

## THÈSE

Pour l'obtention du grade de  
DOCTEUR DE L'UNIVERSITÉ DE POITIERS  
UFR des sciences fondamentales et appliquées  
Laboratoire de mathématiques et applications - LMA (Poitiers)  
(Diplôme National - Arrêté du 25 mai 2016)

École doctorale : Sciences et ingénierie pour l'information, mathématiques - S2IM (Poitiers)  
Secteur de recherche : Mathématiques et leurs interactions

Cotutelle : Université de Carthage, Tunisie

Présentée par :  
Asma Jemai

### **Estimation fonctionnelle non paramétrique au voisinage du bord**

Directeur(s) de Thèse :  
Julien Michel, Faouzi Chaabane, Yousri Slaoui

Soutenue le 16 mars 2018 devant le jury

Jury :

Président	<b>Abderrazek Karoui</b>	Professeur, Université de Carthage
Rapporteur	<b>Jean-François Dupuy</b>	Professeur, INSA, Rennes
Rapporteur	<b>Armelle Guillou</b>	Professeur, IRMA, Université de Strasbourg
Rapporteur	<b>Afif Masmoudi</b>	Professeur, Université de SFAX, Tunisie
Membre	<b>Julien Michel</b>	Professeur, Université de Poitiers
Membre	<b>Faouzi Chaabane</b>	Professeur, ENSTAB, Université de Carthage, Tunisie
Membre	<b>Yousri Slaoui</b>	Maître de conférences, Université de Poitiers
Membre	<b>Elias Ould Said</b>	Professeur, LMPA, Université du littoral, Calais

**Pour citer cette thèse :**

Asma Jemai. *Estimation fonctionnelle non paramétrique au voisinage du bord* [En ligne]. Thèse Mathématiques et leurs interactions. Poitiers : Université de Poitiers, 2018. Disponible sur Internet <<http://theses.univ-poitiers.fr>>



Université de Poitiers



Université de Carthage



Faculté des Sciences de Bizerte

## THÈSE EN CO-TUTELLE

pour l'obtention du titre de docteur de

La Faculté des Sciences de Bizerte, Université de Carthage

L'Université de Poitiers

Présentée par

**Asma JEMAI**

Spécialité : **Mathématiques appliquées**

### Estimation fonctionnelle non-paramétrique au voisinage du bord

Soutenue le : **16 Mars 2018** devant le Jury :

<b>Abderrazek KAROUI</b>	Pr. FSB Université de Carthage	Président
<b>Afif MASMOUDI</b>	Pr. FSS Université de Sfax	Rapporteur
<b>Armelle GUILLOU</b>	Pr. IRMA, Université de Strasbourg	2 <sup>ème</sup> Rapporteur Français
<b>Elias OULD SAÏD</b>	Pr. Université du Littoral, Calais	Examineur
<b>Faouzi CHAABANE</b>	Pr. ENSTAB Université de Carthage	Encadreur
<b>Jean-François DUPUY</b>	Pr. INSA de Rennes	Rapporteur
<b>Julien MICHEL</b>	Pr. Université de Poitiers	Encadreur
<b>Yousri SLAOUI</b>	MC-HDR, Université de Poitiers	Co-encadreur

Invités :

**Wassima DELLAGI** MA, Université de Carthage Co-encadreur

# Remerciements

Durant les années de thèse, j'ai été entourée de personnes qui ont contribué, à leur manière, à la réussite de ce travail. Je tiens très sincèrement à les en remercier.

Je voudrais exprimer ma gratitude à mes directeurs de thèse, Monsieur **Faouzi CHAABANE**, Professeur à l'Ecole Nationale des Sciences et Technologies Avancées de Borj Cedria, et Monsieur **Julien MICHEL**, Professeur à l'Université de Poitiers, et à mes co-directeurs de thèse Madame **Wassima DELLAGI** et Monsieur **Yousri SLAOUI**, Maître de Conférences-HDR à l'Université de Poitiers, qui ont dirigé mes premiers pas dans le monde de la recherche, pour leur disponibilité au cours de ces années et pour la grande liberté qu'ils ont laissée dans mes choix. Grâce à leurs conseils, j'ai pu mener à bien ce projet de thèse.

Je tiens également à exprimer toute ma reconnaissance à Monsieur **Abderrazek KAROUI**, Professeur à la Faculté des Sciences de Bizerte, qui me fait l'honneur de presider le jury de cette thèse.

Je suis reconnaissante à Monsieur **Afif MASMOUDI**, Professeur à la Faculté des Sciences de Sfax, à Monsieur **Jean-François DUPUY**, Professeur à l'Institut National des Sciences Appliquées de Rennes et à Madame **Armelle GUILLOU**, Professeur à l'Institut de Recherche Mathématique Avancée, Université de Strasbourg de l'intérêt qu'ils ont porté à mon travail et d'en être les rapporteurs de cette thèse.

J'exprime ma reconnaissance à Monsieur **Elias OULD SAÏD**, Professeur à l'Université du Littoral Côte d'Opale, d'avoir accepté de participer au jury.

J'adresse également mes remerciements à tous les membres de l'équipe du Laboratoire de Mathématiques et Applications de Poitiers, aux près desquels j'ai trouvé une seconde famille pendant ces années de travail sur ma thèse, et particulièrement l'équipe de Probabilités et Statistiques, pour l'ambiance conviviale qui régnait au sein de notre laboratoire.

Je voudrais aussi exprimer ma reconnaissance à tous mes professeurs qui, tout au long de mon cursus, ont su me transmettre leur passion pour les mathématiques, et qui m'ont ainsi communiqué le désir de travailler dans ce domaine.

Je remercie en particulier mon enseignant du Lycée Pilote de Gabès Monsieur **Hedi ABDERRAHIM**, mes enseignants du cycle agrégatif à l'IPEST et mes enseignants de l'ENS, où j'ai passé des moments inoubliables.

Mes remerciements s'adressent également à mes amis qui ont toujours été là. Je voudrais aussi mentionner mes collègues de l'ISSAT de Gabès en particulier mes "Extra Sisters" **Maroua SOULA** et **Yemna LABIDI**.

Mes remerciements les plus profonds et les plus sincères vont certainement à ma famille à qui je dédie ce travail. Merci à ma chère mère **Mounira MRABET**. Aucune dédicace ne peut traduire l'affection et l'amour que je lui porte. Aucun hommage ne saurait être à la hauteur de ses sacrifices, de sa tendresse et de ses encouragements. Merci également à mon bien-aimé père **Rachid JMAEI** qui peut être fier et trouver ici le résultat de longues années de sacrifices qui m'a toujours aidé à avancer. Merci pour les valeurs nobles, l'éducation et le soutien permanent de mes chers parents. Merci à mes deux soeurs **Nesrine** et **Sarra** qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

J'exprime toute ma gratitude, mon estime et mon attachement à mes grands-parents. Puisse dieu vous accorder santé, longue vie et prospérité. Merci à mes tantes, oncles en particulier ma tante **Najoua** et mon oncle **Hedi** pour leur soutien, leur hospitalité et leur générosité. Enfin, une dédicace spéciale à mon fiancé **Oussama LASSOUED**, je n'oublierais pas les encouragements et la compréhension dont il a fait preuve dans les moments difficiles.

A tous ceux que je viens de citer, et à tous ceux que j'aurais oublié, Merci du fond du cœur.

## Résumé

L'objectif de cette thèse est de construire des estimateurs non-paramétriques d'une fonction de distribution, d'une densité de probabilité et d'une fonction de régression en utilisant les méthodes d'approximations stochastiques afin de corriger l'effet du bord créé par les estimateurs à noyaux continus classiques. Dans le premier chapitre, on donne quelques propriétés asymptotiques des estimateurs continus à noyaux, puis, on présente l'algorithme stochastique de Robbins-Monro qui permet d'introduire les estimateurs récursifs. Enfin, on rappelle les méthodes utilisées par Vitale, Leblanc et Kakizawa pour définir des estimateurs d'une fonction de distribution et d'une densité de probabilité en se basant sur les polynômes de Bernstein. Dans le deuxième chapitre, on a introduit un estimateur récursif d'une fonction de distribution en se basant sur l'approche de Vitale. On a étudié les propriétés de cet estimateur : biais, variance, erreur quadratique intégrée (*MISE*) et on a établi sa convergence ponctuelle faible. On a comparé la performance de notre estimateur avec celle de Vitale et on a montré qu'avec le bon choix du pas et de l'ordre qui lui correspond notre estimateur domine en terme de *MISE*. On a confirmé ces résultats théoriques à l'aide des simulations. Pour la recherche pratique de l'ordre optimal, on a utilisé la méthode de validation croisée. Enfin, on a confirmé les meilleures qualités de notre estimateur à l'aide de données réelles. Dans le troisième chapitre, on a estimé une densité de probabilité d'une manière récursive en utilisant toujours les polynômes de Bernstein. On a donné les caractéristiques de cet estimateur et on les a comparées avec celles de l'estimateur de Vitale, de Leblanc et l'estimateur donné par Kakizawa en utilisant la méthode multiplicative de correction du biais. On a appliqué notre estimateur sur des données réelles. Dans le quatrième chapitre, on a introduit un estimateur récursif et non récursif d'une fonction de régression en utilisant les polynômes de Bernstein. On a donné les caractéristiques de ces estimateurs et on les a comparées avec celles des estimateurs à noyau classiques. Ensuite, on a utilisé nos estimateurs pour interpréter des données réelles.

**Mots clés :** Algorithme stochastique, Effet du bord, Estimation non-paramétrique, Polynômes de Bernstein, Validation croisée.

## Abstract

The aim of this thesis is to construct nonparametric estimators of distribution, density and regression functions using stochastic approximation methods in order to correct the edge effect created by kernels estimators. In the first chapter, we give some asymptotic properties of kernel estimators. Then, we introduce the Robbins-Monro stochastic algorithm which creates the recursive estimators. Finally, we recall the methods used by Vitale, Leblanc and Kakizawa to define estimators of distribution and density functions based on Bernstein polynomials. In the second chapter, we introduced a recursive estimator of a distribution function based on Vitale's approach. We studied the properties of this estimator : bias, variance, mean integrated squared error (*MISE*) and we established a weak pointwise convergence. We compared the performance of our estimator with that of Vitale and we showed that, with the right choice of the stepsize and its corresponding order, our estimator dominates in terms of *MISE*. These theoretical results were confirmed using simulations. We used the cross-validation method to search the optimal order. Finally, we applied our estimator to interpret real dataset. In the third chapter, we introduced a recursive estimator of a density function using Bernstein polynomials. We established the characteristics of this estimator and we compared them with those of the estimators of Vitale, Leblanc and Kakizawa. To highlight our proposed estimator, we used real dataset. In the fourth chapter, we introduced a recursive and non-recursive estimator of a regression function using Bernstein polynomials. We studied the characteristics of these estimators. Then, we compared our proposed estimators with the classical kernel estimators using real dataset.

