f and f' are bounded.

Throughout this chapter, we let $i = 1 \dots m$, $x \in [-1,1]$ for $m \geq 1$, and we consider the following notations:

$\theta_i = (2i-1)\pi/2m$, $\sigma^2(x) = F(x)(1-F(x))$, $x_i = \cos(\theta_i)$: Tchebytchev-Gauss points,

$A_m(x) = \sum_{i=1}^{m} F(x_i)\mathcal{L}_i(x)$, $b(x) = f(x)/2 + f'(x)(x-1)/4 - f''(x)(1+x^2-2x)/12$,

$\mathcal{L}_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{m} \frac{x - x_j}{x_i - x_j}$: Lagrange polynomial, $T_m(x) = \cos(m \arccos(x))$: Tchebytchev polynomial.

5.4 Main Results

Our first result is the following proposition which sets forward the bias and the variance of $\mathcal{F}_{n,m}$.

Proposition 5.2 (Bias and variance of $\mathcal{F}_{n,m}$). *Under assumption $(\mathcal{A}_1)$, we have for $x \in [-1,1]$,*

$$Bias(\mathcal{F}_{n,m}(x)) = \pi m^{-2} T_m(x) b(x) + o(m^{-2}), \tag{5.4.1}$$

$$Var(\mathcal{F}_{n,m}(x)) = n^{-1}\sigma^2(x) + O(n^{-1}m^{-1/2}). \tag{5.4.2}$$

Notice that for $x \in]0,1[$, the bias of the Bernstein estimator $\tilde{F}_{n,\nu}$ and the bias of the kernel estimator $\widehat{F}_n$ are given respectively by

$$Bias(\tilde{F}_{n,\nu}(x)) = \nu^{-1} b(x) + o(\nu^{-1}),$$

$$Bias(\widehat{F}_n(x)) = \frac{1}{2}h^2 f'(x)\mu_2(K) + o(h^2),$$

where $\mu_2(K) = \int_0^1 z^2 K(z)dz$. The previous result implies that, in the case when $\nu = m$, the bias of the estimator $\mathcal{F}_{n,m}$ is $O(m^{-2})$ is smaller than the one obtained using the Bernstein polynomial, which has a bias of order $O(m^{-1})$. On the one hand, if we consider $h = m^{-1}$ and f' is bounded, we notice that the bias of $\mathcal{F}_{n,m}$ is $O(m^{-2}) = O(h^2)$, which is asymptotically similar to the bias obtained using the kernel estimator $\widehat{F}_n$, that is generally $O(h^2)$ except near the boundaries. On the other hand, if f is bounded, it is well known that the variance of the Bernstein estimator and the variance of the kernel estimator are given respectively by

$$\begin{aligned} Var(\tilde{F}_{n,\nu}(x)) &= n^{-1}\sigma^2(x) + O(\nu^{-1/2}n^{-1}), \\ Var(\widehat{F}_n(x)) &= n^{-1}\sigma^2(x) + O(hn^{-1}). \end{aligned}$$

In this respect, another consequence of the previous result is that in the case when $\nu = m$, the variance of $\mathcal{F}_{n,m}$ is asymptotically similar to the variance of the estimator obtained using Bernstein polynomial. On the other side, in order to compare the proposed estimator and the kernel estimator, we consider some classical choices, which are $m = n$ and $h = n^{-1/3}$, this choice is motivated by the optimal bandwidth based on the minimization of the MSE. We notice that in the case where f is bounded and $x \in]0, 1[$, the variance of $\mathcal{F}_{n,m}$ is $n^{-1}\sigma^2(x) + O(n^{-3/2})$, which is asymptotically smaller than the variance obtained using kernel estimator, namely $n^{-1}\sigma^2(x) + O(n^{-4/3})$. In addition, it is well known that

$$MSE(\overline{F}_n(x)) = Var(\overline{F}_n(x)) = n^{-1}\sigma^2(x).$$

In conclusion, regarding the performance of the proposed estimator, we point out that

- The three considered estimators and the empirical distribution $\overline{F}_n$ are asymptotically equivalent in terms of MSE up to the first order.
- The proposed estimator asymptotically dominates the Bernstein estimator $\tilde{F}_{n,\nu}$ in terms of bias and in terms of MSE in the case when f is bounded.
- Under the assumption $(\mathcal{A}_2)$, the proposed estimator is asymptotically similar to the kernel estimator $\widehat{F}_n$ in terms of bias without any additional assumptions, and dominates the kernel estimator in terms of MSE under some classical conditions.

We complete our study with the following proposition which reveals that $\mathcal{F}_{n,m}$ is strongly consistent.

Proposition 5.3 (**Uniform convergence of $\mathcal{F}_{n,m}$**). *Under assumption* $(\mathcal{A}_1)$*, if* $n, m \to \infty$*, then*

$$\|\mathcal{F}_{n,m} - F\|_\infty \to 0 \quad a.s.$$

In this chapter, we prove that the estimator $\mathcal{F}_{n,m}$ satisfied the Chung-Smirnov property, which quantifies its extreme fluctuations about F, as $m \to \infty$, under certain regularity conditions on F. Let G_n be any estimator of the distribution function F. Therefore, G_n is said to satisfy the Chung-Smirnov property when

$$\overline{\lim_{n\to\infty}} \left(\frac{2n}{\log\log n}\right)^{1/2} \sup_{x\in[-1,1]} |G_n(x) - F(x)| \quad \leq \quad 1, \text{ a.s.} \tag{5.4.3}$$

We know that the empirical distribution function $\overline{F}_n$ satisfies the above property. To be more accurate, we have

$$\overline{\lim_{n\to\infty}} \left(\frac{2n}{\log\log n}\right)^{1/2} \sup_{x\in[-1,1]} \left|\overline{F}_n(x) - F(x)\right| = 1. \tag{5.4.4}$$

This was proved by Chung (1949) and Smirnov (1944). The following proposition demonstrates that $\mathcal{F}_{n,m}$ satisfies this property under certain conditions.

Proposition 5.4 (**Chung Smirnov property for $\mathcal{F}_{n,m}$**). *Let* $F \in Lip(\alpha, c)$ *for some* $c > 0$*. If* $m, n \to \infty$ *and* $\sqrt{n}m^{-\alpha/2} \to 0$*, then* $\mathcal{F}_{n,m}$ *satisfies equation (5.4.3).*

Finally, the following proposition indicates the asymptotic normality of the estimator (5.1.3).

Proposition 5.5 (**Asymptotic normality of $\mathcal{F}_{n,m}$**). *Assume* $(\mathcal{A}_1)$ *holds and* $m, n \to \infty$*. For* $x \in (-1, 1)$*, we have*

$$n^{1/2}\left(\mathcal{F}_{n,m}(x) - A_m(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(x)).$$

Note that, under an appropriate choice of bandwidth, a result similar to proposition 5.5 was recorded by Watson and Leadletter (1964) for general kernel estimators, and by Leblanc (2012a) for the Bernstein estimator of distribution functions.

5.5 Application

5.5.1 Simulations

In this section, we investigate the performance of the proposed estimator in estimating different distributions by comparing it to the performances of Bernstein estimator and

of the standard Gaussian kernel estimator. We can apply Bernstein estimator and the proposed estimator when the sample is concentrated on the intervals $[0,1]$ and $[-1,1]$, respectively. In order to enact the comparison between the estimators (5.1.1), (5.1.2) and (5.1.3), applicable in general, we list below suggested transformations in different cases:

(1) Suppose that X is concentrated on a finite support $[a,b]$, then we work with the sample values $Y_1,\ldots,Y_n$ where $Y_i=(X_i-a)/(b-a)$.

(2) For the distributions functions concentrated on $\mathbb{R}$, we can use the transformed sample $Y_i=1/2+\pi^{-1}\arctan(X_i)$ which transforms the range to the interval $[0,1]$.

(3) For the support $\mathbb{R}_+$, we can use the transformed sample $Y_i=X_i/(1+X_i)$ which transforms the range to the interval $[0,1]$.

In our simulation study, six sample sizes are considered, $n=10$, $n=50$, $n=100$, $n=150$, $n=200$, $n=250$ and the following distribution functions:

1− The beta distribution $\mathcal{B}(3,2)$,

2− The beta distribution $\mathcal{B}(2,2)$,

3− The gamma distribution $\mathcal{G}(1,6)$,

4− The mixture beta distribution $0.5\mathcal{B}(2.5,6)+0.5\mathcal{B}(9,1)$.

For each distribution function and sample size n, we compute the Integrated Squared Error (ISE) of the estimator over $N=500$ trials,

$$ISE[\overline{F}]=\int_0^1\Big(\widehat{F}(x)-F(x)\Big)^2\,dx, \tag{5.5.1}$$

where $\widehat{F}$ is an estimator of the distribution F. To select the smoothing parameters m, ν and h , we consider the Monte Carlo procedure for each point $x\in[0,1]$. We determine the parameters m (for $1\leq m\leq 300$), ν (for $1\leq\nu\leq 300$) and h (for $h=i/1000$ with $1\leq i\leq 300$), which minimizes ISE, which is approximated by

$$\frac{1}{N}\sum_{i=1}^{N}ISE_i(\widehat{F}),$$

where $ISE_i(\widehat{F})$ is the value of ISE computed from the ith sample of size n and obtained from (6.4.1).

Table 5.1: ISE for $N = 500$ trials of $\tilde{F}_{n,\nu}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for $n = 10, 50, 100$ I
[ISE for $N = 500$ trials of Bernstein estimator, standard Gaussian Kernel estimator and the proposed estimator $\mathcal{F}_{n,m}$, for $n = 10$, $n = 50$ and $n = 100$. The bold values indicates the smallest values of ISE.]

	n	m_{opt}	Lagrange	Bernstein	Kernel
$\mathcal{B}(3,2)$	10	10	0.032331	**0.013258**	0.019944
	50	50	**0.003819**	0.004411	0.005014
	100	100	**0.002198**	0.002431	0.002999
$\mathcal{B}(2,2)$	10	10	0.009598	**0.006958**	0.012854
	50	50	**0.001302**	0.001717	0.002420
	100	100	$\mathbf{0.564e^{-3}}$	$0.802e^{-3}$	0.001132
$\mathcal{G}(1,6)$	10	10	**0.037654**	0.038798	0.040357
	50	50	**0.005205**	0.006879	0.006393
	100	100	**0.001780**	0.002236	0.002052
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	10	10	0.005359	**0.003579**	0.007807
	50	50	**0.001326**	0.001515	0.001767
	100	100	$\mathbf{0.699e^{-3}}$	$0.727e^{-3}$	$0.820e^{-3}$

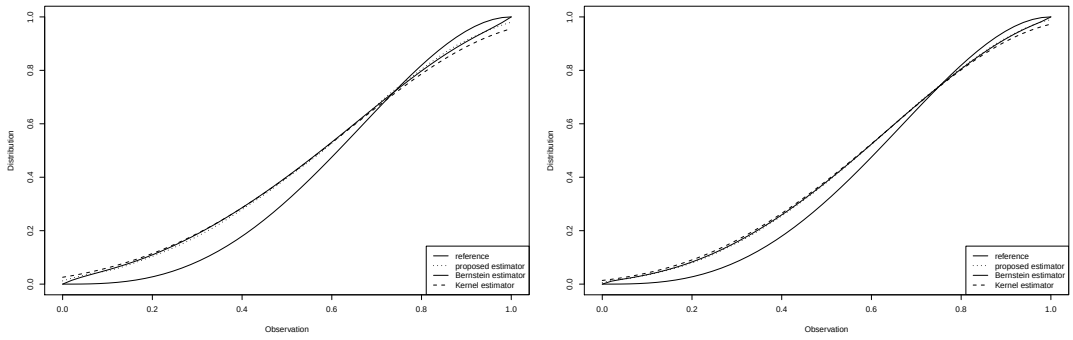

Figure 5.1: Qualitative comparison between $\tilde{F}_{n,\nu}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$ I
[Qualitative comparison between the Bernstein estimator $\tilde{F}_{n,\nu}$, the kernel estimator $\widehat{F}_n$ and the proposed distribution estimator $\mathcal{F}_{n,m}$, for $N = 500$ samples of size $n = 50$ (left panel) and of size $n = 100$ (right panel) for the beta distribution $\mathcal{B}(3,2)$.]

From figures 6.1-6.2 and tables 6.1-6.2, we conclude that

- In the considered distributions (1)-(4), by choosing the appropriate m, ν and h, the ISE of the distribution estimator (5.1.3) is smaller than that of Kernel estimator

Table 5.2: ISE for $N = 500$ trials $\tilde{F}_{n,\nu}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for $n = 150, 200, 250$ II
[ISE for $N = 500$ trials of Bernstein estimator $\overline{F}_{n,\nu}$, standard Gaussian kernel estimator $\widehat{F}_n$ and the proposed estimator $\mathcal{F}_{n,m}$, for $n = 150$, $n = 200$ and $n = 250$. The bold values indicates the smallest values of ISE.]

	n	m_{opt}	Lagrange	Bernstein	Kernel
	150	150	**0.001799**	0.002023	0.002342
$\mathcal{B}(3,2)$	200	200	**0.001596**	0.001782	0.001763
	250	250	**0.001258**	0.001447	0.001462
	150	150	$\mathbf{0.377e^{-3}}$	$0.489e^{-3}$	$0.718e^{-3}$
$\mathcal{B}(2,2)$	200	200	$\mathbf{0.264e^{-3}}$	$0.327e^{-3}$	$0.522e^{-3}$
	250	250	$\mathbf{0.229e^{-3}}$	$0.289e^{-3}$	$0.392e^{-3}$
	150	150	$\mathbf{0.540e^{-3}}$	$0.896e^{-3}$	$0.676e^{-3}$
$\mathcal{G}(1,6)$	200	200	$\mathbf{0.107e^{-3}}$	$0.200e^{-3}$	$0.115e^{-3}$
	250	250	$\mathbf{2.921e^{-5}}$	$4.996e^{-5}$	$5.429e^{-5}$
	150	150	$0.503e^{-3}$	$\mathbf{0.501e^{-3}}$	$0.568e^{-3}$
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	200	200	$\mathbf{0.379e^{-3}}$	$0.380e^{-3}$	$0.486e^{-3}$
	250	250	$\mathbf{0.309e^{-3}}$	$\mathbf{0.309e^{-3}}$	$0.354e^{-3}$

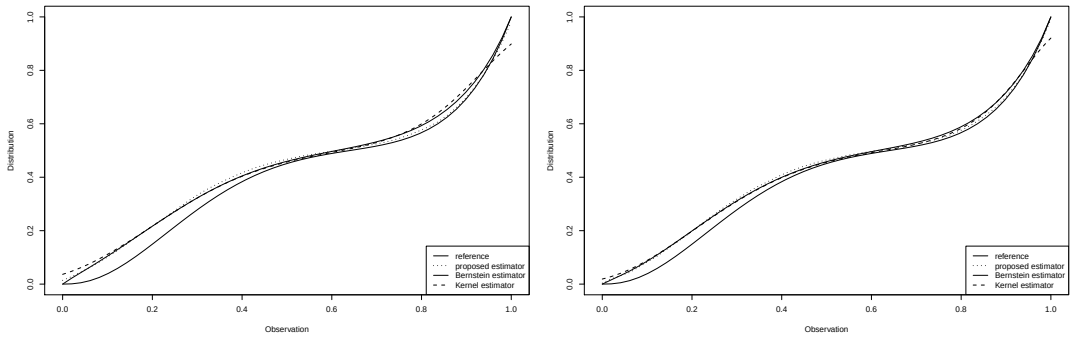

Figure 5.2: Qualitative comparison between $\tilde{F}_{n,\nu}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$ II
[Qualitative comparison between the Bernstein estimator $\tilde{F}_{n,\nu}$, the kernel estimator $\widehat{F}_n$ and the proposed distribution estimator $\mathcal{F}_{n,m}$, for $N = 500$ samples of size $n = 50$ (left panel) and of size $n = 100$ (right panel) for the exponential distribution $0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$.]

(5.1.1) and Bernstein estimator (5.1.2) even when the sample size is very large.

- The ISE decreases as the sample size increases.

5.5.2 Real dataset

We consider two examples that highlight the features of the proposed estimator $\mathcal{F}_{n,m}$:

1. At first time, the data show 50 alignments of a coding `DNA` sequence of the growth factor receptor of a Norwegian rat `EGFR` (Rattus norvegicus egfr gene, partial cds), which is available in the site `https://www.ncbi.nlm.nih.gov/`. For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 50$ for our proposed estimator, $m = 35$ for the Bernstein estimator and $h = 0.636438$ for the kernel estimator.

2. At the second time, we used `Salvister` data which appear in `R` package `kerdiest` (Quintela-del-Río and Estévez-Pérez (2012)). These data contain 85 observations of the annual peak instantaneous flow levels of the Salt River near Roosevelt, AZ, USA, for the period 1924-2009, obtained from the National Water Information System. For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 85$ for our proposed estimator, $m = 80$ for the Bernstein estimator and $h = 0.06$ for the kernel estimator.

3. The third data show the failure time (breakdowns of electronic devices) in operating hours. These data contain 18 observations and are introduced by Wang (2000). For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 18$ for our proposed estimator, $m = 15$ for the Bernstein estimator and $h = 0.20559$ for the kernel estimator.

4. Moreover, we used `attenu` data which appear in `R` package `datasets` (Joyner and Boore (2004)). These data contain 182 observations of the numeric moment magnitude at various stations for 23 earthquakes in California. For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 182$ for our proposed estimator, $m = 180$ for the Bernstein estimator and $h = 0.0305$ for the kernel estimator.

In the real examples, the three estimators are compared with the empirical distribution $\overline{F}_n$. Then, for any considered estimator $\widehat{F}$ of the distribution function F, we propose to compute the ISE defined as:

$$ISE(\widehat{F}) = \int_0^1 \left(\widehat{F}(x) - \overline{F}_n(x)\right)^2 dx.$$

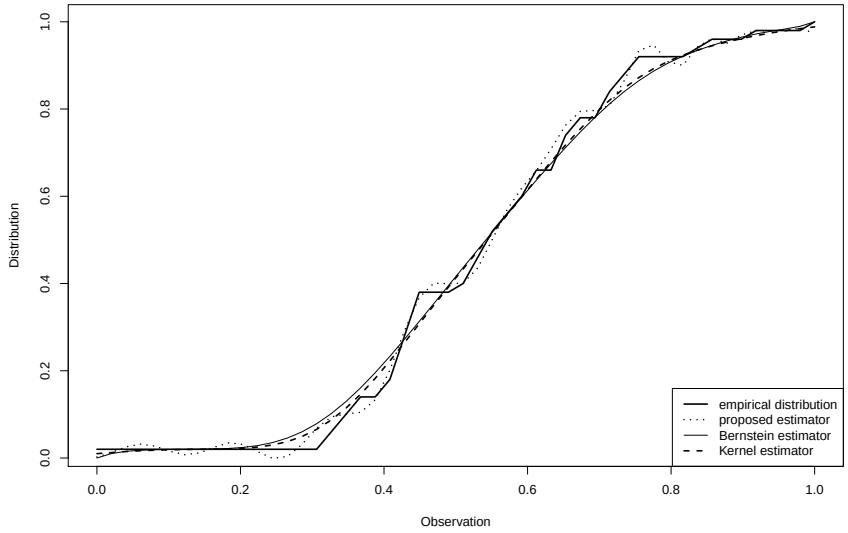

Figure 5.3: Qualitative comparison between $\tilde{F}_{n,\nu}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for `S A DNA` data. [Qualitative comparison between the Bernstein estimator $\tilde{F}_{n,\nu}$, the kernel estimator $\widehat{F}_n$ and the proposed distribution estimator $\mathcal{F}_{n,m}$, for `S A DNA` data.]

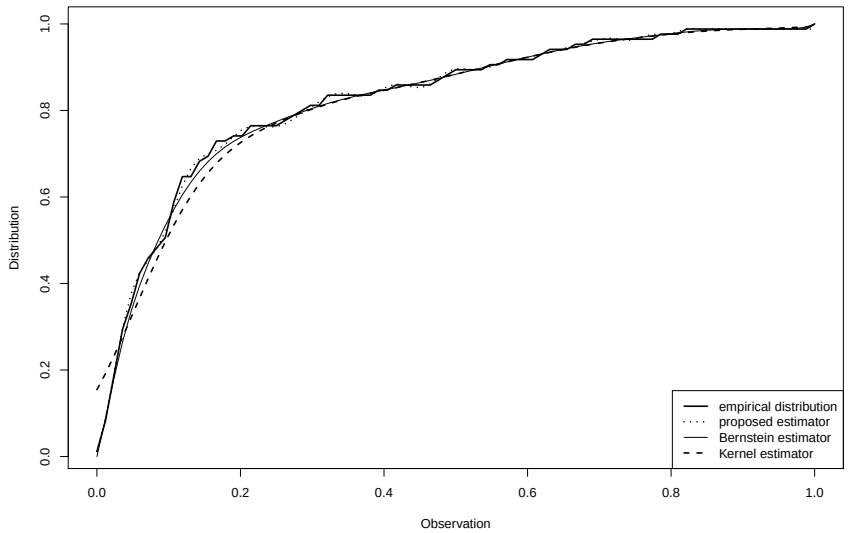

Figure 5.4: Qualitative comparison between $\tilde{F}_{n,\nu}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for `Saltriver` data. [Qualitative comparison between the Bernstein estimator $\tilde{F}_{n,\nu}(x)$, the kernel estimator $\widehat{F}_n$ and the proposed distribution estimator $\mathcal{F}_{n,m}$, for `Saltriver` data.]

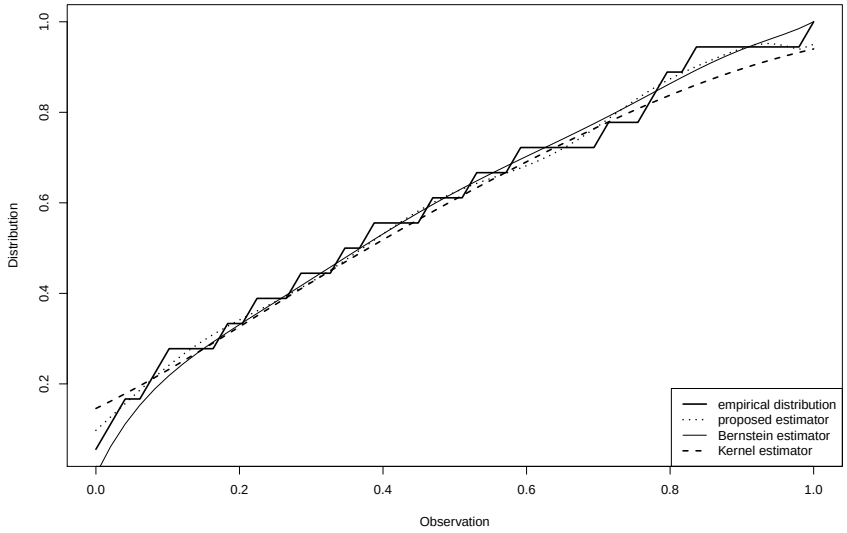

Figure 5.5: Qualitative comparison between $\tilde{F}_{n,\nu}(x)$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for `failure` data. [Qualitative comparison between the Bernstein estimator $\tilde{F}_{n,\nu}(x)$, the kernel estimator $\widehat{F}_n$ and the proposed distribution estimator $\mathcal{F}_{n,m}$, for `failure time` data.]

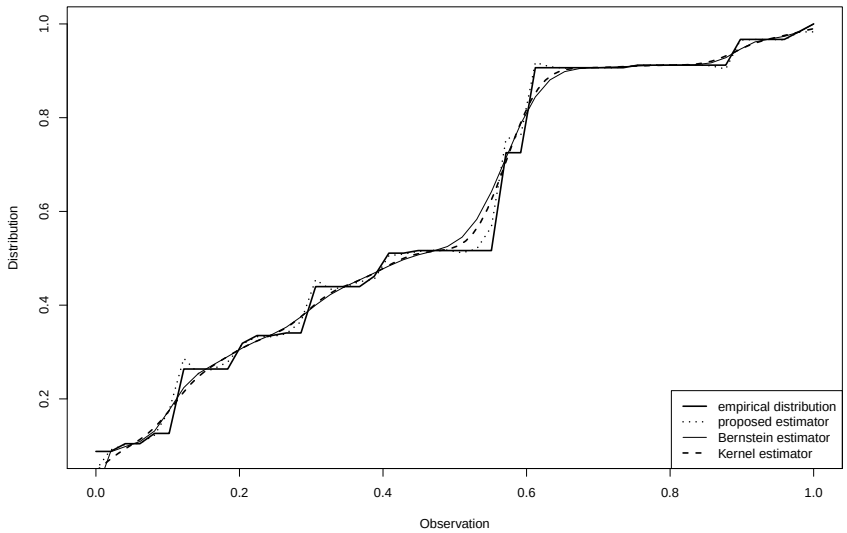

Figure 5.6: Qualitative comparison between $\tilde{F}_{n,\nu}(x)$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for `magnitude` data. [Qualitative comparison between the Bernstein estimator $\tilde{F}_{n,\nu}(x)$, the kernel estimator $\widehat{F}_n$ and the proposed distribution estimator $\mathcal{F}_{n,m}$, for `magnitude` data.]

Table 5.3: ISE of $\tilde{F}_{n,m}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for `S A DNA` data and `Saltriver` data. [ISE of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $\mathcal{F}_{n,m}$, for `S A DNA` data and `Saltriver` data. The bold values indicates the smallest values of ISE.]

Data set	Proposed estimator	Bernstein estimator	Kernel estimator
S A DNA	$\mathbf{0.334860e^{-3}}$	$0.647956e^{-3}$	$0.466513e^{-3}$
Saltriver	$\mathbf{4.586125e^{-5}}$	$0.112049e^{-3}$	$0.756750e^{-3}$

Table 5.4: ISE of $\tilde{F}_{n,m}$, $\widehat{F}_n$ and $\mathcal{F}_{n,m}$, for `magnitude` data and `failure` time data. [ISE of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $\mathcal{F}_{n,m}$, for magnitude data and failure time data. The bold values indicates the smallest values of ISE.]

Data set	Proposed estimator	Bernstein estimator	Kernel estimator
Failure time	**0.000549**	0.000760	0.001088
Magnitude	**0.000223**	0.000929	0.000695

Departing from Tables 6.3-5.4 and figures 5.3-5.6, we infer that the ISE of the proposed estimator is smaller than the ISE of the Bernstein estimator and the ISE of the kernel estimator, thus demonstrating the effectiveness of our considered estimator. We remark also from figures 5.3-5.6, that the curve of the proposed estimator passes exactly through the Lagrange polynomial interpolation points.

5.6 Conclusion

The central focus of this chapter is upon suggesting an estimator of the distribution function using Lagrange polynomials and Tchebytchev-Gauss points. We showed that a few important properties contributing to the popularity of kernel estimator and Bernstein estimator of distribution function are also satisfied by the proposed estimator. The asymptotic laws of the proposed estimator are established under general conditions. We also argued that the proposed estimator asymptotically dominates the Bernstein estimator in terms of bias. Through a simulation study and a simple data set examples, we have demonstrated how the proposed estimator can lead to satisfactory estimates of the distribution function. To sum up, our simulations also suggest that the proposed estimator is

quite promising and interesting as it behaves well when compared with both the Bernstein estimator and the Gaussian kernel estimator. To this extent, we could simply assert that our work is a step may be taken further as it lays the ground and offers new perspectives for future works to extend this investigation by considering a recursive version and compare the obtained estimators to the one adopted by Slaoui (2014) and Jmaei et al. (2017). We plan also to consider the estimation of a density function in a recursive framework and then the estimation of a regression function in a recursive framework by using Lagrange polynomials (see Slaoui (2015c), Slaoui (2016)).