**Keywords** : Stochastic algorithm, Edge effect, Nonparametric estimation, Bernstein polynomials, Cross validation.

# Table des matières

<b>Introduction Générale</b>	<b>12</b>
<b>1 Concepts de Base : Estimateurs et Algorithme</b>	<b>19</b>
1.1 Estimateurs à noyaux continus classiques . . . . .	20
1.1.1 Estimation d'une densité . . . . .	21
1.1.2 Estimation d'une fonction de régression . . . . .	28
1.2 Algorithmes stochastiques et estimateurs récursifs . . . . .	35
1.2.1 Estimateurs récursifs d'une densité . . . . .	36
1.2.2 Estimateurs récursifs d'une fonction de régression . . . . .	42
1.2.3 Problèmes de support . . . . .	46
1.3 Estimateurs fonctionnels utilisant les polynômes de Bernstein . . . . .	48
1.3.1 Estimateur d'une distribution . . . . .	50
1.3.2 Estimateurs d'une densité . . . . .	55
<b>2 Recursive distribution estimator using Bernstein polynomials</b>	<b>64</b>



2.1	Introduction . . . . .	64
2.2	Assumptions and Notations . . . . .	66
2.3	Main Results . . . . .	68
2.4	Numerical studies . . . . .	72
2.4.1	Simulations . . . . .	72
2.4.2	Real dataset . . . . .	73
2.5	Conclusion . . . . .	74
2.6	Proofs . . . . .	75
2.6.1	Proof of Proposition 2.3.1 . . . . .	75
2.6.2	Proof of Proposition 2.3.2 . . . . .	77
2.6.3	Proof of Corollary 2.3.1 . . . . .	77
2.6.4	Proof of Theorem 2.3.1 . . . . .	79
<b>3</b>	<b>Recursive density estimator using Bernstein polynomials</b>	<b>88</b>
3.1	Introduction . . . . .	88
3.2	Assumptions and Notations . . . . .	91
3.3	Main Results . . . . .	93
3.3.1	Within the interval $[0, 1]$ . . . . .	93
3.3.2	The edges of the interval $[0, 1]$ . . . . .	95
3.3.3	<i>MISE</i> of $f_n$ . . . . .	97

3.4	Applications . . . . .	103
3.4.1	Simulations . . . . .	104
3.4.2	Real dataset . . . . .	106
3.5	Conclusion . . . . .	110
3.6	Proofs . . . . .	111
3.6.1	Proof of Proposition 3.3.1 and Proposition 3.3.2 . . . . .	111
3.6.2	Proof of Proposition 3.3.3 . . . . .	113
3.6.3	Proof Corollary 3.3.1 . . . . .	114
3.6.4	Proof Theorem 3.3.1 and Theorem 3.3.2 . . . . .	115
<b>4</b>	<b>Regression estimators using Bernstein polynomials</b>	<b>125</b>
4.1	Introduction . . . . .	125
4.2	Assumptions and Notations . . . . .	126
4.3	Non-Recursive estimator . . . . .	128
4.4	Recursive estimator . . . . .	130
4.4.1	Within the interval $[0, 1]$ . . . . .	131
4.4.2	The edges of the interval $[0, 1]$ . . . . .	133
4.4.3	The <i>MISE</i> of $r_n$ . . . . .	135
4.5	Applications . . . . .	136
4.5.1	Simulations . . . . .	137

4.5.2	Real dataset . . . . .	139
4.6	Conclusion . . . . .	140
4.7	Outlines of the proofs . . . . .	140
4.7.1	Proof of Proposition 4.3.1 . . . . .	142
4.7.2	Proof of Proposition 4.3.2 . . . . .	147
4.7.3	Proof of Theorem 4.3.1 . . . . .	151
4.7.4	Proof of the results obtained for $r_n$ . . . . .	154
	<b>Conclusion et perspectives</b>	<b>178</b>
	<b>Bibliographie</b>	<b>179</b>

# Liste des tableaux

2.1	Average <i>ISE</i> of Bernstein distribution estimators small sizes . . . . .	81
2.2	Average <i>ISE</i> of Bernstein distribution estimators big sizes . . . . .	82
3.1	Average <i>ISE</i> of Vitale's estimator and our density estimator . . .	119
3.2	Average <i>ISE</i> of Bernstein density estimators . . . . .	120
4.1	Average <i>ISE</i> of Bernstein and Kernel regression estimators . . . . .	175

# Table des figures

2.1	Bernstein estimators for a beta distribution . . . . .	83
2.2	Bernstein estimators for a beta mixture distribution . . . . .	84
2.3	Bernstein estimators for a truncated normal distribution . . . . .	85
2.4	Bernstein estimators for a truncated normal mixture distribution . . .	86
2.5	Distribution estimates for the suicide data . . . . .	87
3.1	Bernstein estimators of a beta density . . . . .	121
3.2	Density estimates for the prostate-specific antigen data . . . . .	122
3.3	Density estimates for the Old Faithful data . . . . .	123
3.4	Density estimates for the tuna data . . . . .	124
4.1	Bernstein regression estimators for cos model . . . . .	176
4.2	Regression estimates for daily carbon dioxide measurements data . .	177

# Introduction Générale

La théorie de l'estimation est une des préoccupations majeures des statisticiens. L'estimation de la densité, par exemple est un sujet qui a donné lieu à un grand nombre de travaux à cause de son application dans des divers domaines, comme l'analyse de la régression des séries chronologiques et la théorie de fiabilité.

En effet, Singh [102] s'est basé sur l'estimation de la densité pour approcher l'information de Fisher. Aussi, dans des nombreux domaines (médecine, économie, sociologie, environnement, finances, marketing) surviennent des phénomènes complexes à modéliser. Ces derniers sont souvent décrits par une variable aléatoire réelle (v.a.r.) inconnue, d'où la nécessité d'estimer sa loi de probabilité.

On trouve deux approches d'estimation dans la littérature, la première est l'approche paramétrique qui se résume en l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon. Dans ce cas, on construit les estimateurs en utilisant soit la méthode des moments soit celle du maximum de vraisemblance ou encore celle des moindres carrés. Cependant, l'approche non-paramétrique consiste, généralement à estimer à partir des observations une fonction inconnue élément d'une certaine classe fonctionnelle. Cette branche de la statistique a subi un intérêt croissant autant sur le plan théorique que pratique. La méthode la plus ancienne dans cette approche est l'histogramme qui présente graphiquement la répartition des données. Cette représentation fournit des informations sur les observations étudiées mais elle reste insuffisante et limitée pour estimer une densité raisonnablement lisse.

En fait, les points près du bords et les points près du milieu du support de la densité à estimer reçoivent le même poids d'où la variabilité des interprétations statistiques que l'on peut faire d'un histogramme suivant l'origine choisie. D'où l'introduction de nouveaux estimateurs appelés estimateurs à noyau qui ont rectifié ce problème. Ces estimateurs ont été depuis amélioré en utilisant les méthodes d'approximations stochastiques dans le but de faciliter la mise à jour de l'estimateur lorsque en passe d'un échantillon de taille  $n$  à un échantillon de taille  $n + 1$ .

Malgré leur intérêt apporté au domaine de la statistique, les estimateurs à noyaux connaissent quelques inconvénients parmi lesquels se situent les problèmes de support. Le plus connu est l'effet du bord dans l'estimation des fonctions à support borné au moins d'un côté. Plusieurs solutions à ce problème ont été proposées dans la littérature.

Dans ce cadre, l'objectif de notre travail est de construire des estimateurs non-paramétriques d'une distribution d'une densité et d'une fonction de régression en utilisant les méthodes d'approximations stochastiques afin de corriger l'effet du bord créé par les estimateurs à noyaux continus classiques.

Dans le premier chapitre, on donne quelques propriétés asymptotiques des estimateurs continus à noyaux. Rappelons que Rosenblatt [88] a introduit un estimateur de la densité d'un échantillon de variables aléatoires réelles  $X_1, X_2, \dots, X_n$

$$\forall x \in \mathbb{R}, \quad \widehat{f}_n(x) = \frac{1}{2nh_n} \sum_{i=1}^n \mathbb{I}_{\{x-h_n < X_i \leq x+h_n\}},$$

avec  $(h_n)$  est une suite dite fenêtre vérifiant  $\lim_{n \rightarrow \infty} h_n = 0$ .

Parzen [76] a suggéré une généralisation de cet estimateur

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

où  $K : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction positive, intégrable, telle que  $\int_{\mathbb{R}} K(u) du = 1$ , appelée noyau.

Pour une fonction de régression, il existe plusieurs types d'estimateurs à noyau dont le plus célèbre est celui de Nadaraya [72] et [120]. Pour un échantillon de variables aléatoires réelles  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  vérifiant

$$\forall i \in 1, \dots, n, \quad Y_i = r(X_i) + \varepsilon_i,$$

avec  $\varepsilon_1, \dots, \varepsilon_n$  est une suite de variables aléatoires i.i.d jouant le rôle de bruit. L'estimateur de la fonction de régression  $r$ , est donné par

$$\forall x \in \mathbb{R}, \quad \hat{r}_n^{NW}(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \neq 0 \\ \frac{1}{n} \sum_{i=1}^n Y_i & \text{sinon.} \end{cases}$$

On présente en seconde partie l'algorithme stochastique de Robbins et Monro [87] qui permet d'introduire les estimateurs récursifs de la densité et la fonction de régression.

Pour la densité, on cherche le zéro de la fonction  $h : y \mapsto f(x) - y$  avec  $f$  est la densité à estimer d'un échantillon  $X_1, \dots, X_n$  de variables aléatoires réelles. Cet algorithme est défini comme suit :

- (i) Pour  $x \in \mathbb{R}$ , on fixe  $f_0(x) \in \mathbb{R}$ ,
- (ii) pour  $n \geq 1$ , on a

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

où  $(\gamma_n)$  est une suite positive qui tend vers zéro à l'infini dite pas de l'algorithme.

Pour la fonction de régression, on cherche le zéro de la fonction  $h : y \mapsto f(x)r(x) - f(x)y$  avec  $f$  est la densité de  $X_1, \dots, X_n$  et  $r$  est la régression à estimer.



(i) Pour  $x \in \mathbb{R}$ , on fixe  $r_0(x) \in \mathbb{R}$ ,

(ii) pour  $n \geq 1$ , on obtient

$$\begin{aligned} r_n^{GR}(x) &= (1 - \gamma_n f(x))r_{n-1}(x) + \gamma_n \left[ f(x) - h_n^{-1}K \left( \frac{x - X_n}{h_n} \right) \right] r_{n-1}(x) \\ &\quad + \gamma_n h_n^{-1} Y_n K \left( \frac{x - X_n}{h_n} \right). \end{aligned}$$

L'avantage de ces estimateurs par rapport aux estimateurs non récursifs est le gain de temps de calcul. Malheureusement, ils ne donnent pas une solution pour l'effet du bord. Pour résoudre ce problème il existe d'autres méthodes, parmi lesquelles on trouve les estimateurs de la distribution et de la densité introduits par Vitale [119] en se basant sur le théorème de Weistrass.

Il a estimé la distribution  $F$  par

$$\forall x \in [0, 1], \quad \tilde{F}_{n,m}(x) = \sum_{k=0}^m F_n \left( \frac{k}{m} \right) b_k(m, x),$$

où  $\forall t \in \mathbb{R}$ ,  $F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq t\}}$  est la distribution empirique et

$$\forall k \in \{0, \dots, m\}, \quad b_k(m, x) = \binom{m}{k} x^k (1-x)^{m-k},$$

sont les polynômes de Bernstein d'ordre  $m$  dans  $\mathbb{N}^*$ .