5.7 Appendix

Throughout the proofs, we use the following notations:

$$R_m^{(1)} = \sum_{k=1}^{m} \frac{\sin\theta_k}{\sin(m\theta_k)},\ R_m^{(2)} = \sum_{k=1}^{m} \frac{x_k \sin\theta_k}{\sin(m\theta_k)},\ R_m^{(3)} = \sum_{k=1}^{m} \frac{\cos(2\theta_k)\sin\theta_k}{\sin(m\theta_k)},$$

$J_m(x) = \sum_{k=1}^{m} \mid x_k - x \mid \mathcal{L}_k^2(x)$, $S_m(x) = \sum_{k=1}^{m} \mathcal{L}_k^2(x)$, $\lambda_m(x) = \sum_{k=1}^{m} |\mathcal{L}_k(x)|$ Lebesgue function,

$\Lambda_m = \max_{x\in[-1,1]} \lambda_m(x)$ lebesgue constant, for $j \in \{0,1,2\}$, $P_{j,m}(x) = \sum_{\substack{k=1 \\ k<l}}^{m} (x_k - x)^j \mathcal{L}_k(x)\mathcal{L}_l(x)$.

In order to prove Theorems 6.1-6.3, we establish the following technical lemmas 5.1 and 5.2 stated below.

Lemma 5.1. *For $m \geq 1$, we have*

$$R_m^{(1)} = \sin(\pi/2m),\ R_m^{(2)} = \sin(\pi/m)/2,\ R_m^{(3)} = (\sin(3\pi/2m) - \sin(\pi/2m))/2.$$

Lemma 5.2. *For $x \in [-1,1]$, we have*

i)

$$\sum_{k=1}^{m} (x_k - x)\mathcal{L}_k(x) = -\frac{T_m(x)}{m} R_m^{(1)} = -\frac{\pi}{2m^2} T_m(x) + o(m^{-2}),$$

ii)

$$\sum_{k=1}^{m} (x_k - x)^2 \mathcal{L}_k(x) = \frac{T_m(x)}{m}\left(xR_m^{(1)} - R_m^{(2)}\right) = \frac{\pi}{2m^2} T_m(x)(x-1) + o(m^{-2}),$$

iii)

$$\begin{aligned}\sum_{k=1}^{m}(x_k-x)^3\mathcal{L}_k(x) &= -\frac{T_m(x)}{m}\left(\frac{R_m^{(1)}+R_m^{(3)}}{2}+x^2R_m^{(1)}-2xR_m^{(2)}\right)\\ &= -\frac{\pi}{2m^2}T_m(x)(1+x^2-2x)+o(m^{-2}).\end{aligned}$$

Proof: [Proof of lemma 1] We first note that $R_m^{(1)}$, $R_m^{(2)}$ and $R_m^{(3)}$ can be rewritten as:

$$\begin{aligned}R_m^{(1)} &= \frac{-\cos(\pi/2m)}{2}\sum_{k=1}^{m}\left[\sin\left(k\left(\pi/m+\pi\right)\right)+\sin\left(k\left(\pi/m-\pi\right)\right)\right]\\ &\quad+\sin(\pi/2m)/2\sum_{k=1}^{m}\left[\cos\left(k\left(\pi/m-\pi\right)\right)+\cos\left(k\left(\pi/m+m\right)\right)\right].\end{aligned}$$

$$\begin{aligned}R_m^{(2)} &= \frac{-1}{4}\cos(\pi/m)\sum_{k=1}^{m}\sin(k(2\pi/m+\pi))-\frac{1}{4}\cos(\pi/m)\sum_{k=1}^{m}\sin(k(2\pi/m-\pi))\\ &\quad+\frac{1}{4}\sin(\pi/m)\sum_{k=1}^{m}\cos(k(2\pi/m-\pi))+\frac{1}{4}\sin(\pi/m)\sum_{k=1}^{m}\cos(k(2\pi/m+\pi)).\\ R_m^{(3)} &= -\frac{1}{4}\sum_{k=1}^{m}\sin(3\theta_k+k\pi)+\sin(3\theta_k-k\pi)-\frac{1}{4}\sum_{k=1}^{m}\sin(-\theta_k+k\pi)+\sin(-\theta_k-k\pi).\end{aligned}$$

Using for $t\in\mathbb{R}$,

$$\sum_{k=1}^{m}\cos(kt)=\frac{\cos(mt/2)\sin((m+1)t/2)}{\sin(t/2)}\text{ and }\sum_{k=1}^{m}\sin(kt)=\frac{\sin(mt/2)\sin((m+1)t/2)}{\sin(t/2)},$$

some classical computations provide

$$\begin{aligned}R_m^{(1)} &= \sin(\pi/2m)=\frac{\pi}{2m}+o(m^{-1}),\\ R_m^{(2)} &= \frac{1}{2}\sin(\pi/m)=\frac{\pi}{2m}+o(m^{-1}),\\ R_m^{(3)} &= \frac{1}{2}(\sin(3\pi/2m)-\sin(\pi/2m))=\frac{\pi}{2m}+o(m^{-1}).\end{aligned}$$

Proof: [Proof of lemma 2] First, we have (see Feldheim (1939) page 10)

$$\mathcal{L}_k(x)=\frac{T_m(x)}{T_m'(x_k)(x-x_k)}\tag{5.7.1}$$

and $-\sin\theta_k T_m'(\cos\theta_k)=-m\sin(m\theta_k)$. It follows that

$$\sum_{k=1}^{m}(x_k-x)\mathcal{L}_k(x)=-\frac{T_m(x)}{m}R_m^{(1)}.$$

$$\sum_{k=1}^{m}(x_k - x)^2\mathcal{L}_k(x) = \frac{xT_m(x)}{m}R_m^{(1)} - \frac{T_m(x)}{m}R_m^{(2)},$$

$$\sum_{k=1}^{m}(x_k - x)^3\mathcal{L}_k(x) = \frac{-T_m(x)}{2m}\left(R_m^{(1)} + R_m^{(3)}\right) - \frac{T_m(x)x^2}{m}R_m^{(1)} + 2\frac{T_m(x)}{m}R_m^{(2)}.$$

5.7.1 Proof of proposition 6.1

Clearly, we have

$$\mathbb{E}(\mathcal{F}_{n,m}(x)) = A_m(x).$$

The expansion of Taylor-Young ensures that for $1 \leq k \leq m$,

$$\begin{aligned}\mathbb{E}(\mathcal{F}_{n,m}(x)) &= F(x) + f(x)\sum_{k=1}^{m}(x_k - x)\mathcal{L}_k(x) + \frac{f'(x)}{2}\sum_{k=1}^{m}(x_k - x)^2\mathcal{L}_k(x) \\ &\quad + \frac{f''(x)}{6}\sum_{k=1}^{m}(x_k - x)^3\mathcal{L}_k(x) + o\left(\sum_{k=1}^{m}(x - x_k)^3\mathcal{L}_k(x)\right).\end{aligned}$$

The application of Lemma 5.1 together with Lemma 5.2 yield the equation (5.4.1). Let's now focus on calculating the variance of our estimator.

First, we set $(\eta_i)_{1\leq i\leq n} = \left(\sum_{k=1}^{m}\left(\mathbb{1}_{\{X_i\leq x_k\}} - F(x_k)\right)\mathcal{L}_k(x)\right)_{1\leq i\leq n}$, it comes that

$$\begin{aligned}\mathcal{F}_{n,m}(x) - A_m(x) &= \sum_{k=1}^{m}\left(\overline{F}_n(x_k) - F(x_k)\right)\mathcal{L}_k(x) \\ &= \frac{1}{n}\sum_{k=1}^{m}\left(\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leq x_k\}} - F(x_k)\right)\mathcal{L}_k(x) \\ &= \frac{1}{n}\sum_{i=1}^{n}\eta_i.\end{aligned}$$

Moreover, since $\mathbb{E}\left(\mathcal{F}_{n,m}(x) - A_m(x)\right) = 0$, it follows that

$$\begin{aligned}Var\left(\mathcal{F}_{n,m}(x) - A_m(x)\right) &= \mathbb{E}\left[\left(\mathcal{F}_{n,m}(x) - A_m(x)\right)^2\right] \\ &= Var\left(\mathcal{F}_{n,m}(x)\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}Var(\eta_i) \\ &= \frac{1}{n}\mathbb{E}(\eta_1^2).\end{aligned}$$

Now, we define for any $x \in [-1,1]$ and for $i \geq 1$, $\varphi_i(x) = \mathbb{1}_{\{X_i \leq x\}} - F(x)$. We infer that

$$\begin{aligned}
\mathbb{E}(\eta_i^2) &= \mathbb{E}\left[\left(\sum_{k=1}^{m} \varphi_i(x_k)\mathcal{L}_k(x)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{k,l=1}^{m} \varphi_i(x_k)\mathcal{L}_k(x)\varphi_i(x_l)\mathcal{L}_l(x)\right] \\
&= \sum_{k,l=1}^{m} \mathbb{E}\left[\varphi_i(x_k)\varphi_i(x_l)\right]\mathcal{L}_k(x)\mathcal{L}_l(x). \qquad (5.7.2)
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\mathbb{E}\left[\varphi_i(x_k)\varphi_i(x_l)\right] &= \mathbb{E}\left[\left(\mathbb{1}_{\{X_i \leq x_k\}} - F(x_k)\right)\left(\mathbb{1}_{\{X_i \leq x_l\}} - F(x_l)\right)\right] \\
&= \mathbb{E}\left(\mathbb{1}_{\{X_i \leq x_k\}}\mathbb{1}_{\{X_i \leq x_l\}}\right) - F(x_k)F(x_l) \\
&= \mathbb{E}\left(\mathbb{1}_{\{X_i \leq \min(x_k,x_l)\}}\right) - F(x_k)F(x_l) \\
&= F(\min(x_k,x_l)) - F(x_k)F(x_l) \\
&= \min(F(x_k),F(x_l)) - F(x_k)F(x_l).
\end{aligned}$$

Substituting this result for (5.7.2) leads to

$$\begin{aligned}
\mathbb{E}(\eta_i^2) &= \sum_{k,l=1}^{m} \left[\min(F(x_k),F(x_l)) - F(x_k)F(x_l)\right]\mathcal{L}_k(x)\mathcal{L}_l(x) \\
&= \sum_{k=1}^{m} F(x_k)\mathcal{L}_k^2(x) + 2\sum_{\substack{k=1 \\ k<l}}^{m} F(x_k)\mathcal{L}_k(x)\mathcal{L}_l(x) - A_m(x)^2. \qquad (5.7.3)
\end{aligned}$$

We need now to find an asymptotic expression for (5.7.3). For this reason, we first expand $F(x_k)$ about x to state that for all $0 \leq k \leq m$, $F(x_k) = F(x) + O(|x_k - x|)$. This allows us to write the first term of (5.7.3) as

$$\begin{aligned}
\sum_{k=1}^{m} F(x_k)\mathcal{L}_k^2(x) &= \sum_{k=1}^{m} \left[F(x) + O(|x_k - x|)\right]\mathcal{L}_k^2(x) \\
&= \sum_{k=1}^{m} F(x)\mathcal{L}_k^2(x) + \sum_{k=1}^{m} O(|x_k - x|\,\mathcal{L}_k^2(x)) \\
&= F(x)S_m(x) + O(J_m(x)),
\end{aligned}$$

where $J_m(x) = \sum_{k=1}^{m} \mid x_k - x \mid \mathcal{L}_k^2(x)$.

For the second term of (5.7.3), we instead write $F(x_k)$ as

$$F(x_k) = F(x) + (x_k - x)f(x) + O((x_k - x)^2).$$

Moreover, we have

$$2P_{0,m}(x) + S_m(x) = \sum_{k,l=1}^{m} \mathcal{L}_k(x)\mathcal{L}_l(x).$$

Since

$$\sum_{k,l=1}^{m} \mathcal{L}_k(x)\mathcal{L}_l(x) = 1,$$

it comes that

$$P_{0,m}(x) = \frac{1}{2}(1 - S_m(x)).$$

Then

$$\begin{aligned}
\sum_{\substack{k=1\\k<l}}^{m} F(x_k)\mathcal{L}_k(x)\mathcal{L}_l(x) &= \sum_{\substack{k=1\\k<l}}^{m} \left(F(x) + (x_k - x)f(x) + O((x_k - x)^2)\right)\mathcal{L}_k(x)\mathcal{L}_l(x) \\
&= \sum_{\substack{k=1\\k<l}}^{m} F(x)\mathcal{L}_k(x)\mathcal{L}_l(x) + \sum_{\substack{k=1\\k<l}}^{m} (x_k - x)f(x)\mathcal{L}_k(x)\mathcal{L}_l(x) \\
&\quad + \sum_{\substack{k=1\\k<l}}^{m} O\left((x_k - x)^2\mathcal{L}_k(x)\mathcal{L}_l(x)\right) \\
&= F(x)P_{0,m}(x) + f(x)P_{1,m}(x) + O(P_{2,m}(x)) \\
&= \frac{1}{2}F(x)(1 - S_m(x)) + f(x)P_{1,m}(x) + O(P_{2,m}(x)).
\end{aligned}$$

Moreover, we have

$$2\sum_{\substack{k=1\\k<l}}^{m} F(x_k)\mathcal{L}_k(x)\mathcal{L}_l(x) = F(x)(1 - S_m(x)) + 2f(x)P_{1,m}(x) + O(m^{-4}).$$