Pour la densité, il suffit de dériver  $\tilde{F}_{n,m}$  pour obtenir

$$\forall x \in [0, 1], \quad \tilde{f}_{m,n}(x) = m \sum_{k=0}^{m-1} \left[ F_n \left( \frac{k+1}{m} \right) - F_n \left( \frac{k}{m} \right) \right] b_k(m-1, x).$$

Pour utiliser les polynômes de Bernstein, nous devons considérer des fonctions à support dans  $[0, 1]$ . Cependant, à l'aide des transformations simples, on peut toujours se ramener à l'intervalle unité.

Dans le but de réduire le biais de l'estimateur de Vitale, Leblanc [55] a introduit un nouvel estimateur pour la densité qui est donné par

$$\forall x \in [0, 1], \quad \tilde{f}_{m,m/2,n}(x) = 2\tilde{f}_{m,n}(x) - \tilde{f}_{m/2,n}(x).$$

Kakizawa (2014) a généralisé cet estimateur pour devenir :

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,m/b}(x) = \frac{b}{b-1} \tilde{f}_{n,m}(x) - \frac{1}{b-1} \tilde{f}_{n,m/b}(x),$$

avec  $b = 2, 3, \dots$ ,  $\tilde{f}_{n,m}$  et  $\tilde{f}_{n,m/b}$  sont les estimateurs de Vitale d'ordre  $m$  and  $m/b$  respectivement.

Ces estimateurs ont un biais plus petit que celui de l'estimateur de Vitale, mais ils perdent la positivité. Pour rectifier ce problème, Terrell et Scott [116] ont développé une méthode multiplicative de correction du biais. Cette méthode a été adaptée par Hirukawa [41] sur l'estimateur de noyau bêta introduit par Chen [14] et par Igarashi and Kakizawa [42] dans l'estimation par les polynômes de Bernstein pour obtenir l'estimateur suivant :

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,b,\varepsilon}(x) = \left\{ \tilde{f}_{n,m}(x) \right\}^{b/(b-1)} \left\{ \tilde{f}_{n,m/b}(x) + \varepsilon \right\}^{-1/b-1},$$

avec  $0 < \varepsilon = \varepsilon(m) \rightarrow 0$ .

Cet estimateur  $\tilde{f}_{n,m,b,\varepsilon}$  conserve la positivité mais ne constitue pas une densité.

En fait, l'estimateur  $\tilde{f}_{n,m,b,\varepsilon}$  ne s'intègre généralement pas à l'unité. Pour le rectifier, Igarashi and Kakizawa [42] ont proposé l'estimateur normalisé suivant :

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,b,\varepsilon}^N(x) = \frac{\tilde{f}_{n,m,b,\varepsilon}(x)}{\int_0^1 \tilde{f}_{n,m,b,\varepsilon}(y) dy}.$$

Dans le deuxième chapitre, on a introduit un estimateur récursif d'une distribution en utilisant l'algorithme de Robbins-Monro et en se basant sur l'approche de Vitale. Cet estimateur est défini par :

- (i)  $\forall x \in [0, 1], F_0(x) \in \mathbb{R}$  fixé,
- (ii) pour  $n \geq 1$ , on a

$$F_n(x) = (1 - \gamma_n)F_{n-1}(x) + \gamma_n \sum_{k=0}^m \mathbb{I}_{\{X_n \leq \frac{k}{m}\}} b_k(m, x).$$

On a étudié les propriétés de cet estimateur : biais, variance, erreur quadratique intégrée (*MISE*) et on a établi sa convergence ponctuel faible. On a déterminé l'ordre  $m$  optimal qui minimise la *MISE*. On a comparé la performance de notre estimateur avec celui de Vitale et on a montré qu'avec le bon choix du pas ( $\gamma_n$ ) et l'ordre ( $m_n$ ) qui lui correspond, notre estimateur domine en terme de *MISE*.

On a confirmé ces résultats théoriques à l'aide des simulations qui ont montré que la

courbe de notre estimateur est plus proche de la vraie courbe que celle de l'estimateur de Vitale. Pour la recherche pratique de l'ordre optimal  $(m_n)$ , on a utilisé la méthode de validation croisée. Enfin, on a confirmé les meilleures qualités de notre estimateur à l'aide des données réelles de suicide fournies par Silverman [101].

Ces données se composent de durées (en jours) de traitement psychiatrique pour 86 patients utilisés comme témoins dans une étude sur les risques de suicide.

Dans le troisième chapitre, on a approximé une densité  $f$  d'une manière récursive en posant :

(i)  $\forall x \in [0, 1]$ ,  $f_0(x) \in \mathbb{R}$  fixé

(ii) pour  $n \geq 1$ , on a

$$f_n(x) = (1 - \gamma_n)f_{n-1}(x) + \gamma_n \left\{ 2m \sum_{k=0}^{m-1} \left[ \mathbb{I}_{\{X_n \leq \frac{k+1}{m}\}} - \mathbb{I}_{\{X_n \leq \frac{k}{m}\}} \right] b_k(m-1, x) - \frac{m}{2} \sum_{k=0}^{\frac{m}{2}-1} \left[ \mathbb{I}_{\{X_n \leq \frac{2(k+1)}{m}\}} - \mathbb{I}_{\{X_n \leq \frac{2k}{m}\}} \right] b_k\left(\frac{m}{2}-1, x\right) \right\}.$$

On a donné les caractéristiques de cet estimateur et on les a comparées avec ceux de l'estimateur de Vitale  $f_{m,n}$  de l'estimateur de Leblanc  $f_{n,m,m/2}$  et l'estimateur donné par Kakizawa en utilisant la méthode multiplicative de correction du biais. On a montré que notre estimateur de densité constitue une bonne approximation près des bords. En plus, cet estimateur récursif  $f_n$  a un taux réduit du biais (un taux de  $m^{-2}$ ) que celui de l'estimateur de Vitale (un taux de  $m^{-1}$ ) et il a une *MISE* plus petite que celle de  $f_{m,n}$ . Pour mettre en évidence notre estimateur proposé, on a examiné d'abord les données provenant d'une étude sur le cancer de prostate (Stamey et al. [112]). Ces données ont été divisées en deux parties : un ensemble de formation avec 67 observations et un ensemble de tests avec 30 observations. On a appliqué notre estimateur  $f_n$  sur les données de formation pour l'antigène spécifique de prostate. Comme deuxième exemple, on a pris les données "Old Faithful" de Silverman [101]. En dernier, on a étudié les "tuna data" de Chen [12].

Dans le quatrième chapitre, on a proposé deux estimateurs d'une fonction de régression. Le premier est l'estimateur non-récuratif donné pour tout  $x \in [0, 1]$  tel que  $f(x) \neq 0$  par

$$\widehat{r}_n(x) = \frac{\sum_{i=1}^n Y_i \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x)}{\sum_{i=1}^n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x)}.$$

Le deuxième, on l'a construit d'une manière récursive en posant : Pour  $x \in [0, 1]$  tel que  $f(x) \neq 0$ , on fixe  $r_0(x) \in \mathbb{R}$  et pour  $n \geq 1$ , on obtient

$$\begin{aligned} r_n(x) &= (1 - \gamma_n f(x)) r_{n-1}(x) + \gamma_n \left( f(x) - m_n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x) \right) r_{n-1}(x) \\ &\quad + \gamma_n m_n Y_n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x). \end{aligned}$$

On a donné les caractéristiques de cet estimateur et on les a comparées avec ceux de l'estimateur de Nadaraya-Watson  $r_n^{NW}$  et l'estimateur Révész généralisé  $r_n^{GR}$ . On a montré que notre estimateur non récursif  $\widehat{r}_n$  a la plus petite *MISE* dans la plupart des modèles. Enfin, on a considéré le jeu de données CO2 disponible dans le R package `StatData2` qui contient 237 observations avec deux variables : Jour et CO2. Des Scientifiques à une station de recherche à Brotjacklriegel en Allemagne, ont enregistré les niveaux de CO2, en parties par million, dans l'atmosphère pour chaque jour du début d'avril jusqu'à novembre 2001. On a appliqué nos estimateurs pour interpréter ces données réelles et on a montré qu'ils peuvent mener à des estimations très satisfaisantes, particulièrement près du bord.

Les perspectives visées après la thèse sont : établir la convergence presque sûre et appliquer le principe de grandes déviations et de déviations modérées pour nos estimateurs et la généralisation de ces estimateurs dans le cas multidimensionnel. Par ailleurs, on va continuer la recherche dans le cadre d'estimation semi-réursive d'une fonction de régression.

# Chapitre 1

## Concepts de Base : Estimateurs et Algorithme

On s'intéresse aux problèmes d'estimation d'une distribution, d'une densité et d'une fonction de régression. C'est à dire trouver une approximation de ces fonctions à partir des données observées. La première approche est celle de l'estimation paramétrique lorsque la loi inconnue est décrite par un paramètre de dimension finie. Dans ce cas, on construit les estimateurs en utilisant soit la méthode ou encore celle des moindres carrés. L'avantage de cette approche est qu'elle utilise des algorithmes simples. Par opposition, dans l'approche non-paramétrique, on ne fait aucune hypothèse sur la fonction à estimer. Dans cette approche la méthode la plus ancienne et la plus naturelle est celle de l'histogramme qui est une représentation graphique particulière de la répartition des observations. Cette approche convient bien pour des analyses relativement grossières. Néanmoins ses discontinuités n'apparaissent pas très naturelles. En fait, pour des densités raisonnablement régulières, par exemple, deux fois continûment différentiable, l'histogramme est sévèrement limité. Pour remédier à ce problème, Rosenblatt [88] a introduit la méthode d'estimation par les noyaux. Cette méthode a été améliorée après par Parzen [76]. De même, en utilisant les noyaux, Nadaraya [72] et [120] ont introduit un estimateur pour une fonction de régression.

Dans ce chapitre on rappelle quelques résultats concernant les estimateurs à noyaux continus d'une densité et d'une fonction de régression. Dans un premier temps, on présente leurs propriétés de base puis les méthodes de choix des fenêtres. Dans un second temps, on parle de l'effet du bord l'inconvénient majeur de cette méthode d'estimation lorsque le support de la fonction à estimer n'est plus  $\mathbb{R}$ . On rappelle après les estimateurs utilisant les polynômes de Bernstein pour corriger cet effet.

## 1.1 Estimateurs à noyaux continus classiques

On présente, dans cette section, la méthode d'estimation par les noyaux continus pour une densité et une fonction de régression. On définit un noyau continu par :

### Définition 1.1.1.

Une fonction  $K$  à support  $\mathbb{S}$  est dite un noyau si elle est une fonction positive, intégrable, telle que  $\int_{\mathbb{S}} \mathbf{K}(\mathbf{u}) d\mathbf{u} = \mathbf{1}$  ( on dit qu'elle est une densité de probabilité), symétrique ( $\mathbf{K}(-\mathbf{u}) = \mathbf{K}(\mathbf{u})$ ), de moyenne  $\mu_K$  nulle ( $\mu_{\mathbf{K}} = \int_{\mathbb{S}} \mathbf{u} \mathbf{K}(\mathbf{u}) d\mathbf{u} = \mathbf{0}$ ), de variance  $\sigma_K^2$  finie ( $\sigma_{\mathbf{K}}^2 = \int_{\mathbb{S}} \mathbf{u}^2 \mathbf{K}(\mathbf{u}) d\mathbf{u} < +\infty$ ) et de carré intégrable ( $\int_{\mathbb{S}} \mathbf{K}^2(\mathbf{u}) d\mathbf{u} < +\infty$ ).

On donne quelques exemples de noyaux à support  $\mathbb{S} = [-1, 1]$  ou bien  $\mathbb{S} = \mathbb{R}$ .

$$K(u) = \frac{1}{2} \mathbb{I}_{\{|u| \leq 1\}} : \text{noyau rectangulaire,}$$

$$K(u) = (1 - |u|) \mathbb{I}_{\{|u| \leq 1\}} : \text{noyau trinagulaire,}$$

$$K(u) = \frac{3}{4} (1 - |u|^2) \mathbb{I}_{\{|u| \leq 1\}} : \text{noyau parabolique ou d'Epanechnikov,}$$

$$K(u) = \frac{15}{16} (1 - |u|^2)^2 \mathbb{I}_{\{|u| \leq 1\}} : \text{noyau "biweight",}$$

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) : \text{noyau gaussien,}$$

$$K(u) = \frac{1}{2} \exp\left(-\frac{|u|}{\sqrt{2}}\right) \sin\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right) : \text{noyau de Silverman.}$$

Le lecteur peut éventuellement se référer à Scott et al. [99], Epanechnikov [28], Tsybakov [118] et Silverman [101] pour plus de détails.

### 1.1.1 Estimation d'une densité

Pour l'estimation d'une densité, on considère un échantillon  $X_1, \dots, X_n$  de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de densité  $f$  à support  $\mathbb{T}$ .

#### Définition 1.1.2.

Soit la fenêtre de lissage  $(h_n)$ , une suite réelle positive qui tend vers 0 lorsque  $n \rightarrow +\infty$ , et  $K$  la fonction noyau. L'estimateur à noyau continu de  $f$  est défini en un point  $x \in \mathbb{T}$  par :

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right). \quad (1.1.1)$$

Dans l'expression (1.1.1), le choix de la fonction noyau a peu d'impact sur la qualité d'estimation. Cependant, le paramètre de lissage  $h_n$  a un impact très significatif sur les résultats. Lorsqu'il s'approche de 0, la fonction  $K$  se comporte comme la fonction de Dirac ce qui nous donne plus d'informations (sur lissage); tandis que, lorsque  $h_n$  devient grand, on a moins d'informations (sous lissage). Ainsi, pour obtenir une estimation plus proche de la réalité on doit avoir un choix optimal de  $h_n$ .

#### (A) Erreur Quadratique Moyenne.

Il y a un certain nombre de critères qui permettent d'évaluer la qualité de l'estimateur  $\widehat{f}_n$ . Parmi ces critères proposés dans la littérature, figure l'erreur quadratique moyenne ponctuelle ("Mean Squared Error",  $MSE$ ).

Ainsi, pour l'estimateur  $\widehat{f}_n$  défini en (1.1.1), on a :

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

En développant cette expression, on obtient

$$\forall x \in \mathbb{T}, \quad MSE \left[ \widehat{f}_n(x) \right] = \text{Biais}^2 \left[ \widehat{f}_n(x) \right] + \text{Var} \left[ \widehat{f}_n(x) \right],$$

avec  $Biais \left[ \widehat{f}_n(x) \right]$  et  $Var \left[ \widehat{f}_n(x) \right]$  sont respectivement le biais et la variance ponctuels de l'estimateur  $\widehat{f}_n$  définis pour tout point  $x \in \mathbb{T}$  par :

$$\begin{aligned} Biais \left[ \widehat{f}_n(x) \right] &= \widehat{f}_n(x) - \mathbb{E} \left[ \widehat{f}_n(x) \right], \\ Var \left[ \widehat{f}_n(x) \right] &= \mathbb{E} \left[ \left\{ \widehat{f}_n(x) - \mathbb{E} \left[ \widehat{f}_n(x) \right] \right\}^2 \right] = \mathbb{E} \left[ \widehat{f}_n^2(x) \right] - \mathbb{E}^2 \left[ \widehat{f}_n(x) \right]. \end{aligned}$$

L'espérance et la variance de l'estimateur  $\widehat{f}_n$  en un point  $x$  de  $\mathbb{T}$  sont exprimées en fonction du noyau classique  $K$  et de la densité à estimer  $f$  comme suit :

$$Biais \left[ \widehat{f}_n(x) \right] = \frac{1}{h_n} \mathbb{E} \left[ K \left( \frac{x - X_1}{h_n} \right) \right] = \frac{1}{h_n} \int_{\mathbb{T}} K \left( \frac{x - t}{h_n} \right) f(t) dt, \quad (1.1.2)$$

et

$$\begin{aligned} Var \left[ \widehat{f}_n(x) \right] &= \frac{1}{nh_n^2} \left\{ \mathbb{E} \left[ K^2 \left( \frac{x - X_1}{h_n} \right) \right] - \mathbb{E}^2 \left[ K \left( \frac{x - X_1}{h_n} \right) \right] \right\}, \\ &= \frac{1}{nh_n^2} \int_{\mathbb{T}} K^2 \left( \frac{x - t}{h_n} \right) f(t) dt \\ &\quad - \frac{1}{nh_n^2} \left( \int_{\mathbb{T}} K \left( \frac{x - t}{h_n} \right) f(t) dt \right)^2. \end{aligned} \quad (1.1.3)$$

Une mesure, globale, de l'efficacité de l'estimateur  $\widehat{f}_n$  est obtenue en intégrant la  $MSE$  sur tout le support  $\mathbb{T}$  de  $f$ . Il s'agit de "l'Erreur Quadratique Moyenne Intégrée" ("Means Integrated Squared Error",  $MISE$ ). Elle s'écrit :

$$\begin{aligned} MISE(\widehat{f}_n) &= \int_{\mathbb{T}} MSE(x) dx \\ &= \int_{\mathbb{T}} Biais^2 \left[ \widehat{f}_n(x) \right] dx + \int_{\mathbb{T}} Var \left[ \widehat{f}_n(x) \right] dx. \end{aligned}$$

En dépit des conditions vérifiées par le noyau classique dans la Définition 1.1.1, nous supposons dans la suite que la densité à estimer  $f$  est bornée, deux fois différentiable, et  $f''$  est bornée. En reprenant l'expression de l'espérance de l'estimateur  $\widehat{f}_n$  dans (1.1.2) et la variance dans (1.1.3) puis en posant  $t = x - h_n u$ , on a :

$$\mathbb{E} \left[ \widehat{f}_n(x) \right] = \int_{\mathbb{T}} K(u) f(x - h_n u) du, \quad (1.1.4)$$

et

$$\begin{aligned} Var \left[ \widehat{f}_n(x) \right] &= \frac{1}{nh_n} \int_{\mathbb{T}} K^2(u) f(x - h_n u) du \\ &\quad - \frac{1}{nh_n^2} \mathbb{E}^2 \left[ K \left( \frac{x - X_1}{h_n} \right) \right]. \end{aligned} \quad (1.1.5)$$



Le développement en séries de Taylor de  $f(x - h_n u)$  au voisinage de  $x$  est

$$f(x - h_n u) = f(x) - h_n u f'(x) + \frac{1}{2} h_n^2 u^2 f''(x) + o_x(h_n^2 u^2).$$

Il convient de noter qu'ici, et tout au long de cette thèse, on utilise  $o$  lorsque l'erreur d'approximation est uniforme par rapport à  $x$  et on écrit  $o_x$  pour souligner que l'erreur dépend de  $x$ .

En injectant le dernier développement dans (1.1.4) et (1.1.5), on obtient :

$$\mathbb{E} [\widehat{f}_n(x)] = f(x) + \frac{1}{2} h_n^2 f''(x) \sigma_K^2 + o(h_n^2), \quad (1.1.6)$$

et

$$Var [\widehat{f}_n(x)] = \frac{1}{nh_n} f(x) \int_{\mathbb{S}} K^2(u) du + o\left(\frac{1}{nh_n}\right). \quad (1.1.7)$$

De ces deux derniers résultats (1.1.6) et (1.1.7), on peut déduire les formes approximées et asymptotiques de  $MSE$  et de  $MISE$  notées respectivement  $AMSE$  et  $AMISE$  par :

$$AMSE [\widehat{f}_n(x)] = \frac{1}{nh_n} f(x) \int_{\mathbb{S}} K^2(u) du + \frac{1}{4} h_n^4 \sigma_K^4 (f''(x))^2, \quad (1.1.8)$$

et

$$AMISE (\widehat{f}_n) = \frac{1}{nh_n} \int_{\mathbb{S}} K^2(u) du + \frac{1}{4} h_n^4 \sigma_K^4 \int_{\mathbb{T}} (f''(x))^2 dx. \quad (1.1.9)$$

L'une des qualités recherchées pour un estimateur est l'étude des types de convergence.

Le paragraphe suivant nous donne quelques résultats de convergence des estimateurs à noyaux classiques.

### (B) Convergence des estimateurs à noyaux classiques.

Rappelons d'abord les hypothèses par lesquelles ces résultats ont été établis.

$$\begin{cases} \int_{\mathbb{S}} K(u) du = 1, \\ \sup_{u \in \mathbb{S}} K(u) < +\infty, \end{cases} \quad (1.1.10)$$

$$\lim_{n \rightarrow +\infty} h_n = 0. \quad (1.1.11)$$

$$\lim_{n \rightarrow +\infty} nh_n = +\infty. \quad (1.1.12)$$

$$\lim_{n \rightarrow +\infty} nh_n^2 = +\infty. \quad (1.1.13)$$

$$\lim_{n \rightarrow +\infty} \frac{nh_n}{\ln(n)} = +\infty. \quad (1.1.14)$$

(i) **Convergence en moyenne quadratique ponctuelle et intégrée.**

Ces résultats sont établis dans les travaux respectivement de Parzen [76] et Tiago de Oliveira [114]. Le premier montre la convergence de  $MSE$  et le second donne celle de  $MISE$ .

**Théorème 1.1.1.** ([76])

Soit  $f$  une densité continue sur  $\mathbb{T} = \mathbb{R}$  et  $\widehat{f}_n$  son estimateur à noyau  $K$  vérifiant (1.1.10). Si le paramètre  $h_n$  satisfait (1.1.11) et (1.1.12) alors

$$MSE \left[ \widehat{f}_n(x) \right] \underset{n \rightarrow +\infty}{\xrightarrow{\mathbb{P}}} 0, \quad \forall x \in \mathbb{R},$$

où  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}}$  désigne la convergence en probabilité.