Therefore,

$$\begin{aligned}
\mathbb{E}(\eta_i^2) &= F(x) + 2f(x)P_{1,m}(x) + O(J_m(x)) + O(P_{2,m}(x)) - A_m^2(x) \\
&= F(x)(1 - F(x)) + 2f(x)P_{1,m}(x) + O(J_m(x)) + O(m^{-4})
\end{aligned}$$

$$= \sigma^2(x) + 2f(x)P_{1,m}(x) + O(J_m(x)) + O(P_{2,m}(x)). \tag{5.7.4}$$

Now, using Cauchy-Schwartz's inequality combined with the fact that $|\mathcal{L}_k(x)| \leq 1$, we get

$$\begin{aligned} |J_m(x)| &= \left| \sum_{k=1}^{m} |x_k - x| \, \mathcal{L}_k^2(x) \right| \\ &\leq \left| \left(\sum_{k=1}^{m} (x_k - x)^2 \mathcal{L}_k(x) \right) \right|^{1/2} \left| \left(\sum_{k=1}^{m} \mathcal{L}_k^3(x) \right) \right|^{1/2} \\ &\leq \left[\left(\frac{\pi}{m^2} + o\left(m^{-2}\right) \right) S_m(x) \right]^{1/2}. \end{aligned}$$

On the other side, using the fact that $\Lambda_m \leq \frac{2}{\pi} \ln(m+1) + 1$ (see Brutman (1978)), we obtain

$$\begin{aligned} S_m(x) &\leq \left(\sum_{k=1}^{m} |\mathcal{L}_k(x)| \right)^2 \\ &\leq \Lambda_m^2 \leq \frac{4}{\pi^2} \ln(m+1)^2 + 1 + \frac{4}{\pi} \ln(m+1). \end{aligned}$$

As a matter of fact, we infer that $J_m(x) = O(m^{-1/2})$. Now, it follows from (5.7.1), that

$$\begin{aligned} P_{2,m}(x) &= [(1-x)^2 + O(m^{-2})] \sum_{\substack{k=1 \\ k<l}}^{m} \mathcal{L}_k(x)\mathcal{L}_l(x) \\ &= [(1-x)^2 + O(m^{-2})] \sum_{\substack{k=1 \\ k<l}}^{m} O(m^{-2}). \end{aligned}$$

It follows that $P_{2,m}(x) = O(m^{-1})$. Moreover, we have

$$\begin{aligned} P_{1,m}(x) &= \sum_{k=1}^{m} (x_k - x)\mathcal{L}_k(x) \sum_{l=k+1}^{m} \mathcal{L}_l(x) \\ &= [1 - x + O(m^{-2})] \sum_{\substack{k=1 \\ k<l}}^{m} O(m^{-2}). \end{aligned}$$

Hence, we obtain $P_{1,m}(x) = O(m^{-1})$ and equation (5.4.2) follows.

5.7.2 Proof of proposition 5.3

We first use the fact that

$$\|\mathcal{F}_{n,m} - F\| \leq \|\mathcal{F}_{n,m} - A_m\| + \|A_m - F\|.$$

The use of Jackson's theorem, ensures that

$$\lim_{m\to\infty} \|A_m - F\| = 0.$$

Moreover, we have

$$\mathcal{F}_{n,m}(x) - A_m(x) = \sum_{k=1}^{m} \left(\overline{F}_n(x_k) - F(x_k)\right) \mathcal{L}_k(x),$$

it comes that

$$\|\mathcal{F}_{n,m} - A_m\| \leq \max_{1\leq k\leq m} \left|\overline{F}_n(x_k) - F(x_k)\right|.$$

In addition, the application of Clivenco-Cantelli's theorem, ensures that

$$\lim_{n\to\infty} \left\|\overline{F}_n - F\right\| = 0,$$

which conclude the proof.

5.7.3 Proof of proposition 5.4

First, we note that for all $m \geq 1$,

$$\|\mathcal{F}_{n,m} - F\| \leq \left\|\overline{F}_n - F\right\| + \|A_m - F\|.$$

Moreover, as $F \in Lip(\alpha, c)$, Jackson (1911, 1912) implies that

$$\|A_m - F\| = O\left(\frac{\log(m)}{m^{\alpha/2}}\right).$$

It follows that

$$\overline{\lim_{n\to\infty}} u_n \|\mathcal{F}_{n,m} - F\| \leq \overline{\lim_{n\to\infty}} u_n \left\|\overline{F}_n - F\right\| + \overline{\lim_{n\to\infty}} u_n \|A_m - F\|,$$

where $u_n = (2n/\log\log n)^{1/2}$, for all $n \geq 1$. Now, using equation (5.4.4), we obtain

$$\overline{\lim_{n\to\infty}} u_n \left\|\overline{F}_n - F\right\| = 1 \text{ a.s.}$$

Moreover, since $n^{1/2}m^{-\alpha/2} \to 0$ when $n,\ m \to \infty$, we have

$$\begin{aligned}
\overline{\lim_{n\to\infty}} u_n \|A_m - F\| &= \overline{\lim_{n\to\infty}} \frac{(2n)^{1/2}}{(\log\log n)^{1/2}} \frac{\log m}{m^{\alpha/2}} \\
&= \overline{\lim_{n\to\infty}} \frac{\sqrt{n}}{m^{\alpha/2}} \\
&= 0.
\end{aligned}$$

It comes that, $\overline{\lim_{n\to\infty}} u_n \|\mathcal{F}_{n,m} - F\| \leq 1$. This completes the proof of proposition 5.4.

5.7.4 Proof of proposition 6.3

Since we have

$$\mathcal{F}_{n,m}(x) - A_m(x) = \frac{1}{n}\sum_{i=1}^{n} \eta_i.$$

It follows that,

$$n^{1/2}\left(\mathcal{F}_{n,m}(x) - A_m(x)\right) = \sum_{i=1}^{n} \frac{\eta_i}{n^{1/2}}.$$

Now, in order to check the Lindeberg condition, we notice for all $n \geq 1$ and for $i = 1, \ldots, n$

$$X_{i,n} = \frac{\eta_i}{n^{1/2}} \quad \text{and} \quad s_n^2 = \sum_{i=1}^{n} \mathbb{E}(X_{i,n}^2).$$

We have

$$n^{1/2}\left(\mathcal{F}_{n,m}(x) - A_m(x)\right) = \sum_{i=1}^{n} X_{i,n},$$

with $(X_{i,n})_{i\geq 1}$ is a sequence of i.i.d. random variables such that $\mathbb{E}(X_{i,n}) = 0$. Further, we have for $n \geq 1$,

$$s_n^2 = \sum_{i=1}^{n} \mathbb{E}(X_{i,n}^2) = \sum_{i=1}^{n} \frac{1}{n}\mathbb{E}(\eta_i^2) = \mathbb{E}(\eta_1^2).$$

However, in the light of (5.7.4), we have $\lim\limits_{n\to\infty} s_n^2 = \sigma^2(x)$. Indeed, using the Cauchy-Schwarz inequality, $\mathcal{L}_k(x) \leq 1$ and by inferring the proof of proposition 6.1, we get $\lim\limits_{n\to\infty} J_m(x) = 0$ and $\lim\limits_{n\to\infty} P_{1,m}(x) = 0$. Moreover, since $\sum\limits_{k=0}^{m} \mathcal{L}_k(x) = 1$, we have

$$\begin{aligned} |\eta_1| &= \left|\sum_{k=1}^{m} \left(\mathbb{1}_{\{X_i \leq x_k\}} - F(x_k)\right)\mathcal{L}_k(x)\right| \leq \sum_{k=1}^{m} \left|\mathbb{1}_{\{X_i \leq x_k\}} - F(x_k)\right|\mathcal{L}_k(x) \\ &\leq \sum_{k=1}^{m} (1+1)\mathcal{L}_k(x) = 2. \end{aligned}$$

It comes that

$$\begin{aligned} X_{1,n}^2 \mathbb{1}_{\frac{|X_{1,n}|}{s_n} > \varepsilon} &= \frac{\eta_1^2}{n}\mathbb{1}_{\{|\eta_1| > s_n n^{1/2}\varepsilon\}} \\ &\leq \frac{4}{n}\mathbb{1}_{\{|\eta_1| > s_n n^{1/2}\varepsilon\}}. \end{aligned}$$

Hence,

$$\sum_{i=1}^{n} \mathbb{E}\left[X_{i,m}^2 \mathbb{1}_{\frac{|X_{i,m}|}{s_n}>\varepsilon}\right] \leq \frac{4}{n} \sum_{i=1}^{n} \mathbb{1}_{\{|\eta_i|>s_n n^{1/2}\varepsilon\}}.$$

Moreover, we have $s_n^2 \to \sigma^2(x)$ when $n \to \infty$, then Lindeberg's condition

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left[X_{i,n}^2 \mathbb{1}_{\frac{|X_{i,n}|}{s_n}>\varepsilon}\right] = 0,$$

is fulfided. thus by Lindeberg-Feller's central limit theorem, we get

$$n^{1/2}\left(\mathcal{F}_{n,m}(x) - A_m(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(x)),$$

which concludes the proof.

Chapter 6

Recursive distribution estimator based on stochastic algorithm

Key words and phrases: Asymptotic properties, Distribution estimation, Lagrange polynomial, Stochastic approximation algorithm, Tchebychev-Gauss points.

Abstract 4: *In this chapter, our central focus is upon introducing a recursive distribution estimator using Robbins Monro's algorithm and Lagrange polynomial. Some asymptotic properties of the proposed estimator are investigated, such as its asymptotic bias and variance. The asymptotic normality is also established. We show that the proposed estimator outperforms the non recursive distribution function in terms of the asymptotic mean squared error. Generally, the proposed estimator compares favorably with other competitors in theoretical comparisons, a simulation study and a real data sets examples.*

6.1 Introduction

Distribution estimation is undoubtedly a beneficial tool of data analysis, occurring in multiple domains like actuarial sciences or finance, life sciences, engineering, under different names such as lifetime, survival distributions. Let $X_1 \dots X_n$ be a sequence of *i.i.d.* random variables having an underlying unknown distribution function F and associated density function f. Within the *parametric* framework, the model structure is already determined. For instance, the normal distribution form $\mathcal{N}(\mu, \sigma^2)$. The unique target is to estimate the parameters, namely μ and σ. In the *non parametric* setting, the model structure is not identified. In this chapter, our central focus is upon introducing a recursive distribution

estimator using Robbins Monro's algorithm and Lagrange polynomial. Some asymptotic properties of the proposed estimator are investigated, such as its asymptotic bias and variance. The asymptotic normality is also established. We show that the proposed estimator outperforms the non recursive distribution function in terms of the asymptotic mean squared error. Generally, the proposed estimator compares favorably with other competitors in theoretical comparisons, a simulation study and a real data sets examples. The application of Robbins–Monro's procedure to construct a stochastic approximation algorithm was introduced by Révész (1973, 1977) and extended by Tsybakov (1990). As examples for recursive estimators of the distribution function, we display the recursive kernel distribution estimator identified by Slaoui (2014)

$$F_{n,K}(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathcal{K}\left(\frac{x - X_k}{h_k}\right), \tag{6.1.1}$$

where $\mathcal{K}(z) = \int_{-\infty}^{z} K(u)du$, K is a kernel and (h_n) is a bandwidth; and the recursive distribution estimator using the Bernstein polynomial defined by Jmaei et al. (2017)

$$F_{n,B}(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathcal{B}_k(x), \tag{6.1.2}$$

where $\mathcal{B}_n(x) = \sum_{k=0}^{m} \mathbb{1}_{\{X_n \leq k/m\}} b_k(m, x)$ and $b_k(m, x) = C_m^k x^k (1-x)^{m-k}$ is the Bernstein polynomial of order $m > 0$. In this chapter, we construct a stochastic algorithm, which approximates the function F using Lagrange polynomial, at a given point x. We suppose that f is supported on $[-1, 1]$ and we define an algorithm of search of the zero of the function $\phi : y \to F(x) - y$. Using Robbins and Monro (1951) recursive scheme, we set:

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

(ii) For all $n \geq 1$, we set $F_n(x) = F_{n-1}(x) + \gamma_n J_n(x)$, where the stepsize (γ_n) is a positive sequence of real numbers decreasing to zero and (J_n) is a sequence of functions $J_n : \mathbb{R} \to \mathbb{R}$ defined by $J_n(x) = \phi(F_{n-1}(x)) - W_n + \beta_n$. Using the assumption of Robbins-Monro's procedure defined by $\mathbb{E}(W_n|\mathcal{F}_{n-1}) = 0$, where $\mathcal{F}_{n-1}$ stands for the σ-algebra of the events occurring up the time $n-1$, we have $\mathbb{E}(J_n(x)) = F(x) - F_{n-1}(x) + \beta_n$. Following the approach of Révész (1973, 1977) and noting that $\mathbb{E}\left(\sum_{k=1}^{m} \mathbb{1}_{\{X_n \leq x_k\}} L_k(x)\right) = F(x) + \xi_n(x)$, where $\xi_n(x)$ goes to zero as n goes to infinity, we set $J_n(x) = \sum_{k=1}^{m} \mathbb{1}_{\{X_n \leq x_k\}} L_k(x) - F_{n-1}(x)$.

Therefore, the recursive estimator F_n of the distribution function F at the point x can be started as

$$F_n(x) = (1-\gamma_n)F_{n-1}(x) + \gamma_n \sum_{k=1}^{m} \mathbb{1}_{\{X_n \leq x_k\}} L_k(x). \tag{6.1.3}$$

Moreover, we suppose that $F_0(x) = 0$. We let

$$\Pi_n = \prod_{j=1}^{n}(1-\gamma_j) \text{ and } Z_n(x) = \sum_{k=1}^{m} \mathbb{1}_{\{X_n \leq x_k\}} L_k(x).$$

Then, it follows from the Equation (6.1.3) that F_n can be rewritten as

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k Z_k(x). \tag{6.1.4}$$

The basic objective of this chapter is to investigate the properties of the recursive distribution estimator defined by the stochastic approximation algorithm (6.1.4). The assessment of its performance is fulfilled through enacting its comparison with the non recursive distribution estimator defined by Helali and Slaoui (2020) and the recursive estimators (6.1.1) - (6.1.2). The remainder of the chapter is organized as follows. In Section 6.2, we highlight the notations and assumptions. In section 6.3, we display the main results: the bias, variance and asymptotic normality. Section 6.4 is devoted to the application results obtained, first through simulations (subsection 6.4.1) and second using a real data set (subsection 6.4.2). We draw the conclusion in Section 6.5, and Section 6.6 exhibits the proofs of the theoretical results.

6.2 Assumptions and notations

We first define the following class of regularly varying sequences.

Definition 6.1. *Let $\gamma \in \mathbb{R}$ and $(\psi_n)_{n\geq 1}$ be a nonrandom positive sequence. We write $(\psi_n) \in \mathcal{GS}(\gamma)$, if*

$$\lim_{n\to\infty} n\left[1 - \frac{\psi_{n-1}}{\psi_n}\right] = \gamma. \tag{6.2.1}$$

Condition (6.2.1) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1973)). Note that the acronym $\mathcal{GS}$ stands for (Galambos and Seneta). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\log n)^b$, $n^{\gamma}(\log\log n)^b$, and so on. To investigate the asymptotic behaviours of the estimator (6.1.4), we make the following assumptions:

(A1) F is of class C^2 on $[-1,1]$.

(A2) f, f' and f'' are bounded.

(A3) $(\gamma_n) \in \mathcal{GS}(-\alpha)$, $\alpha \in \left(\frac{1}{2}, 1\right]$.

(A4) $(m_n) \in \mathcal{GS}(a)$, $a \in (0,1)$.

(A5) $\lim_{n\to\infty} n\gamma_n \in (\min(2a, (\alpha+a)/2), \infty)$.

Discussion on the assumptions:

- Assumptions **(A1)**-**(A2)** are standard in density estimation.
- Assumptions **(A3)**-**(A4)** gives conditions for the stepsize and the degree of the Lagrange polynomial which allows getting the bias and variance of the proposed estimator.
- Assumption **(A5)** on the limit of $(n\gamma_n)$ as $n \to \infty$ is usual within the framework of stochastic approximation algorithms. It implies, in particular, that the limit of $([n\gamma_n]^{-1})$ is finite.

Throughout this chapter, we let $i = 1 \dots m$, $x \in [-1,1]$ for $m \geq 1$, and we consider the following notations:

$$\Pi_n = \prod_{j=1}^{n}(1-\gamma_j),\ Z_n(x) = \sum_{k=1}^{m} \mathbb{1}_{\{X_n \leq x_k\}} \mathcal{L}_k(x),\ \xi = \lim_{n\to\infty}(n\gamma_n)^{-1},\ \theta_i = (2i-1)\pi/2m,$$

$$\sigma^2(x) = F(x)(1-F(x)),\ x_i = \cos(\theta_i)\text{: Tchebytchev-Gauss points},\ A_m(x) = \sum_{i=1}^{m} F(x_i)\mathcal{L}_i(x),$$

$$b(x) = f(x)/2 + f'(x)(x-1)/4 - f''(x)(1+x^2-2x)/12,\ \mathcal{L}_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{m} \frac{x-x_j}{x_i-x_j}\text{: Lagrange}$$

polynomial, $T_m(x) = \cos(m \arccos(x))$: Tchebytchev polynomial.

6.3 Main results

The first result is the bias, variance, and mean squared error (MSE) of F_n.

Proposition 6.1 (**Bias and variance of** F_n). *Let Assumptions* ***(A1)–(A5)*** *hold.*

(1) *If* $0 < a \leq \frac{\alpha}{3}$, *then*

$$Bias(F_n(x)) = O(m_n^{-2}). \tag{6.3.1}$$

If $\frac{\alpha}{3} < a < 1$

$$Bias(F_n(x)) = o\left(\sqrt{\gamma_n m_n^{-1}}\right). \tag{6.3.2}$$

(2) *If* $\frac{\alpha}{3} \leq a < 1$, *then*

$$Var(F_n(x)) = \frac{\sigma^2(x)}{2-\alpha\xi}\gamma_n + O(\gamma_n m_n^{-1}). \tag{6.3.3}$$

If $\frac{\alpha}{4} \leq a < \frac{\alpha}{3}$, *then*

$$Var(F_n(x)) = \sigma^2(x)\frac{1}{2-\alpha\xi}\gamma_n + o(\gamma_n). \tag{6.3.4}$$

If $0 < a < \frac{\alpha}{4}$, *then*

$$Var(F_n(x)) = o(m_n^{-4}). \tag{6.3.5}$$

The bias and the variance of the estimator F_n defined by the stochastic approximation algorithm, depend heavily on the choice of the stepsize (γ_n) and the degree (m_n). Let us first state the following proposition, which provides the MSE of the recursive estimator defined in (6.1.4).

Proposition 6.2 (**MSE of** F_n). *Let Assumptions* ***(A1)–(A5)*** *hold.*

(1) If $a \in]0, \alpha/4[$,

$$MSE(F_n(x)) = O(m_n^{-4}).$$

(2) If $a \in [\alpha/4, \alpha/3[$,

$$MSE(F_n(x)) = \frac{\sigma^2(x)}{2-\alpha\xi}\gamma_n + o(\gamma_n) + O(m_n^{-4}).$$

(3) If $a = \alpha/3$,

$$MSE(F_n(x)) = \frac{\sigma^2(x)}{2-\alpha\xi}\gamma_n + O(m_n^{-4}) + O(\gamma_n m_n^{-1}).$$

(4) If $a \in]\alpha/3, 1[$,

$$MSE(F_n(x)) = \frac{\sigma^2(x)}{2-\alpha\xi}\gamma_n + O(\gamma_n m_n^{-1}).$$

The following corollary is a by-product of the previous proposition, which affords the optimal orders (m_n) and (γ_n) of the estimator F_n minimizing the MSE.

Corollary 6.3.1. *Let Assumptions **(A1)**–**(A5)** hold. To minimise the MSE of F_n, we need to choose $(m_n) \in \mathcal{GS}(1/3)$ and $(\gamma_n) \in \mathcal{GS}(-1)$.*

The previous result implies that, to minimize $MSE(F_n(x))$, a must be equal to $\alpha/3$. In this case,

$$MSE(F_n(x)) = \frac{\sigma^2(x)\gamma_0^2}{2\gamma_0 - 1} n^{-1} + O(n^{-4/3}), \tag{6.3.6}$$

where $\gamma_n = \gamma_0 n^{-1}$. The following remarks show that, for a special choice of the stepsize (γ_n) and the degree (m_n), the proposed distribution estimator F_n can dominate the non recursive estimator $\mathcal{F}_{n,m}$ and the recursive estimators $F_{n,K}$ and $F_{n,B}$ in terms of MSE.

Remark 6.3.1 (**Comparison with estimators (6.1.1) and (6.1.2)**). *We note that the MSE of the recursive kernel estimator (6.1.1) (see Nadaraya (1964)) and the MSE of the recursive estimator using the Bernstein polynomial (6.1.2) (see Jmaei et al. (2017)) are established respectively by*

$$MSE(F_{n,K}(x)) = \left[\frac{\sigma^2(x)}{2-\alpha\xi}\gamma_n - \frac{f(x)\phi(K)}{2-(a+\alpha)\xi}\gamma_n h_n + \frac{\mu_2^2(K)f'^2(x)}{4(1-2a\xi)^2}h_n^4 + o(h_n^4)\right],$$

and

$$\begin{aligned} MSE(F_{n,B}(x)) = &\left[\frac{c^2(x)}{(1-a\xi)^2}m_n^{-2} + \frac{\sigma^2(x)}{2-\alpha\xi}\gamma_n - \frac{2V(x)}{4-(2\alpha+a)\xi}\gamma_n m_n^{-1/2}\right. \\ &\left. + o\left(m_n^{-2} + \gamma_n m_n^{-1/2}\right)\right], \end{aligned}$$

where $\mu_2(K) = \int_{\mathbb{R}} z^2 K(z)dz$, $\phi(K) = 2\int_{\mathbb{R}} zK(z)\mathcal{K}(z)dz$, $c(x) = \frac{x(1-x)f'(x)}{2}$ and $V(x) = f(x)\left[\frac{2x(1-x)}{\pi}\right]^2$. Following Slaoui (2014), we recall that to minimize the MSE of $F_{n,K}(x)$, we need to choose $(\gamma_n) \in \mathcal{GS}(-1)$ and $(h_n) \in \mathcal{GS}(-1/3)$. Departing from Corollary 3.1 in Jmaei et al. (2017), we note that to minimize the MSE of $F_{n,B}(x)$, we need to choose $(\gamma_n) \in \mathcal{GS}(-1)$ and $(m_n) \in \mathcal{GS}(2/3)$. In conclusion, if we consider that $h_n = m_n^{-1}$ as the "bandwidth" of the estimator (6.1.1), and regarding the performance of the proposed estimator, we point out that the three considered estimators (6.1.1), (6.1.2) and (6.1.4) are asymptotically equivalent in terms of MSE. Besides, with an optimal choice of γ_0, the proposed estimator can be a competitor of the estimators (6.1.2) and (6.1.1).

Remark 6.3.2 (**Comparison with the non recursive estimator**). *The non recursive estimator using Lagrange polynomial is defined by*

$$\mathcal{F}_{n,m}(x) = \sum_{i=1}^{m} \overline{F}_n(x_i)\mathcal{L}_i(x). \tag{6.3.7}$$

Helali and Slaoui (2020) showed that the MSE of (6.3.7) is provided by

$$MSE(\mathcal{F}_{n,m}(x)) = \pi^2 m^{-4} T_m^2(x) b^2(x) + n^{-1}\sigma^2(x) + O(n^{-1}m^{-1/2}) + o(m_n^{-4}).$$

Noteworthy that when n goes to infinity, $MSE(F_n(x)) \sim m_n^{-4} + n^{-1} + n^{-1}m_n^{-1}$ and $MSE(\mathcal{F}_{n,m}(x)) \sim m_n^{-4} + n^{-1} + n^{-1}m_n^{-1/2}$. It follows that, the proposed estimator asymptotically dominates the non recursive estimator in terms of MSE. On the other setting, the main merit of the proposed estimator, is that its update, when new sample points are available, requires less computational cost than the non recursive estimator.

Finally, the following proposition shows the asymptotic normality of the proposed recursive estimator (6.1.4).

Proposition 6.3 (**Asymptotic normality of F_n**). *Let Assumptions **(A1)–(A5)** hold. If $\lim_{n\to+\infty} \gamma_n^{-1} m_n^{-4} = 0$, then*

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

Here $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution and $\mathcal{N}$ corresponds to the Gaussian distribution.

Note that, under an appropriate choice of bandwidth, a result similar to Theorem 1 was shown by Slaoui (2014) for recursive kernel estimators, and by Jmaei et al. (2017) for results from recursive Bernstein estimator of distribution functions. The following corollary is a consequence of the previous proposition which provides an asymptotic confidence interval of the distribution F, for a confidence level $0 < \alpha < 1$.

Corollary 6.3.2. *If $\lim_{n\to+\infty} \gamma_n^{-1} m_n^{-4} = 0$, then the $100(1-\alpha)\%$ asymptotic confidence interval of $F(x)$ is given by*

$$\left(F_n(x) \pm \frac{z_{1-\frac{\alpha}{2}}\sigma(x)\gamma_n^{1/2}}{\sqrt{2-\alpha\xi}}\right),$$

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

In the next section, we shall explore the performance of the proposed recursive estimator in estimating different distributions by comparing it to the performances of the recursive Bernstein estimator (6.1.2), as well as that of the recursive Gaussian kernel estimator (6.1.1) and of the non recursive estimator using Lagrange polynomials (6.3.7).

6.4 Numerical studies

6.4.1 Simulations

In this section, we demonstrate the performance of the proposed estimator (6.1.4) in estimating different distributions by comparing it to the performances of the recursive (Gaussian) kernel estimator (6.1.1), of the recursive estimator using Bernstein polynomial (6.1.2) as well as that of the non recursive estimator using Lagrange polynomial (6.3.7).

We can apply the Bernstein estimator (6.1.2) and the Lagrange estimators (6.3.7) and (6.1.4) when the sample is concentrated on the intervals $[0,1]$ and $[-1,1]$, respectively. Then, in order to enact the comparison between the Bernstein estimator (6.1.2), the Lagrange estimators (6.1.4)-(6.3.7) and the kernel approach (6.1.1), which are applicable in general, we elaborate below suggested transformations in different cases:

(1) Suppose that X is concentrated on a finite support $[a,b]$, then we work with the sample values $Y_1,\ldots,Y_n$ where $Y_i=(X_i-a)/(b-a)$.

(2) For the distributions functions concentrated on $\mathbb{R}$, we can use the transformed sample $Y_i=1/2+\pi^{-1}\arctan(X_i)$ which transforms the range to the interval $[0,1]$.

(3) For the support $\mathbb{R}_+$, we can use the transformed sample $Y_i=X_i/(1+X_i)$ which transforms the range to the interval $[0,1]$.

In our simulation study, three sample sizes are considered, $n=50,100,200$ according to the following distribution functions F:

1− The beta distribution $\mathcal{B}(3,2)$,

2− The gamma distribution $\mathcal{G}(1,6)$,

3− The mixture beta distribution $0.5\mathcal{B}(2.5,6)+0.5\mathcal{B}(9,1)$.

For each distribution function F and sample size n, we compute the Integrated Squared Error (ISE) of the estimator over $N=500$ trials,

$$ISE[\widehat{F}]=\int_0^1\Big(\widehat{F}(x)-F(x)\Big)^2\,dx, \tag{6.4.1}$$

where $\widehat{F}$ is an estimator of the true distribution F. To select the smoothing parameters m and h, we consider the Monte Carlo procedure for each point $x\in[0,1]$. We determine

the parameters m (for $1 \leq m \leq 300$), ν and h (for $h = i/1000$ with $1 \leq i \leq 300$), which minimizes ISE, that is approximated by

$$\frac{1}{N} \sum_{i=1}^{N} ISE_i(\widehat{F}),$$

where $ISE_i(\widehat{F})$ is the value of ISE computed from the ith sample of size n and obtained from (6.4.1).