**Théorème 1.1.2.** (Tiago de Oliveira [114])

Soit  $f$  une densité de puissance  $p^{\text{ème}}$ -intégrable et  $\widehat{f}_n$  son estimateur à noyau  $K$  vérifiant (1.1.10). Si le paramètre  $h_n$  satisfait (1.1.11) et (1.1.12) alors

$$MISE \left( \widehat{f}_n \right)^p \underset{n \rightarrow +\infty}{\xrightarrow{\mathbb{P}}} 0.$$

(ii) **Consistances faible et forte.**

Les trois résultats donnés ici sont issus des travaux de Parzen [76], Nadaraya [73] et Silverman [101]. Les deux premiers montrent la consistance faible tandis que le dernier propose la consistance forte.

**Théorème 1.1.3.** (Parzen [76])

Soit  $f$  la densité à estimer et  $\widehat{f}_n$  son estimateur à noyau  $K$  vérifiant (1.1.10). Si le paramètre  $h_n$  satisfait (1.1.13) et si, de plus, la transformé de Fourier du noyau  $\mathcal{TF}(z) = \int \exp(-izu)K(u)du$  est absolument intégrable, alors

$$\sup_{x \in \mathbb{R}} \left| \widehat{f}_n(x) - f(x) \right| \underset{n \rightarrow +\infty}{\xrightarrow{\mathbb{P}}} 0.$$

**Théorème 1.1.4.** (Nadaraya [73])

Soit  $f$  une densité uniformément continue et  $\widehat{f}_n$  son estimateur à noyau  $K$  à variations bornées. Pour tout  $h_n$  vérifiant (1.1.11) tel que

$$\sum_{k \geq 1} \exp(-\varepsilon n h_n^2) < +\infty, \text{ alors}$$

$$\sup_{x \in \mathbb{R}} \left| \widehat{f}_n(x) - f(x) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

**Théorème 1.1.5.** (Silverman [101])

Soit  $f$  une densité uniformément continue et  $\widehat{f}_n$  son estimateur à noyau  $K$  à variations bornées. Pour tout  $h_n$  vérifiant (1.1.11) et (1.1.14), alors

$$\sup_{x \in \mathbb{R}} \left| \widehat{f}_n(x) - f(x) \right| \xrightarrow[n \rightarrow +\infty]{p.s} 0,$$

où  $\xrightarrow[n \rightarrow +\infty]{p.s}$  désigne la convergence presque-sûre.

(iii) **Normalité asymptotique.**

Ce dernier résultat est tiré des travaux de Parzen (1962). Il montre que l'estimateur à noyau est asymptotiquement normal.

**Théorème 1.1.6.** (Parzen [1962])

Soit  $f$  une densité continue sur  $\mathbb{T} = \mathbb{R}$  et  $\widehat{f}_n$  son estimateur à noyau  $K$  vérifiant (1.1.10). Si  $h_n$  vérifie (1.1.11) et (1.1.12), alors

$$\frac{\widehat{f}_n(x) - \mathbb{E} \left[ \widehat{f}_n(x) \right]}{\sqrt{\text{Var} \left[ \widehat{f}_n(x) \right]}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1), \quad \forall x \in \mathbb{R}$$

où  $\xrightarrow[n \rightarrow +\infty]{\mathcal{L}}$  désigne la convergence en loi et  $\mathcal{N}(0, 1)$  désigne la loi normale centrée réduite.

### 1.1.1.1 Choix de fenêtre

Comme nous l'avons précédemment signalé, pour avoir une bonne estimation par la méthode des noyaux, il faut bien choisir le paramètre de lissage  $h_n$  puisque celui-ci joue un rôle essentiel dans le processus. Les méthodes existantes pour le choix de  $h_n$  peuvent être classées en deux catégories.

La première catégorie est purement théorique. Elle consiste à minimiser l'erreur quadratique moyenne intégrée (*MISE*). En effet, la valeur optimale théorique de  $h_n$  notée  $h_{opt}$  s'obtient en minimisant la *MISE* asymptotique donnée en (1.1.9). Ainsi, pour un échantillon de taille  $n$  donné et pour un noyau  $K$  fixé, cette valeur optimale de  $h_n$  est donnée par

$$h_{opt} = \left\{ \frac{\int_{\mathbb{S}} K^2(u) du}{\sigma_K^4 \int_{\mathbb{T}} (f''(x))^2 dx} \right\}^{1/5} n^{-1/5}. \quad (1.1.15)$$

où  $\sigma_K^4 = \left( \int_{\mathbb{S}} u^2 K(u) du \right)^2$ . On ne peut pas utiliser ce paramètre de lissage optimal  $h_{opt}$  car il dépend encore de la quantité inconnue  $\int_{\mathbb{T}} (f''(x))^2 dx$ .

La deuxième catégorie est celle dite des méthodes pratiques. Elle utilise seulement les observations. Plusieurs auteurs ont travaillé sur ces méthodes parmi lesquels on peut citer Scott et al. [99], Rudemo [90], Stone [113], Bowman [9], Marron [61], Berline et Devroye [4], Park et Marron [75], Sarda et Vieu [92], Cuevas et al. [21] ainsi que Youndjé et al. [126, 127].

On s'intéresse dans la suite à deux méthodes pratiques : la méthode de ré-injection ("Plug-in") et la méthode de validation croisée par les moindres carrés ("Least Squares Cross Validation").

#### (A) Méthode de Plug-in.

Il s'agit ici de trouver un estimateur à noyau de la quantité inconnue  $\int_{\mathbb{T}} (f''(x))^2 dx$  dans l'expression de  $h_{opt}$  donnée en (1.1.15). Cette approche est celle de Sheather et al. [100] appelée communément plug-in à trois étapes.

Elle consiste à remplacer  $\int_{\mathbb{T}}(f''(x))^2dx$  dans (1.1.15) par son estimateur

$$\widehat{R}_{a_n} = \int_{\mathbb{T}}(\widehat{f''}_n(x))^2dx = \frac{1}{n^2a_n^5} \sum_{i,j=1}^n L^{(4)}\left(\frac{X_i - X_j}{a_n}\right),$$

où  $L^{(4)}$  désigne la dérivée quatrième d'un noyau suffisamment lisse  $L$  et  $a_n$  est un nouveau paramètre de lissage appelé paramètre pilote. Cet estimateur est obtenu en remarquant que sous des conditions de régularités suffisantes sur la densité  $f$ , on a

$$\int_{\mathbb{T}}(f''(x))^2dx = \int_{\mathbb{T}}f^{(4)}(x)f(x)dx.$$

Le nouveau paramètre de lissage  $a_n$  minimisant la quantité

$$\mathbb{E} \left[ \left\{ \widehat{R}_{a_n} - \int_{\mathbb{T}}(f''(x))^2dx \right\}^2 \right],$$

s'écrit

$$a_n = \left\{ \frac{2L^{(4)}(0)}{\sigma_L^2 \int_{\mathbb{T}}(f^{(3)}(x))^2dx} \right\}^{1/7} n^{-1/7}.$$

On constate que, une autre fois,  $a_n$  dépend de la quantité inconnue  $\int_{\mathbb{T}}(f^{(3)}(x))^2dx$  qu'on estime de nouveau par

$$\widehat{R}_{b_n} = \int_{\mathbb{T}}(\widehat{f^{(3)}}_n(x))^2dx = -\frac{1}{n^2b_n^7} \sum_{i,j=1}^n L^{(6)}\left(\frac{X_i - X_j}{a_n}\right),$$

avec  $b_n = 0.912\widehat{\lambda}_n n^{-1/9}$ , où  $\widehat{\lambda}_n$  est l'écart inter-quartile de l'échantillon  $X_1, \dots, X_n$ .

## (B) Méthode de la validation croisée par les moindres carrés.

L'idée principale de cette méthode de validation croisée est celle de la minimisation par rapport à  $h_n$  d'un estimateur de la *MISE*. Pour plus de détails, on peut se référer à de nombreux auteurs tels que Bowman [9], Marron [61], Rudemo [90] et Stone [113].

L'approche de validation croisée par les moindres carrés consiste à minimiser un estimateur convenable de la *MISE* qui s'écrit

$$\begin{aligned} MISE(\widehat{f}_n) &= \mathbb{E} \left[ \int_{\mathbb{T}} \left\{ \widehat{f}_n(x) - f(x) \right\}^2 dx \right], \\ &= \mathbb{E} \left[ \int_{\mathbb{T}} \widehat{f}_n^2(x) dx - 2 \int_{\mathbb{T}} \widehat{f}_n(x) f(x) dx \right] + \int_{\mathbb{T}} f^2(x) dx. \end{aligned}$$

Comme la dernière intégrale  $\int_{\mathbb{T}} f^2(x)dx$  ne dépend pas de  $h_n$ , la valeur  $h_{opt}$  qui minimise la *MISE* est aussi celle qui minimise la fonction

$$J(h_n) = \mathbb{E} \left[ \int_{\mathbb{T}} \widehat{f}_n^2(x)dx - 2 \int_{\mathbb{T}} \widehat{f}_n^2(x)f(x)dx \right].$$

Pour trouver  $h_{opt}$ , il suffit d'estimer sans biais  $J(h_n)$  par :

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

avec

$$\widehat{f}_{n,-i}(x) = \frac{1}{(n-1)h_n} \sum_{j \neq i} K \left( \frac{x - X_j}{h_n} \right), \quad (1.1.16)$$

est l'estimateur calculé à partir de l'échantillon privé de l'observation  $X_i$ . La valeur optimale de  $h_n$  est donnée par :

$$h_{opt} = \arg \min_{h_n > 0} CV(h_n).$$

### 1.1.2 Estimation d'une fonction de régression

Le modèle de la régression est l'un des modèles les plus fréquemment rencontrés en statistique paramétrique et non paramétrique.

On considère  $(X_1, Y_1), \dots, (X_n, Y_n)$  des couples de variables aléatoires indépendantes réelles et de même loi que  $(X, Y)$ . Dans le modèle de régression non paramétrique on suppose l'existence d'une fonction  $r$  qui exprime la valeur moyenne de la variable à expliquer  $Y$  en fonction de la variable explicative  $X$ , c'est à dire :

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

où  $\varepsilon$  est une variable centrée ( $\mathbb{E}(\varepsilon) = 0$ ) et indépendante de  $X$ .

On suppose que la loi du couple  $(X, Y)$  admet une densité conjointe  $g(x, y)$ , la variable aléatoire  $X$  admet une densité marginale  $f$  et la variable aléatoire  $Y$  est intégrable ( $\mathbb{E}(|Y|) < +\infty$ ). Alors, la fonction de régression  $r$ , de support  $\mathbb{T}$ , est définie par :

$$\forall x \in \mathbb{T} \text{ tel que } f(x) \neq 0, \quad r(x) = \mathbb{E}(Y|X = x) = \frac{\int_{\mathbb{R}} yg(x, y)dy}{f(x)}.$$

La fonction  $r$  réalise la meilleure approximation de  $Y$  sachant  $X = x$  au sens des moindres carrés.

On donne l'estimateur de  $r$  introduit par Nadaraya [72] et [120].

**Définition 1.1.3.**

Soit la fenêtre de lissage  $(h_n)$  et  $K$  la fonction noyau vérifiant la Définition 1.1.1.