Table 6.1: ISE for $N = 500$ trials of *Recursive 1*, *Recursive 2* and *Recursive 3*

Density	n	*Recursive 1*	*Recursive 2*	*Recursive 3*
		$c = 0.05$		
	50	**0.003660**	0.004273	0.046308
$\mathcal{B}(3,2)$	100	**0.002188**	0.002306	0.018131
	200	**0.000706**	0.001306	0.011315
	50	**0.006846**	0.009218	0.131480
$\mathcal{G}(1,6)$	100	**0.002186**	0.002285	0.072287
	200	**0.000272**	0.000310	0.019873
	50	0.001168	**0.000995**	0.023376
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	100	0.000531	**0.000506**	0.010810
	200	0.000354	**0.000345**	0.002522
		$c = 0.1$		
	50	**0.003833**	0.004285	0.041341
$\mathcal{B}(3,2)$	100	**0.002193**	0.002458	0.018925
	200	**0.000681**	0.001232	0.010338
	50	**0.006211**	0.008690	0.047800
$\mathcal{G}(1,6)$	100	**0.001622**	0.002355	0.034586
	200	**0.000160**	0.000205	0.015976
	50	0.001110	**0.000953**	0.030482
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	100	**0.000546**	0.000662	0.010861
	200	**0.000458**	0.000513	0.005940
		$c = 0.2$		
	50	**0.004148**	0.004549	0.072466
$\mathcal{B}(3,2)$	100	**0.002306**	0.002538	0.022216
	200	**0.001368**	0.001633	0.012922
	50	**0.005435**	0.007110	0.092982
$\mathcal{G}(1,6)$	100	**0.001820**	0.001773	0.050594
	200	$\mathbf{8.042e^{-5}}$	0.000155	0.018933
	50	**0.001081**	0.001122	0.027487
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	100	**0.000636**	0.000657	0.011531
	200	0.000435	**0.000323**	0.006048

Here, *Recursive 1* correspond to the proposed estimator (6.1.4), *Recursive 2* corre-

spond to the recursive kernel estimator (6.1.1) and *Recursive 3* correspond to the recursive Bernstein estimator (6.1.2), for $n = 50$, $n = 100$ and $n = 200$, for $(\gamma_n) = ([2/3 + c]n^{-1})$, for $c = 0.05$, $c = 0.1$ and $c = 0.2$. The bold values indicate the smallest values of ISE.

Table 6.2: ISE for $N = 500$ trials of $\mathcal{F}_{n,m}$ and F_n, for $n = 50, 100, 200$. [ISE for $N = 500$ trials of the non recursive estimator (6.3.7) and the proposed estimator (6.1.4), for $n = 50$, $n = 100$ and $n = 200$, for the case $(\gamma_n) = ([2/3 + \epsilon]n^{-1})$ with $\epsilon = 0.05, 0.1, 0.2$ and $(\gamma_n) = (n^{-1})$. The bold values indicate the smallest values of ISE.]

			Recursive estimator		
			$(\gamma_n) = ([2/3 + \epsilon]n^{-1})$		
	n	Non recursive	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.2$
$\mathcal{B}(3, 2)$	50	0.003798	**0.003660**	0.003714	0.004148
	100	0.002471	**0.002188**	0.002193	0.002306
	200	0.001459	0.000706	**0.000681**	0.001368
$\mathcal{G}(1, 6)$	50	0.004998	0.006846	**0.006211**	0.005435
	100	0.001477	0.002186	**0.001622**	0.001820
	200	0.001264	0.000272	0.000160	$\mathbf{8.042e^{-5}}$
$0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$	50	0.001203	**0.001110**	0.001119	0.001081
	100	0.000718	**0.000531**	0.000546	0.000636
	200	0.000412	**0.000354**	0.000458	0.000435

Departing from Tables 6.1 and 6.2, we infer that:

(i) In all the cases, the ISE of the proposed distribution estimator (6.1.4) with an appropriate choice of the stepsize (γ_n) is smaller than that of the other recursive estimators (6.1.1)-(6.1.2) and the non recursive distribution estimator (6.3.7).

(ii) The ISE decreases as the sample size increases.

(iii) Relying upon Table (6.1), we deduce that $1-$ for small and moderate samples size ($n = 50$ and $n = 100$), the proposed recursive estimator displays better results compared to the two other recursive methods in terms of ISE, except for the mixture beta distribution in the cases $c = 0.05$, $c = 0.1$ for $n = 50$. $2-$ for large samples size $n = 200$, the proposed recursive estimators have a smaller ISE compared to the other recursive estimators, except for the mixture beta distribution in the cases $c = 0.05$ and $c = 0.2$.

(iv) Referring to Table (6.2), we notice that the proposed recursive estimator presents better results compared to the non-recursive distribution estimator in all the cases using an appropriate choice of the stepsize (γ_n).

6.4.2 Real dataset

The basic target of this subsection is to compare the performance of the proposed estimator defined in (6.1.4) to that of the non recursive estimator defined in (6.3.7). We consider two examples that highlight the features of the proposed estimator F_n:

1. We analyze the suicide data portrayed in Table 2.1 of Silverman (1986). This data set consists of duration (in days) of psychiatric treatment for 86 patients used as experiments in a study of suicide risks. The maximum treatment duration is 800 days (the data are such that $\min(x_i) = 1$ and $\max(x_i) = 737$). Finally, we used the Monte Carlo method to obtain $m = 86$ for the proposed estimator where $c = 0.05$, and $m = 86$ for the non recursive estimator.

2. We use `Salvister` data which appear in the `R` package `kerdiest` (Quintela-del-Río and Estévez-Pérez (2012)). This data involve 85 observations of the annual peak instantaneous flow levels of the Salt River near Roosevelt, AZ, USA, for the period $1924 - 2009$, reported from the National Water Information System. Finally, we used the Monte Carlo method to obtain $m = 80$ for the proposed estimator where $c = 0.3$, and $m = 85$ for the non recursive estimator.

In the real examples, the two estimators (6.3.7) and (6.1.4) are compared with the empirical distribution $\overline{F}_n$. Then, for any considered estimator $\widehat{F}$ of the distribution function F, we propose to compute the ISE defined as:

$$ISE(\widehat{F}) \quad = \quad \int_0^1 \left(\widehat{F}(x) - \widehat{F}_n(x)\right)^2 dx.$$

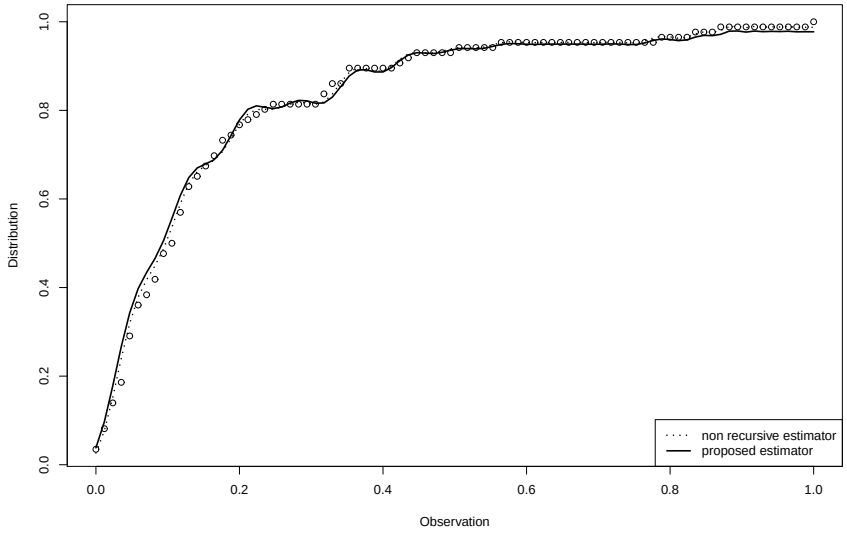

Figure 6.1: Qualitative comparison between $\mathcal{F}_{n,m}$ and F_n for `Suicide` data. [Qualitative comparison between the non recursive estimator $\mathcal{F}_{n,m}$ and the proposed recursive estimator F_n for `Suicide` data.]

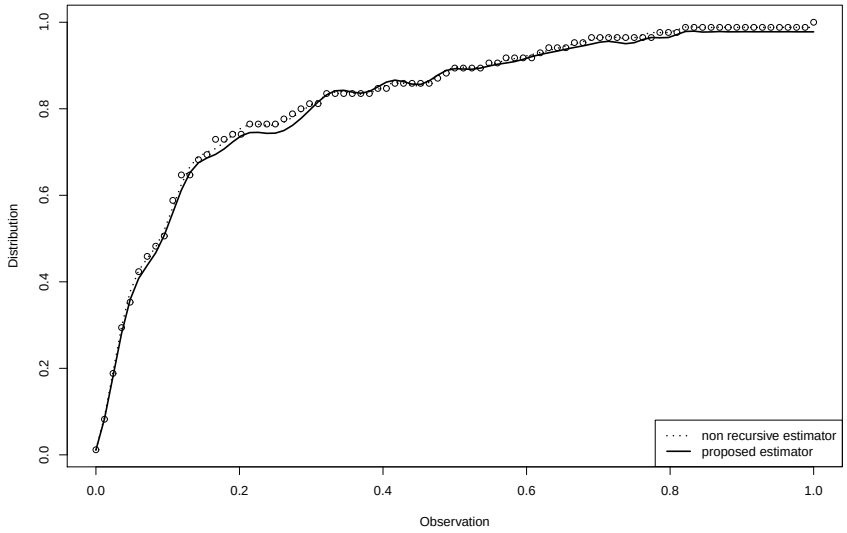

Figure 6.2: Qualitative comparison between $\mathcal{F}_{n,m}$ and F_n for `Salvister` data. [Qualitative comparison between the non recursive estimator $\mathcal{F}_{n,m}$ and the proposed recursive estimator F_n for `Salvister` data.]

Table 6.3: ISE of $\mathcal{F}_{n,m}$ and F_n, for `Saltriver` data and `Suicide` data.
[ISE of the non recursive estimator $\tilde{F}_{n,m}$ and the proposed estimator F_n, for `Saltriver` data and `Suicide` data. The bold values indicate the smallest values of ISE.]

Data set	Proposed estimator	Non recursive estimator
Saltriver	$\mathbf{4.5519e^{-8}}$	$6.1417e^{-8}$
Suicide	$\mathbf{4.1893e^{-6}}$	$4.4607e^{-6}$

Visually, from Figures 6.1-6.2, we conclude that the non-recursive estimator (6.3.7) and the proposed recursive estimator (6.1.4) seem to be close to the true empirical distribution functions. Resting on Table 6.3, we infer that that the ISE of the proposed estimator is smaller than that of the non recursive estimator with optimal parameters (γ_n) and (m_n), thus demonstrating the effectiveness of the considered estimator.

6.5 Conclusion

This work elaborates a recursive distribution estimator using Lagrange polynomial with Tchebytchev-Gauss points. The proposed estimator asymptotically follows normal distribution. The ISE of the proposed estimator (6.1.4), with an appropriate choice of the stepsize (γ_n) and the order (m_n), is smaller than that of the non recursive estimator (6.3.7) as well as the recursive estimators (6.1.1)-(6.1.2). In addition, a major advantage of the proposed estimator, lies in the fact that its up date, when new sample points are available, requires less computational cost than the non recursive estimator. To sum up, using the proposed recursive estimator F_n, we can get better results than those obtained by estimators (6.3.7), (6.1.1) and (6.1.2). A future research direction would be to extend our findings to the recursive regression estimation using Lagrange polynomial. We plan also to work on the choice of the order (m_n) through a plug-in method (see Slaoui (2016)).

6.6 Proofs

Let us first state the following technical lemma.

Lemma 6.1 (Mokkadem et al (2009)). *Let $(w_n) \in \mathcal{GS}(w^\star)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, and let $l > 0$ such that $l - w^\star\xi > 0$ where $\xi = \lim_{n\to\infty}(n\gamma_n)^{-1}$. We have*

$$\lim_{n\to\infty} w_n \Pi_n^l \sum_{k=1}^{n} \Pi_k^{-l} \frac{\gamma_k}{w_k} = \frac{1}{l - w^\star\xi}.$$

Furthermore, for all positive sequence (α_n) such that $\lim_{n\to\infty}\alpha_n = 0$, and all $\rho \in \mathbb{R}$;

$$\lim_{n\to\infty} w_n \Pi_n^l \left[\sum_{k=1}^{n} \Pi_k^{-l} \frac{\gamma_k}{w_k}\alpha_k + \rho\right] = 0.$$

Lemma 6.1 is widely applied throughout the proofs and it is proved in Mokkadem et al (2009). Its application requires assumption **(A5)** on the limit of $n\gamma_n$ as n goes to infinity.