L'estimateur à noyau de  $r$  est défini en un point  $x \in \mathbb{T}$  par :

$$\widehat{r}_n^{NW}(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \neq 0 \\ \frac{1}{n} \sum_{i=1}^n Y_i & \text{sinon.} \end{cases} \quad (1.1.17)$$

**1.1.2.1 Caractéristiques**

D'une manière analogue à l'estimateur de Rosenblatt  $\widehat{f}_n$ , nous étudions dans cette partie les propriétés de  $\widehat{r}_n^{NW}$ . Pour étudier le biais et la variance de l'estimateur  $\widehat{r}_n^{NW}$ , on utilise l'expression

$$\widehat{r}_n^{NW}(x) = \frac{\phi_n(x)}{\widehat{f}_n(x)}, \quad \text{où } \phi_n(x) = \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right).$$

**(A) Etude asymptotique de la variance.**

On suppose que  $\mathbb{E}(Y^2) < +\infty$ , alors en chaque point de continuité des fonctions  $r$ ,  $f$  et  $\sigma^2 : x \mapsto \text{Var}(Y|X = x)$ ,  $x \in \mathbb{T}$  on a

$$\text{Var} [\widehat{r}_n^{NW}(x)] = \frac{1}{nh_n} \frac{\sigma^2(x)}{f(x)} \int_{\mathbb{T}} K^2(u) du + o\left(\frac{1}{nh_n}\right). \quad (1.1.18)$$

En effet en posant  $\psi(x) = \int_{\mathbb{R}} y^2 g(x, y) dy$ , on a

$$\begin{aligned}
\text{Var} [\phi_n(x)] &= \frac{1}{nh_n^2} \left\{ \mathbb{E} \left[ Y^2 K^2 \left( \frac{x-X}{h_n} \right) \right] - \left( \mathbb{E} \left[ Y K \left( \frac{x-X}{h_n} \right) \right] \right)^2 \right\}, \\
&= \frac{1}{nh_n} \int_{\mathbb{T}} K^2(u) \psi(x - h_n u) du \\
&\quad - \frac{1}{n} \left( \int_{\mathbb{T}} K(u) f(x - uh_n) r(x - uh_n) du \right)^2, \\
&= \frac{1}{nh_n} \psi(x) \int_{\mathbb{T}} K^2(u) du + o \left( \frac{1}{nh_n} \right).
\end{aligned}$$

D'autre part, on a

$$\begin{aligned}
\text{Cov} \left( \widehat{f}_n(x), \phi_n(x) \right) &= \mathbb{E} \left\{ \left( \widehat{f}_n(x) - \mathbb{E} \left[ \widehat{f}_n(x) \right] \right) (\phi_n(x) - \mathbb{E} [\phi_n(x)]) \right\}, \\
&= \mathbb{E} \left[ \widehat{f}_n(x) \phi_n(x) \right] - \mathbb{E} \left[ \widehat{f}_n(x) \right] \mathbb{E} [\phi_n(x)], \\
&= \frac{1}{nh_n^2} \mathbb{E} \left[ Y K^2 \left( \frac{x-X}{h_n} \right) \right] - \frac{n(n-1)}{n^2 h_n^2} \mathbb{E}^2 \left[ Y K \left( \frac{x-X}{h_n} \right) \right] \\
&\quad - \mathbb{E} \left[ \widehat{f}_n(x) \right] \mathbb{E} [\phi_n(x)], \\
&= \frac{1}{nh_n^2} \mathbb{E} \left[ Y K^2 \left( \frac{x-X}{h_n} \right) \right] - \frac{1}{n} \mathbb{E} \left[ \widehat{f}_n(x) \right] \mathbb{E} [\phi_n(x)], \\
&= \frac{1}{nh_n} \int_{\mathbb{T}} K^2(u) f(x - uh_n) r(x - uh_n) du - \frac{1}{n} \mathbb{E} \left[ \widehat{f}_n(x) \right] \mathbb{E} [\phi_n(x)], \\
&= \frac{1}{nh_n} r(x) f(x) \int_{\mathbb{T}} K^2(u) du + o \left( \frac{1}{nh_n} \right).
\end{aligned}$$

On rappelle que

$$\text{Var} \left[ \widehat{f}_n(x) \right] = \frac{1}{nh_n} f(x) \int_{\mathbb{T}} K^2(u) du + o \left( \frac{1}{nh_n} \right).$$

Posons

$$\begin{aligned}
B_n(x) &= \left( \widehat{f}_n(x), \phi_n(x) \right), \\
A(x) &= \left( \frac{-r(x)}{\{f(x)\}^2}, \frac{1}{f(x)} \right).
\end{aligned}$$

La matrice de covariance de  $B_n(x)$  est alors donnée par l'expression suivante

$$\Sigma = \frac{1}{nh_n} \begin{pmatrix} f(x) & r(x) \\ r(x) & \psi(x) \end{pmatrix} \int_{\mathbb{T}} K^2(u) du + o \left( \frac{1}{nh_n} \right).$$

En remarquant, que

$$\begin{aligned}
\text{Var} \left[ \widehat{r}_n^{NW}(x) \right] &= A \Sigma A^t, \\
&= \frac{1}{nh_n} \left[ \frac{\psi(x)}{\{f(x)\}^2} - \frac{r^2(x)}{\{f(x)\}^3} \right] \int_{\mathbb{T}} K^2(u) du + o \left( \frac{1}{nh_n} \right),
\end{aligned}$$



où  $A^t$  désigne la transposé de  $A$ . On obtient alors la variance de  $\widehat{r}_n^{NW}$  donnée par l'équation (1.1.18). Pour plus détails, nous pouvons citer par exemple les articles de Collomb [19, 18].

### (B) Etude asymptotique du biais.

Le traitement du biais est purement analytique et repose essentiellement sur un développement de Taylor. Il nous faut supposer certaines conditions de régularité sur les fonctions  $r$  et  $f$  qui détermineront l'ordre du biais asymptotique en fonction du paramètre de lissage ( $h_n$ ). L'estimateur de Nadaraya Watson se présente sous la forme d'un quotient aléatoire, c'est pourquoi on utilise généralement comme terme de centrage l'approximation suivante

$$\widetilde{\mathbb{E}} [\widehat{r}_n^{NW}(x)] = \frac{\mathbb{E} [\phi_n(x)]}{\mathbb{E} [\widehat{f}_n(x)]}.$$

Cette formule est plus facile à manipuler et permet notamment la linéarisation de la déviation

$$\widetilde{d}_n(x) = \widehat{r}_n^{NW}(x) - \widetilde{\mathbb{E}} [\widehat{r}_n^{NW}(x)].$$

Afin de justifier le choix du terme de centrage, Nadaraya [74] a montré :

(i) si la variable  $Y$  est bornée et  $nh_n \rightarrow +\infty$ , alors

$$\mathbb{E} [\widehat{r}_n^{NW}(x)] = \widetilde{\mathbb{E}} [\widehat{r}_n^{NW}(x)] + O\left(\frac{1}{nh_n}\right).$$

(ii) si  $\mathbb{E}(Y^2) < +\infty$  et  $nh_n^2 \rightarrow +\infty$ , alors

$$\mathbb{E} [\widehat{r}_n^{NW}(x)] = \widetilde{\mathbb{E}} [\widehat{r}_n^{NW}(x)] + O\left(\frac{1}{\sqrt{nh_n}}\right).$$

D'autre part, en supposant que  $f$  et  $r$  sont bornées, deux fois différentiables, et  $f''$  et  $r''$  sont bornées ; on a

$$\begin{aligned} \mathbb{E} [\phi_n(x)] &= \frac{1}{h_n} \int_{\mathbb{R}} \int_{\mathbb{T}} y K\left(\frac{x-z}{h_n}\right) dy dz, \\ &= \int_{\mathbb{T}} \left[ \int_{\mathbb{R}} y f(x - uh_n) dy \right] K(u) du, \\ &= \int_{\mathbb{T}} r(x - uh_n) f(x - uh_n) K(u) du. \end{aligned}$$

Le développement en série de Taylor de  $r(x - uh_n)f(x - uh_n)$  au voisinage de  $x$  est

$$\begin{aligned} r(x - uh_n)f(x - uh_n) &= \left[ r(x) - uh_n r'(x) + \frac{(uh_n)^2}{2} r''(x) + o_x((uh_n)^2) \right] \\ &\times \left[ f(x) - uh_n f'(x) + \frac{(uh_n)^2}{2} f''(x) + o_x((uh_n)^2) \right], \\ &= r(x)f(x) - h_n u (r'(x)f(x) + r(x)f'(x)) \\ &+ \frac{h_n^2}{2} u^2 (r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)) + o_x(u^2 h_n^2). \end{aligned}$$

Ainsi, on a

$$\mathbb{E}[\phi_n(x)] = r(x)f(x) + \frac{h_n^2}{2} \{r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)\} \sigma_K^2 + o(h_n^2).$$

où  $\sigma_K^2 = \int_{\mathbb{S}} u^2 K(u) du$ .

On rappelle de plus que l'espérance de  $\widehat{f}_n(x)$  est

$$\mathbb{E}[\widehat{f}_n(x)] = f(x) + \frac{h_n^2}{2} \sigma_K^2 + o(h_n^2).$$

Ainsi, en ajoutant aux conditions utilisées pour le calcul du biais de  $\phi_n$  le fait que la variable  $Y$  est bornée et que  $h_n$  vérifie (1.1.11) et (1.1.12) le biais asymptotique de l'estimateur de Nadaraya-Watson est donné par :

$$\mathbb{E}[\widehat{r}_n^{NW}(x)] - r(x) = \frac{h_n^2}{2} \left\{ r''(x) + 2 \frac{r'(x)f'(x)}{f(x)} \right\} \sigma_K^2 + o(h_n^2). \quad (1.1.19)$$

Lorsque la fonction de régression admet des conditions de régularité supplémentaires, il est possible de réduire le biais asymptotique de l'estimateur Nadaraya-Watson en utilisant un noyau d'ordre supérieur.

En effet, pour  $q$  un entier naturel fixé et en supposant que  $r$  et  $f$  soient de classe  $C^q(\mathbb{T})$  et que le noyau  $K$  vérifiant

$$\int_{\mathbb{T}} K(u) du = 1, \quad \int_{\mathbb{T}} u^j K(u) du = 0, \quad j \in \{1, \dots, q-1\} \quad \text{et} \quad \int_{\mathbb{T}} u^q K(u) < +\infty,$$

le biais asymptotique de l'estimateur  $\widehat{r}_n^{NW}$ , lorsque  $h_n \xrightarrow[n \rightarrow +\infty]{} 0$  et  $nh_n \xrightarrow[n \rightarrow +\infty]{} +\infty$ , est donné par :

$$\mathbb{E}[\widehat{r}_n^{NW}(x)] - r(x) = \frac{h_n^q}{q!} \left\{ r^{(q)}(x) + 2r^{(q-1)}(x) \frac{f'(x)}{f(x)} \right\} \int_{\mathbb{T}} u^q K(u) du + o(h_n^q).$$

Pour plus détails, on peut citer par exemple les articles de Müller [71], Choi et al. [16] et Hall et Müller [36].

(C) **Erreur Quadratique Moyenne ponctuelle et intégrée.**

Des deux expressions (1.1.18) et (1.1.19) respectivement de la variance et du biais de l'estimateur  $\widehat{r}_n^{NW}$ , on peut déduire les formes approximées et asymptotiques de la  $MSE$  et de la  $MISE$  par :

$$\begin{aligned} AMSE [\widehat{r}_n^{NW}(x)] &= \frac{h_n^4}{4} \left\{ r''(x) + 2 \frac{r'(x)f'(x)}{f(x)} \right\}^2 \sigma_K^4 \\ &+ \frac{1}{nh_n} \frac{\sigma^2(x)}{f(x)} \int_{\mathbb{T}} K^2(u) du, \end{aligned} \quad (1.1.20)$$

et

$$\begin{aligned} AMISE(\widehat{r}_n^{NW}) &= \frac{h_n^4}{4} \sigma_K^4 \int_{\mathbb{T}} \left\{ r''(x) + 2 \frac{r'(x)f'(x)}{f(x)} \right\}^2 dx \\ &+ \frac{1}{nh_n} \int_{\mathbb{T}} \frac{\sigma^2(x)}{f(x)} dx \int_{\mathbb{T}} K^2(u) du. \end{aligned} \quad (1.1.21)$$

(D) **Normalité asymptotique.**

Schuster [97] était le premier à établir la normalité asymptotique de l'estimateur  $\widehat{r}_n^{NW}$ . On se réfère également aux Théorèmes 1.3 et 1.4 de [74, p. 117-120] et au Théorème 4.2.1 de Härdle [39, p. 99], qui reposent sur d'autres méthodes de démonstration.

Le noyau  $K$  est supposé borné, à support compact et de classe  $C^2(\mathbb{R})$ .

**Théorème 1.1.7.** (Härdle [39])

Soit  $(X, Y)$  un couple de variables aléatoires réelles de densité conjointe  $g(x, y)$ . On suppose que la variable aléatoire  $X$  admet une densité marginale  $f$  et la variable aléatoire  $Y$  est intégrable ( $\mathbb{E}(|Y|) < +\infty$ ) et possède un moment d'ordre  $l > 2$ . Alors, la fonction de régression  $r$ , de support  $\mathbb{T}$ , est définie par :

$$\forall x \in \mathbb{T} \text{ tel que } f(x) \neq 0, \quad r(x) = \mathbb{E}(Y|X = x) = \frac{\int_{\mathbb{R}} yg(x, y)dy}{f(x)}.$$

Les fonctions  $r : x \mapsto \mathbb{E}(Y|X = x)$  et  $f$  sont supposées deux fois continûment dérivables sur  $\mathbb{T}$ . Pour tout point de continuité  $x \in \mathbb{T}$  de  $\sigma^2(x) = \text{Var}(Y|X = x)$ , tel que  $f(x) > 0$ , on a

$$\sqrt{nh_n} [\widehat{r}_n^{NW}(x) - r(x)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(B(x), v^2(x)),$$

avec  $B(x) = \left\{ r''(x) + 2 \frac{r'(x)f'(x)}{f(x)} \right\} \int_{\mathbb{S}} u^2 K(u) du$  et  $v^2(x) = \frac{\sigma^2(x)}{f(x)} \int_{\mathbb{T}} K^2(u) du$ .

### 1.1.2.2 Choix de fenêtre

Comme on a fait pour l'estimateur de Rosenblatt de la densité  $\widehat{f}_n$ , le choix théorique du paramètre de lissage  $h_n$  consiste à minimiser l'erreur quadratique moyenne intégrée (*MISE*). En effet, la valeur optimale théorique de  $h_n$  notée  $h_{opt}$  s'obtient en minimisant la *MISE* asymptotique donnée en (1.1.21). Ainsi, pour un échantillon de taille  $n$  donné et pour un noyau  $K$  fixé, cette valeur optimale de  $h_n$  est égale à

$$h_{opt} = \left\{ \frac{\int_{\mathbb{T}} \frac{\sigma^2(x)}{f(x)} dx \int_{\mathbb{S}} K^2(u) du}{\sigma_K^4 \int_{\mathbb{T}} \left[ r''(x) + 2 \frac{r'(x)f'(x)}{f(x)} \right] dx} \right\}^{1/5} n^{-1/5},$$

où  $\sigma_K^4 = \left( \int_{\mathbb{S}} u^2 K(u) du \right)^2$ .

De nouveau, la fenêtre optimale dépend de paramètres inconnus et n'est donc pas utilisable en pratique. On se propose de remédier à cet obstacle via la méthode de validation croisée (Härdle et Marron [37] et Härdle [39]) qui consiste à minimiser par rapport à  $h_n$  la fonction

$$CV(h_n) = \frac{1}{n} \sum_{i=1}^n [Y_i - \widehat{r}_{n,-i}^{NW}(X_i)]^2 w(X_i),$$

avec  $w$  est une fonction de poids (le plus naturel est de prendre  $w(x) = f(x)$  voir Härdle et Kelly [38]) et  $\widehat{r}_{n,-i}^{NW}$  est l'estimateur dénommé "leave-one-out", donné par

$$\widehat{r}_{n,-i}^{NW} = \frac{\sum_{j \neq i}^n Y_j K\left(\frac{x - X_j}{h_n}\right)}{\sum_{j \neq i}^n K\left(\frac{x - X_j}{h_n}\right)}.$$

Cet estimateur est simplement l'estimateur de Nadaraya-Watson construit avec les  $(n-1)$  couples aléatoires  $(X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), \dots, (X_n, Y_n)$ .

Les références principales à ce sujet sont Hall [35], Härdle et Marron [37], Härdle et Kelly [38], concernant l'estimation non-paramétrique de la fonction de régression.

La procédure de validation croisée peut s'interpréter comme étant le meilleur choix de  $h_n$ . Pour la méthode de plug-in, on cite les articles de Mack et Müller [60] ainsi que Müller et Prewitt [69], qui ont construits des estimateurs de type Nadaraya-Watson avec une fenêtre asymptotiquement optimale.

Dans le paragraphe suivant, on parlera des estimateurs récursifs construits en utilisant l'algorithme stochastique de [87].

## 1.2 Algorithmes stochastiques et estimateurs récursifs

L'utilisation la plus célèbre des algorithmes stochastiques dans le cadre des statistiques non paramétriques est le travail de Kiefer et Wolfowitz [52].

Ces deux auteurs ont construit un algorithme qui permet l'approximation du mode d'une fonction de régression.

Cet algorithme à été beaucoup discuté et leur travail prolongé dans plusieurs directions (citons, parmi beaucoup d'autres, Blum [6], Fabian [29], Kushner et Clark [54], Hall et Heyde [34], Ruppert [91], Chen [11], Spall [110], Polyak et Tsybakov [81], Dippon et Renz [25], Spall [111], Chen et al. [13], Dippon [26], et Mokkaïem et Pelletier [64]).

Des algorithmes stochastiques d'approximation ont été également présentés par Révész [85, 86] pour estimer une fonction de régression en un point donné, et par Tsybakov [117] pour approximer le mode d'une densité de probabilité.

On introduit dans le paragraphe suivant, les estimateurs récursifs construits en utilisant l'algorithme de Robbins et Monro [87].

### 1.2.1 Estimateurs récurrents d'une densité

Les méthodes d'approximation stochastiques permettent la recherche du zéro  $\theta^*$  d'une fonction inconnue  $h : \mathbb{R} \rightarrow \mathbb{R}$  qui peut être difficilement calculé directement. L'algorithme le plus connu est celui de Robbins et Monro [87]. On procède de la façon suivante :

- (i) On choisit  $\theta_0 \in \mathbb{R}$  ;
- (ii) Pour  $n \geq 1$ , on construit la suite  $(\theta_n)$  par la relation récurrente

$$\theta_n = \theta_{n-1} + \gamma_n W_n,$$

où  $W_n$  est une observation de la fonction  $h$  au point  $\theta_{n-1}$  et  $(\gamma_n)$  est une suite de réels positifs qui tend vers zéro appelée pas de l'algorithme.

Soit  $(X_1, \dots, X_n)$  un échantillon de densité de probabilité  $f$ . Pour construire un estimateur récurrent de  $f$  en un point  $x$  par la méthode des algorithmes stochastiques, Mokkadem et al. [66] ont défini un algorithme de recherche du zéro de la fonction  $h : y \mapsto f(x) - y$ . La procédure est la suivante :

- (i) Soit  $f_0(x) \in \mathbb{R}$  fixé ;
- (ii) pour  $n \geq 1$ , on a

$$f_n(x) = f_{n-1}(x) + \gamma_n W_n(x),$$

où  $W_n(x)$  est une observation de la fonction  $h$  au point  $f_{n-1}(x)$ . On peut estimer  $f(x)$  par  $Z_n(x) = h_n^{-1} K\left(\frac{x - X_n}{h_n}\right)$ , ce qui permet de poser  $W_n(x) = Z_n(x) - f_{n-1}(x)$ . Ainsi, l'estimateur de la densité  $f$  en un point  $x$  s'écrit récursivement sous la forme :

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right). \quad (1.2.1)$$

Notons que si on pose  $(\gamma_n) = (n^{-1})$ , alors l'estimateur  $f_n$  donné par l'algorithme (1.2.1) devient

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{x - X_i}{h_i}\right),$$

dans ce cas,  $f_n$  est l'estimateur récursif introduit par Wolverton et Wagner [124]. D'autre part, dans le cas où  $(\gamma_n) = \left( h_n \left[ \sum_{i=1}^n h_i \right]^{-1} \right)$ , l'estimateur  $f_n$  donné par l'algorithme (1.2.1) s'écrit alors

$$f_n(x) = \frac{1}{\sum_{i=1}^n h_i} \sum_{i=1}^n \frac{1}{h_i} K \left( \frac{x - X_i}{h_i} \right).$$

C'est l'estimateur récursif introduit par Deheuvels [23] et étudié par Dufflo [27].

Pour donner les caractéristiques de l'estimateur  $f_n$ , on introduit une classe de suites à variations régulières que nous utiliserons dans nos hypothèses tout au long des chapitres suivants.

**Définition 1.2.1.**

Soit  $\gamma \in \mathbb{R}$  et  $(v_n)_{n \geq 1}$  une suite réelle positive. On dit que  $(v_n) \in \mathcal{GS}(\gamma)$  si

$$\lim_{n \rightarrow +\infty} n \left[ 1 - \frac{v_{n-1}}{v_n} \right] = \gamma. \quad (1.2.2)$$

La condition (1.2.2) a été introduite par Galambos et Seneta [30] pour définir les suites à variations régulières. Des exemples classiques de suites dans  $\mathcal{GS}(\gamma)$  sont, pour  $c > 0$  et  $b \in \mathbb{R}$ ,  $(v_n) = (cn^\gamma)$ ,  $(v_n) = (cn^\gamma [\ln(n)]^b)$ ,  $(v_n) = (cn^\gamma [\ln \ln(n)]^b)$ , etc.

On donne d'abord le biais et la variance de l'estimateur  $f_n$ , puis l'erreur quadratique intégrée (*MISE*) et le choix théorique du pas et fenêtre qui la minimise. Enfin, on rappelle la vitesse de convergence ponctuelle et uniforme de l'estimateur  $f_n$ .