6.6.1 Proof of Proposition 6.1

Based on equation (6.1.3), for $x \in [-1, 1]$, we have

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

It follows that

$$\mathbb{E}(F_n(x)) - F(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k [\mathbb{E}(Z_k(x)) - F(x)] + \Pi_n (F_0(x) - F(x)). \tag{6.6.1}$$

In addition, we have,

$$\mathbb{E}(Z_n(x)) \;=\; \mathbb{E}\left(\sum_{k=1}^{m} \mathbb{1}_{\{X_n \leq x_k\}} \mathcal{L}_k(x) \right) = \sum_{k=1}^{m} F(x_k)\, \mathcal{L}_k(x) = A_m(x).$$

Using Proposition 3.1 in Helali and Slaoui (2020) (and its proof), we obtain

$$\mathbb{E}(Z_n(x)) - F(x) = \pi b(x) T_m(x) m_n^{-2} + o(m_n^{-2}).$$

Substituting this result into (6.6.1), leads to

$$\mathbb{E}(F_n(x)) - F(x) \;=\; \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k (\pi b(x) T_{m_k}(x) m_k^{-2} [1 + o(1)] + \Pi_n (F_0(x) - F(x))$$

For the case $0 < a \leq \frac{\alpha}{3}$, we have $\xi < 1/2a$. The application of lemma 6.1 entails

$$\mathbb{E}(F_n(x)) - F(x) \;=\; O(m_n^{-2}).$$

which follows (6.3.1). In the case $\frac{\alpha}{3} < a < 1$, we have $\xi < 2/(\alpha + a)$ and $m_n^{-2} = o\left(\sqrt{\gamma_n m_n^{-1}}\right)$. The application of lemma 6.1 provides

$$\mathbb{E}(F_n(x)) - F(x) \;=\; o\left(\sqrt{\gamma_n m_n^{-1}}\right), \text{ which gives (6.3.2).}$$

Next, we have for $x \in [-1, 1]$ and $n \geq 1$,

$$Var(F_n(x)) \;=\; \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 Var(Z_k(x)), \tag{6.6.2}$$

and

$$\begin{aligned} Var(\mathcal{F}_{n,m}(x)) \;&=\; \frac{1}{n^2} Var\left(\sum_{i=1}^{n} \sum_{k=1}^{m} \mathbb{1}_{\{X_i \leq x_k\}} \mathcal{L}_k(x) \right) \\ &=\; \frac{1}{n} Var(Z_n(x)). \end{aligned}$$

Grounded on Proposition 3.1 in Helali and Slaoui (2020), we obtain

$$Var(Z_n(x)) = \sigma^2(x) + O(m_n^{-1}). \tag{6.6.3}$$

Substituting this result into (6.6.2), leads to

$$Var(F_n(x)) \;=\; \sigma^2(x)\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 + \Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 O(m_k^{-1}).$$

In the case $\frac{\alpha}{3} \leq a < 1$, we have $\lim_{n\to\infty} n\gamma_n > (a+\alpha)/2$. The application of Lemma 6.1 gives

$$Var(F_n(x)) \;=\; \sigma^2(x)\frac{1}{2-\alpha\xi}\gamma_n + O(\gamma_n m_n^{-1}),$$

which proves (6.3.3). In the case $\frac{\alpha}{4} \leq a \leq \frac{\alpha}{3}$, we have $\lim_{n\to\infty} n\gamma_n > 2a > \frac{\alpha}{2}$ and $\gamma_n m_n^{-1} = o(m_n^{-4})$. The application of Lemma 6.1 gives

$$\begin{aligned} Var(F_n(x)) &= \sigma^2(x)\frac{1}{2-\alpha\xi}\gamma_n + \Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 o(m_k^{-4}) \\ &= \sigma^2(x)\frac{1}{2-\alpha\xi}\gamma_n + o(\gamma_n). \end{aligned}$$

which gives (6.3.4). In the case $0 < a < \frac{\alpha}{4}$, we have $\lim_{n\to\infty} n\gamma_n > 2a$ and $\gamma_n = o(m_n^{-4})$. The application of Lemma 6.1 gives (6.3.5).

6.6.2 Proof of Corollary 6.3.1

Let

$$G(\xi) = \frac{\sigma^2(x)}{2-\alpha\xi}.$$

It follows from Proposition 6.2 that

$$MSE(F_n(x)) = \begin{cases} O(m_n^{-4}) & \text{if } a \in]0, \alpha/4[\\ G(\xi)\gamma_n + o(\gamma_n) + O(m_n^{-4}) & \text{if } a \in [\alpha/4, \alpha/3[\\ G(\xi)\gamma_n + O(\gamma_n m_n^{-1}) + O(m_n^{-4}) & \text{if } a = \alpha/3 \\ G(\xi)\gamma_n + O(\gamma_n m_n^{-1}) & \text{if } a \in]\alpha/3, 1[. \end{cases}$$

By (**A3**)-(**A4**), we have $MSE(F_n(x)) \in \mathcal{GS}(-4a)$ if $a \in]0, \alpha/4[$ and $MSE(F_n(x)) \in \mathcal{GS}(-\alpha)$ if $a \in [\alpha/4, 1[$. Then to minimize the MSE of F_n, the parameter a must be chosen in the interval $[\alpha/4, 1[$. Therefore, we shall focus the comparison on MSE, when $a \in [\alpha/4, 1[$. We have $G(\xi)\gamma_n + o(\gamma_n) \in \mathcal{GS}(-\alpha)$. Moreover, if $a \in [\alpha/4, \alpha/3[$, $O(m_n^{-4}) \in \mathcal{GS}(-4a)$. If $a = \alpha/3$, $G(\xi)\gamma_n + O(\gamma_n m_n^{-1}) + O(m_n^{-4}) \in \mathcal{GS}(-4/3\alpha)$ and if $a \in]\alpha/3, 1[$, $O(\gamma_n m_n^{-1}) \in \mathcal{GS}(-(a+\alpha))$. It follows that, for a given $\alpha \in]1/2, 1]$, to minimize the $MSE(F_n(x))$, the parameter a must be chosen equal to $\alpha/3$ and $\alpha = 1$. In other words, to minimize the MSE of F_n, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and the degree (m_n) in $\mathcal{GS}(1/3)$.

6.6.3 Proof of Proposition 6.3

To prove Proposition 6.3, we shall use the following lemma, which will be proved later.

Lemma 6.2. *If $a \geq \frac{\alpha}{4}$, then*

$$\gamma_n^{-1/2}[F_n(x) - \mathbb{E}(F_n(x))] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2(x)}{2-\alpha\xi}\right).$$

We have $\gamma_n^{-1/2}(F_n(x) - F(x)) = \gamma_n^{-1/2}[F_n(x) - \mathbb{E}(F_n(x))] + \gamma_n^{-1/2}[\mathbb{E}(F_n(x)) - F(x)]$. If $\lim_{n\to\infty} \gamma_n^{-1} m_n^{-4} = 0$, then $a \geq \alpha/4$. In the case $\alpha/4 \leq a \leq \alpha/3$, Proposition 6.3 follows from lemma 6.3. In the case $\alpha/3 < a < 1$, Proposition 6.3 follows from Equation (6.3.2) and lemma 6.3.

Proof: [Proof of lemma 6.2] Let $a \geq \alpha/4$. Relying on equation (6.1.3), we have

$$\begin{aligned} F_n(x) - \mathbb{E}(F_n(x)) &= (1-\gamma_n)[F_{n-1}(x) - \mathbb{E}(F_{n-1}(x))] + \gamma_n[Z_n(x) - \mathbb{E}(Z_n(x))] \\ &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k[Z_k(x) - \mathbb{E}(Z_k(x))]. \end{aligned}$$

Set $R_k(x) = \Pi_k^{-1}\gamma_k[Z_k(x) - \mathbb{E}(Z_k(x))]$ and $\tau_n^2 = \sum_{k=1}^{n} Var(R_k(x))$. Then,

$$F_n(x) - \mathbb{E}(F_n(x)) = \Pi_n \sum_{k=1}^{n} R_k(x) \text{ and } \tau_n^2 = \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var(Z_k(x)).$$

Resting upon equation (6.6.3), we get $\tau_n^2 = \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2(\sigma^2(x) + o(1))$. As in the case $a \geq \alpha/4$, an application of Lemma 6.1 ensures that

$$\begin{aligned} \tau_n^2 &= \Pi_n^{-2}\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2[\sigma^2(x) + o(1)] \\ &= \Pi_n^{-2}\gamma_n\left[\frac{1}{2-\alpha\xi}\sigma^2(x) + o(1)\right]. \end{aligned}$$

On the other side, we have for all $p > 0$, $\mathbb{E}(\mid Z_k(x) \mid^{2+p}) = O(1)$. It follows that

$$\begin{aligned} \sum_{k=1}^{n} \mathbb{E}\left[\mid R_k(x) \mid^{p+2}\right] &= O\left(\sum_{k=1}^{n} \Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left(\mid Z_k(x) \mid^{2+p}\right)\right) \\ &= O\left(\sum_{k=1}^{n} \Pi_k^{-2-p}\gamma_k^{2+p}\right). \end{aligned}$$

Since $\lim_{n\to\infty} n\gamma_n > \alpha/2$, then there exists $p > 0$ such that $\lim_{n\to+\infty} n\gamma_n > \frac{1+p}{2+p}\alpha$. An application of Lemma 6.1 yields

$$\sum_{k=1}^{n} \mathbb{E}\left[\mid R_k(x) \mid^{p+2}\right] = O\left(\gamma_n^{p+1}\Pi_n^{-(2+p)}\right).$$

We, thus, obtain

$$\begin{aligned}\frac{1}{\tau_n^{2+p}}\sum_{k=1}^{n}\mathbb{E}\left[\mid R_k(x) \mid^{2+p}\right] &= \frac{1}{\frac{\gamma_n^{1+\frac{p}{2}}}{\Pi_n^{2+p}}\left(\frac{1}{2-\alpha\xi}\sigma^2(x)+o(1)\right)^{1+\frac{p}{2}}}O\left(\gamma_n^{p+1}\Pi_n^{-(2+p)}\right)\\ &= \frac{1}{\left(\frac{1}{2-\alpha\xi}\sigma^2(x)+o(1)\right)^{1+\frac{p}{2}}}O(\gamma_n^{\frac{p}{2}}) = O(\gamma_n^{\frac{p}{2}}) = o(1).\end{aligned}$$

Hence, Lemma 6.2 follows from the application of Lyapounov's theorem.

Chapter 7

Conclusion and perspectives

The manuscript proposed new nonparametric estimation techniques to deal with the problem of edge effect. The aim of first contribution was to build a new semi parametric structure to estimate a density function. Indeed, we proposed a shrinkage estimator of a density function based on the Bernstein polynomial and using a finite Mixture Gaussian model. The asymptotic properties of this estimator were established. Afterwards, we demonstrated the effectiveness of the proposed method using some simulated and real data.

In the second contribution, we investigated a recursive nonparametric regression estimator to overcome the edge estimation problem based on Bernstein polynomials and stochastic algorithm with two-time-scale. The proposed estimator asymptotically follows normal distribution. Moreover, our proposed estimators attained the asymptotic convergence rate $O(n^{-4/5})$ within the interval $(0, 1)$ and $O(n^{-2/3})$ near the edges $\{0, 1\}$.

In third contribution, we introduced an estimator of the distribution function using Lagrange polynomials and Tchebytchev-Gauss points. We argued that the proposed estimator asymptotically dominates the Bernstein estimator in terms of bias. Through a simulation study and a simple data set examples, we have demonstrated the proposed estimator is quite promising and interesting as it behaves well when compared with both the Bernstein estimator and the Gaussian kernel estimator.

In our last work, we elaborated a recursive distribution estimator using Lagrange polynomial with Tchebytchev-Gauss points. Using the proposed recursive estimator, we can get better results than those obtained by the previous non recursive estimator, the recursive kernel approach and the recursive Bernstein estimator, with an appropriate choice of the stepsize (γ_n) and the order (m_n). In addition, a major advantage of the proposed estimator, lies in the fact that its up date, when new sample points are available,

requires less computational cost than the non recursive estimator.

Our current work revolves around these axes:

⋆ The first axis consists on studying the following estimation of density function base on the the Lagrange polynomial and Tchebytchev-Gauss points, defined as follows:

$$\tilde{f}_{n,m}(x) = \sum_{i=1}^{m} \sum_{j=1, j\neq i}^{m} \frac{F_n(x_i)}{x_i - x_j} \prod_{k=1, k\neq(i,j)}^{m} \frac{x - x_k}{x_i - x_k},$$

where F_n is the empirical distribution function and for all $i = 1 \dots m$, x_i are the Tchebytchev-Gauss points.

⋆ The second axis consists on studying the following estimation of regression function base on the the Lagrange polynomial and Tchebytchev-Gauss points, defined as follows:

$$\tilde{r}_{n,m}(x) = \frac{\sum_{i=1}^{m} Y_i \sum_{k=1}^{m} 1_{X_i \leq x_k} \left[\sum_{\substack{j=0 \\ j\neq k}}^{m} \frac{1}{x_k - x_j} \prod_{\substack{h=0 \\ h\neq(j,k)}}^{m} \frac{x - x_h}{x_k - x_h} \right]}{\sum_{i=1}^{m} \sum_{k=1}^{m} 1_{X_i \leq x_k} \left[\sum_{\substack{j=0 \\ j\neq k}}^{m} \frac{1}{x_k - x_j} \prod_{\substack{h=0 \\ h\neq(j,k)}}^{m} \frac{x - x_h}{x_k - x_h} \right]}.$$

⋆ An other direction in progress is to study a recursive conditional distribution estimator using Bernstein polynomials and kernel method defined as follows:

$$F_{x,n}(y) = \begin{cases} \frac{a_n(x,y)}{f_n(x)}, & \text{if } f_n(x) \neq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where for any $x \in \mathbb{R}$ and $y \in [0,1]$,

$$a_n(x,y) = (1-\beta_n) a_{n-1}(x,y) + \beta_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right) \sum_{k=0}^{m_n} \mathbb{1}_{\{Y_n \leq k/m_n\}} b_{k,m_n}(y),$$

and for any $x \in \mathbb{R}$,

$$f_n(x) = (1-\gamma_n) f_{n-1}(x) + \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right),$$

with (γ_n) and (β_n) sequences of positive real numbers there go to zero. We let $Q_n = \prod_{i=1}^{n}(1-\beta_i)$, $\Pi_n = \prod_{i=1}^{n}(1-\gamma_i)$ and $(a_0(x,y), f_0(x)) = (0,0)$.

At this stage of analysis, we would assert that our research work is a step that may be extended, built upon and taken further as it lays the ground and paves the way for constructive and fruitful applications. In future research, we may study the following points:

- Applying Lagrange polynomials to estimate the pairwise comparison function.
- Explore the possibility to consider Lagrange polynomial and Tchebytchev-Gauss points in the case of censored data, it is also possible to explore the case of functional data.
- Extend our proposal in Bayesian non parametric estimation framework.

Bibliography

Altman, N. and Léger, C., (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inference*, **46**, 195–214.

Austin, A. P., (2016). Some New Results On, and Applications Of, Interpolation in Numerical Computation. *Doctoral dissertation, University of Oxford.*

Babu, G. J., Canty, A. J., and Chaubey, Y.P., (2002). Application of Bernstein Polynomials for smooth estimation of a distribution and density function. *J. Statist. Plann. Inference*, **105**, 377–392.

Babu, G. J. and Chaubey, Y. P., (2006). Smooth estimation of a distribution function and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statist. Probab. Lett.*, **76**, 959—969.

Bhatnagar, S., Fu, M. C., Marcus S. I. and Bathnagar, S., (2001). Two timescale algorithms for simulation optimization of hidden Markov models. *IEEE Trans.*, **3**, 245-258.

Benveniste, A. M. Métivier, and P. Priouret., (1912). Adaptive Algorithms and Stochastic Approximations. Applications of Mathematics. *Springer Verlag.*

Bernstein, S., (1912). Démonstration du théorème de Weierstrass fondeée sur le calcul des probabiliteés. *Communications of the Kharkov Mathematical Society*, **13**, 1–2.

Bojanic R. and Seneta E., (1973). A Unified Theory of Regularly Varying Sequences. *Math.Z.*, **134**, 91-106.

Borkar, V. S., (1997). Stochastic approximation with two time scales. *Systems Control Lett.*, **29**, 291–294.

Bouezmarni, T. and J. Rolin, (2007). Bernstein estimator for unbounded density function. *Nonparametr. Stat.*, **19**, 145—161.

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

Brutman, L., (1978). On the Lebesgue function for polynomial interpolation. *SIAM J. Numer. Anal.*, **15**, 694–704.

Cheney, E. W. (1966). Introduction to approximation theory. *Mc Graw-Hill, New York.*

Chen, S. X., (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, **83**, 329—341.

Chen, S. X., (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.*, **31**, 131-–145.

Chen, S.X., (2000). Probability density function estimation using gamma kernels. *Ann. Inst. Statist. Math.* **52**, 471—480.

Chung, K. L., (1949). An estimate concerning the kolmogroff limit distribution. *Trans. Amer. Math. Soc.*, **67**, 36–50.

Delaigle, A., and Gijbels, I., (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal*, **45**, 249–267.

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.

Duflo, M., (1997). Random Iterative Models. *Collect. Appl. math., Spring.*

Feldheim, E., (1939). Théorie de la convergence des procédés d'interpolation et de quadrature mécanique.

Falk, M., (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. **37**, 73–83.

Eubank, R., (1988). *Spline Smoothing and Nonparametric Regression.* Dekker, New York.

Galambos, J., and Seneta, E., (1973). Regularly Varying Sequences, *Proc. Amer. Math. Soc.* **41**, 110–116.

Ghosal, S., (2000). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29**, 1264—1280.

Gruber, M., (1998). Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators. *Routledge.*

Härdle, W., (1990). *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Härdle, W., (1991). Smoothing techniques with implementation in S. *S. Springer Science and Business Media.*

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A., (2004). *Nonparametric and Semiparametric Models.* Springer Series in Statistics. Springer.

Hadrich, A., Zribi, M., and Masmoudi, A., (2014). A proposed nonparametric mixture density estimation using B-splines functions. *International Journal of Enhanced Research in Science, Technology and Engineering*, **5**, 2319–7463.

Hassen, H. B., Masmoudi, K., and Masmoudi, A., (2019). Model selection in biological networks using a graphical EM algorithm. *Neurocomputing*, **349**, 271–280.

Hastie, T. and Tishirani, R., (1990) *Generalized Additive Models.*, Chapman and Hall, New York.

Helali, S., and Slaoui, Y., (2020). Estimation of a distribution function using Lagrange polynomials with Tchebychev–Gauss points. *Stat. Interface*, **13**, 399–410.

Hjort N. and Claeskens G., (2003). Frequentist Model Average Estimators. *J. Amer. Statist. Assoc.*, **98**, 879–899.

Igarashi, G. and Kakizawa, Y., (2014). On improving convergence rate of Bernstein polynomial density estimator. *J. Nonparametr. Stat.*, **26**, 61–84.

Jackson, D., (1911). Über die Genauigkeit der Annäherung stetiger Funktionen durch ganze rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung. Dieterich.

Jackson, D., (1912). On approximation by trigonometric sums and polynomials. *Trans. Amer. Math. Soc.*, **13**, 491–515.

James, W., and Stein, C., (1992). Estimation with quadratic loss. *Breakthroughs in statistics (pp. 443-460). Springer, New York, NY.*

Jmaei, A., Slaoui, Y., and Dellagi, W., (2017). Recursive distribution estimators defined by stochastic approximation method using Bernstein polynomials. *J. Nonparametr. Stat.*, **29**, 792–805.

Jones, M. C., (1993). Simple boundary correction for density estimation kernel. *Stat. Comput.*, **3**, 135—146.

Joyner, W. B., and Boore, D. M., (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake.. *Bulletin of the seismological Society of America*, **71**, 2011-2038.

Kakizawa, Y., (2004). Bernstein polynomial probability density estimation. *J. Nonparametr. Stat.*, **16** , 709–729.

Kakizawa, Y., (2011). A note on generalized bernstein polynomial density estimators. *Stat. Methodol.*, **8**, 136—153.

Khardani, S. and Slaoui, Y., (2019). Recursive kernel density estimation and optimal bandwidth selection under alpha-mixing data. *J. Stat. Theory Pract.*, doi: 10.1007/s42519-018-0031-6.

Kiefer J. and Wolfowitz J., (1952). Stochastic Estimation of the Maximum of a Regression Function. *Ann. Math. Statist.*, **23**, 462–466.

Konda, V. R. and Borkar, V. S., (1999). Actor-Critic–Type Learning Algorithms for Markov Decision Processes. *SIAM J. Control Optim.*, **38**, 94-123.

Konda, V. R. and Tsitsiklis, J.N., (2004). Convergence rate of linear two-time-scale stochastic approximation. *Ann. of Appl. Prob ab.*, **14**, 796-819.

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.*

Ljung L., (1978). Strong Convergence of a Stochastic Approximation Algorithm. *Ann. Statist.*, **6**, 680–696.

Leblanc, A., (2009). Chung–Smirnov property for Bernstein estimators of distribution functions. *J. Nonparametr. Stat.*, **21**, 133–142.

Leblanc, A., (2010). A bias-reduced approach to density estimation using Bernstein polynomials. *Nonparametr. Stat.*, **22**, 459–475.

Leblanc, A., (2012a). On estimationg distribution function using Bernstein polynomials. *Ann. Inst. Statist. Math.*, **64**, 919–943.

Leblanc, A., (2012b). On the boundary properties of bernstein polynomial estimators of density and distribution functions. *J. Statist. Plann. Inference*, **142**, 2762—2778.

Lejeune, M. and Sarda, P., (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.*, **14**, 457—471.

Lehmann, E. L., and Casella, G., (2006). Theory of point estimation. *Springer Science and Business Media.*

Leroux B., (1992). Consistent estimation of a mixing distribution. *Ann. Stat.*, **20**, 1350–1360.

Li, J. Q., and Barron, A. R., (2000). Mixture density estimation. *Adv. Neural Inf. Process. Syst.*, **pp**, 279–285.

Lorentz, G. G., (1986). Bernstein polynomials (2nd ed.). *New York : Chelsea Publishing.*

McLachlan, G., and Peel, D., (2004). Finite mixture models. *John Wiley and Sons.*

Mokkadem, A. and Pelletier, M., (2006a). Convergence rate and averaging of nonlinear twotime-scale stochastic approximation algorithms. *Ann. Appl. Probab.*, **16**, 1671–1702

Mokkadem, A. Pelletier, M. and Slaoui, Y. , (2009). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference*, **139**, 2459—2478.

Müller, H.-G., (1988). *Nonparametric Analysis of Longitudinal Data*, Springer-Verlag, Berlin.

Müller, H. G., (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78**, 521–530.

Müller, H.-G., (1993). On the boundary kernel method for nonparametric curve estimation near endpoints. *Scand. J. Statist.*, **20**, 313–328.

Muller, H. G. and Wang, J. L., (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 61-76

Oman, S. D., (1982a). Contracting towards subspaces when estimating the mean of a multivariate normal distribution. *J. Multivariate Anal.*, **12**, 270–290.

Oman, S. D., (1982b). Shrinking towards subspaces in multiple linear regression. *Technometrics*, **24**, 307–311.

Nadaraya, E. A., (1964). Some new estimates for distribution functions. *Theory Probab. Appl.*, **9**, 497–500.

Nadaraya, E. A., (1964). On estimating regression. *Theory Probab. Appl.*, **10**, 186–190.

Peter D, H., (1985). Kernel estimation of a distribution function. *Comm. Statist. Theory Methods*, **14**, 605–620.

Parzen, E., (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Statist.*, **33**, 1065–1076.

Pearson, K., (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, **185**, 71–110.

Quintela-del-Río, A., and Estévez-Pérez, G., (2012). Non parametric kernel distribution function estimation with kerdiest: an R package for bandwidth choice and applications. *J. Statist. Softw*, **50**, 1–21.

Rao, B. L. S. P., (2005). Estimation of distribution and density functions by generalized Bernstein polynomials. *Indian J. Pure Appl. Math.*, **36**, 63–88.

Redner, R. A., and Walker, H. F., (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, **26**, 195–239.

Reiss, R. D., (1981). Nonparametric estimation of smooth distribution functions. *Scand. J. Stat.*, 116–119.

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., (1973). How to Apply the Method of Stochastic Approximation in the Nonparametric Estimation of a Regression Function. *Statistics*, **8**, 119–126.

Roeder K. and Wasserman L., (1997). Practical Bayesian density estimation using mixtures of normals. *J. Am. Stat. Assoc.*, **92**, 894–902.

Robbins, H. and Monro, S.,(1951). A Stochastic Approximation Method. *Ann. Math. Statist.*, **22**, 400–407.

Rosenblatt, M., (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.*, **27**, 832–837.

Ruppert, (1982). Almost Sure Approximations to the Robbins–Monro and Kiefer–Wolfowitz Processes with Dependent Noise. *Ann. Probab.*, **10**, 178-187.

Sarda, P., (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plann. Inference*, **35**, 65–75.

Salzer, H. E., (1972). Lagrangian interpolation at the Chebyshev points X_n, $\nu \equiv \cos(\nu\pi/n)$, $\nu = o(1)n$; some unnoted advantages. *Comput.J*, **15**, 156–159.

Serfling, R. J., (2009). Approximation theorems of mathematical statistics (Vol. 162). *John Wiley and Sons*

Schuster, E. F., (1985). Incorporating support constraints into nonparametric estimators of densities.. *Comm. Statist. Theory Methods*, **14**, 1123–1136.

Schwarz, G., (1978). Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464

Silverman, B. W., (1986). Density estimation for statistics and data analysis. *Chapman and Hall, London.*

Singh, R. S., Gasser, T., and Prasad, B., (1983). Nonparametric estimates of distribution functions. *Comm. Statist. Theory Methods*, **12**, 2095–2108.

Slaoui, Y., (2006). Application des méthodes d'approximations stochastiques à l'estimation de la densité et de la régression (Doctoral dissertation).

Slaoui, Y., (2014). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.*, **23**, 306–325.

Slaoui, Y., (2015a). Moderate deviation principles for recursive regression estimators defined by stochastic approximation method. *Int. J. Math. Stat.*, **16**, 51–60.

Slaoui, Y., (2015b). Large and moderate deviation principles for recursive regression estimators defined by stochastic approximation method. *Serdica Math. J.*, **41**, 307–328.

Slaoui, Y., (2015c). Plug-In Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.*, **69**, 483–509.

Slaoui, Y., (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface* **9**, 375–388.

Slaoui Y., (2018). Bias reduction in kernel density estimation. *J. Nonparametr. Stat.*, **30**, 505–522.

Slaoui, Y., (2019). Wild Bootstrap Bandwidth Selection of Recursive Nonparametric Relative Regression for Independent Functional Data, *J. Multivariate Anal.* **173**, 494–511.

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., (2020). Recursive nonparametric regression estimation for independent functional data *Statist. Sinica*, **30**, 417–437.

Smirnov, N. V. E., (1944). Approximate laws of distribution of random variables from empirical data. *Uspekhi Mat. Nauk*, **10**, 179–206.

Stein C., (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135–1151.

Tenbusch, A., (1994). Two-dimensional Bernstein polynomial density estimators. *Metrika*, **41** 233—253.

Tenbusch, A., (1997). Nonparametric curve estimation with Bernstein estimates. *Metrika*, **45** 1—30.

Tsybakov A. B. , (1990). Recurrent Estimation of the Mode of a Multidimensional Distribution. *Probl. Inf. Transm.*, **8**, 119-126.

Tsybakov, A. B., (2009). Lower bounds on the minimax risk. In Introduction to Nonparametric Estimation. *Springer, New York, NY.*, 77–135.

Vitale, R. A., (1975). A bernstein polynomial approach to density function estimation. *Stat. Inference Relat. Topics*, **2** 87—99..

Wand, M. P., and Jones, M. C., (1994). Kernel smoothing. *CRC press.*

Wang, F. K., (2000). A new model with bathtub-shaped failure rate using an additive Burr XII distribution. *Reliability Engineering and System Safety*, **70**, 305–312.

Watson, G. S., (1964). Smooth regression analysis. *Sankhyã. A*, **26**, 359–372.

Watson, G. S., and Leadbetter, M. R., (1964). Hazard analysis II. *Sankhya A*, **26**, 101–116.

Winter, B. B., (1973). Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.*, **1**, 247–253.

Winter, B. B., (1979). Convergence rate of perturbed empirical distribution functions. *J. Appl. Probab.*, **16**, 163–173.

Wooldridge, J. M., (2000). Basic regression analysis with time series data. *Introductory econometrics: a modern approach. Cincinnati, Ohio: South-Western College Publishing.*

Wu, C. J., (1983). On the convergence properties of the EM algorithm. *Ann. Statist.*, **11**, 95–103.

Yamato, H., (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.*, **15**, 69–78.

Zitouni, M., Zribi, M., and Masmoudi, A., (2018). Asymptotic properties of the estimator for a finite mixture of exponential dispersion models. *Filomat*, **32**, 6575–6598.