### 1.2.1.1 Biais et variance de l'estimateur $f_n$

On se place sous les hypothèses suivantes :

(H1)  $K : \mathbb{R} \rightarrow \mathbb{R}$  est la fonction noyau continue, bornée et vérifiant les propriétés de la définition 1.1.1.

(H2) i)  $(\gamma_n) \in \mathcal{GS}(-\alpha)$  avec  $\alpha \in ]\frac{1}{2}, 1]$ .

ii)  $(h_n) \in \mathcal{GS}(-a)$  avec  $a \in ]0, \frac{\alpha}{2}]$ .

iii)  $\lim_{n \rightarrow +\infty} (n\gamma_n) \in ]\min \left\{ 2a, \frac{1-a}{2} \right\}, +\infty]$ .

(H3)  $f$  est bornée, deux fois différentiables et  $f''$  est bornée.

L'hypothèse (H2) iii) implique que la limite de  $(n\gamma_n)^{-1}$  est finie. On note alors  $\xi = \lim_{n \rightarrow +\infty} (n\gamma_n)^{-1}$ .

**Proposition 1.2.1.** (*Mokkadem et al. [66]*)

*Supposons que les hypothèses (H1)-(H3) soient vérifiées et que  $f''$  soit continue au point  $x$ .*

1. Si  $0 < a \leq \frac{\alpha}{5}$ , alors

$$\mathbb{E}[f_n(x)] - f(x) = \frac{1}{2(1-2a\xi)} h_n^2 \sigma_K^2 f''(x) + o(h_n^2),$$

si  $a > \frac{\alpha}{5}$ , alors

$$\mathbb{E}[f_n(x)] - f(x) = o\left(\sqrt{\gamma_n h_n^{-1}}\right).$$

2. Si  $a \geq \frac{\alpha}{5}$ , alors

$$\text{Var}[f_n(x)] = \frac{1}{2 - (1-a)\xi} \frac{\gamma_n}{h_n} f(x) \int_{\mathbb{R}} K^2(u) du + o\left(\frac{\gamma_n}{h_n}\right),$$

si  $0 < a \leq \frac{\alpha}{5}$ , alors

$$\text{Var}[f_n(x)] = o(h_n^4).$$



### 1.2.1.2 Erreur Quadratique Moyenne Intégrée de $f_n$

Dans la proposition suivante, on donne la *MISE* (Erreur quadratique moyenne intégrée) de  $f_n$ .

**Proposition 1.2.2.** (*Mokkadem et al. [66]*)

Supposons que les hypothèses (H1)-(H3) soient vérifiées, et que  $f''$  soit continue sur  $\mathbb{R}$ .

1. Si  $0 < a < \frac{\alpha}{5}$ , alors

$$MISE(f_n) = \frac{1}{2(1-2a\xi)} h_n^4 \sigma_K^4 \int_{\mathbb{R}} f''(x) dx + o(h_n^4).$$

2. Si  $a = \frac{\alpha}{5}$ , alors

$$\begin{aligned} MISE(f_n) &= \frac{1}{2(1-2a\xi)} h_n^4 \sigma_K^4 \int_{\mathbb{R}} f''(x) dx \\ &+ \frac{1}{2-(1-a)\xi} \frac{\gamma_n}{h_n} \int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} K^2(u) du + o\left(h_n^4 + \frac{\gamma_n}{h_n}\right). \end{aligned}$$

3. Si  $a > \frac{\alpha}{5}$ , alors

$$MISE(f_n) = \frac{1}{2-(1-a)\xi} \frac{\gamma_n}{h_n} \int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} K^2(u) du + o\left(\frac{\gamma_n}{h_n}\right).$$

De cette proposition, on déduit le choix optimal du pas qui permet de minimiser la *MISE* de  $f_n$ .

**Corollaire 1.2.1.** (*Mokkadem et al. [66]*)

Supposons que les hypothèses (H1)-(H3) soient vérifiées et que  $f''$  soit continue sur  $\mathbb{R}$ . Pour minimiser la *MISE* de  $f_n$ , le pas  $(\gamma_n)$  doit être choisi dans  $\mathcal{GS}(-1)$  et tel que  $\lim_{n \rightarrow +\infty} n\gamma_n = 1$  et la fenêtre  $(h_n)$  doit être égale

$$\left( \left[ \frac{3}{10} \frac{\int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} K^2(u) du}{\sigma_K^4 \int_{\mathbb{R}} (f''(x))^2 dx} \right]^{1/5} \gamma_n^{1/5} \right).$$

Dans ce cas, la *MISE* de  $f_n$  est égale à

$$MISE(f_n) = \frac{5^{11/5}}{4^{7/5} 3^{6/5}} \left[ \sigma_K^2 \int_{\mathbb{R}} (f''(x))^2 dx \right]^{2/5} \left[ \int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} K^2(u) du \right]^{4/5} n^{-4/5} + o(n^{-4/5}).$$

Un exemple classique de pas  $(\gamma_n)$  appartenant à  $\mathcal{GS}(-1)$  et vérifiant  $\lim_{n \rightarrow +\infty} n\gamma_n = 1$  est  $(\gamma_n) = (n^{-1})$ . Pour ce choix de pas, l'estimateur  $f_n$  est égal à l'estimateur récursif introduit par Wolverton et Wagner [124], puis étudié entre autres par Yamato [125], Davies [22], Devroye [24], Wegman et Davies [122], Menon et al. [63], Wertz [123], Roussas [89] et Duflo [27].

La *MISE* des estimateurs à noyau récursif est plus grande que celle de l'estimateur à noyau non-récursif de Rosenblatt alors il vaut mieux utiliser l'estimateur non récursif.

Comme dans le cas non récursif, on doit utiliser des méthodes pratiques pour choisir la fenêtre  $h_n$ . Dans ce cadre, le lecteur peut voir le travail de Slaoui [104] qui a utilisé la méthode de plug-in pour l'estimation du paramètre de lissage  $(h_n)$ .

### 1.2.1.3 Vitesse de Convergence de $f_n$

On donne dans le Théorème suivant la vitesse de convergence faible de l'estimateur  $f_n$  défini par l'algorithme (1.2.1).

**Théorème 1.2.1.** *(Mokkadem et al. [66])*

*Supposons que les hypothèses (H1)-(H3) soient vérifiées. Pour  $x \in \mathbb{R}$  tel que  $f(x) > 0$  et  $f''$  soit continue au point  $x$ , on a*

1. *S'il existe  $c \geq 0$  tel que  $\gamma_n^{-1}h_n^5 \xrightarrow[n \rightarrow +\infty]{} c$ , alors*

$$\sqrt{\gamma_n^{-1}h_n} [f_n(x) - f(x)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left( \frac{c^{1/2}}{2(1-2a\xi)} \sigma_K^2 f''(x), \frac{1}{2-(1-a)\xi} \int_{\mathbb{R}} K^2(u) du f(x) \right).$$

2. *Si  $\gamma_n^{-1}h_n^5 \xrightarrow[n \rightarrow +\infty]{} +\infty$ , alors*

$$\frac{1}{h_n^2} [f_n(x) - f(x)] \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{1}{2(1-2a\xi)} \sigma_K^2 f''(x).$$

On donne maintenant la vitesse de convergence presque sûre ponctuelle de  $f_n$ , ainsi qu'une majoration de sa vitesse de convergence presque sûre uniforme.

**Théorème 1.2.2.** (Mokkadem et al. [66])

Supposons que les hypothèses (H1)-(H3) soient vérifiées, et que  $f''$  soit continue au point  $x$ . On pose  $s_n = \sum_{k=1}^n \gamma_k$ .

1. S'il existe  $c_1 \geq 0$  tel que  $\gamma_n^{-1} h_n^5 / (\ln s_n) \rightarrow c_1$ , alors, avec probabilité un, la suite

$$\left( \sqrt{\frac{\gamma_n^{-1} h_n}{2 \ln s_n}} [f_n(x) - f(x)] \right)$$

est relativement compacte et l'ensemble de ses valeurs d'adhérence est l'intervalle

$$\left[ \sqrt{\frac{c_1}{2}} \frac{1}{2(1-2a\xi)} \sigma_K^2 f''(x) - \sqrt{\frac{f(x)}{2 - (1-a)\xi}} \int_{\mathbb{R}} K^2(u) du, \right. \\ \left. \sqrt{\frac{c_1}{2}} \frac{1}{2(1-2a\xi)} \sigma_K^2 f''(x) + \sqrt{\frac{f(x)}{2 - (1-a)\xi}} \int_{\mathbb{R}} K^2(u) du \right].$$

2. Si  $\gamma_n^{-1} h_n^5 / (\ln s_n) \xrightarrow[n \rightarrow +\infty]{} +\infty$ , alors

$$\frac{1}{h_n^2} [f_n(x) - f(x)] \xrightarrow[n \rightarrow +\infty]{p.s.} 0 \frac{1}{2(1-2a\xi)} \sigma_K^2 f''(x),$$

$$\text{avec } \sigma_K^2 = \int_{\mathbb{S}} u^2 K(u) du.$$

Pour établir une majoration de la vitesse de convergence uniforme, on rajoute l'hypothèse suivante.

(H4)  $K$  est une fonction lipschitzienne.

**Théorème 1.2.3.** (Mokkadem et al. [66])

Soit  $I$  un intervalle borné de  $\mathbb{R}$ . Supposons que les hypothèses (H1)-(H4) soient vérifiées, et que  $f''$  soit uniformément continue sur  $I$ .

1. S'il existe  $c \geq 0$  tel que  $\gamma_n^{-1} h_n^5 / (\ln n)^2 \xrightarrow[n \rightarrow +\infty]{} c$ , alors

$$\sup_{x \in I} |f_n(x) - f(x)| = O\left(\sqrt{\gamma_n h_n^{-1} \ln n}\right) p.s.$$

2. Si  $\gamma_n^{-1} h_n^5 / (\ln n)^2 \xrightarrow[n \rightarrow +\infty]{} +\infty$ , alors

$$\sup_{x \in I} |f_n(x) - f(x)| = O(h_n^2) p.s.$$

## 1.2.2 Estimateurs récursifs d'une fonction de régression

Soient  $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$  des variables aléatoires i.i.d à valeurs dans  $\mathbb{R}^2$ . On note  $g(x, y)$  la densité du couple  $(X, Y)$ ,  $f(x)$  la densité marginale de  $X$  et  $r(x) = \mathbb{E}[Y|X = x]$  la régression de  $Y$  sur  $X$ . Révész [85] a utilisé les méthodes d'approximation stochastiques pour construire un estimateur à noyau récursif de  $r$ . Il a défini un algorithme de recherche du zéro de la fonction

$h : y \rightarrow f(x)r(x) - f(x)y$  de la façon suivante

- (i) Il a fixé  $r_0(x) \in \mathbb{R}$ ;
- (ii) Pour tout  $n \geq 1$ , il a posé

$$r_n(x) = r_{n-1}(x) + \frac{1}{n}W_n(x),$$

où  $W_n(x)$  est une observation de la fonction  $h$  au point  $r_{n-1}(x)$ .

En posant  $h_n^{-1}Y_nK\left(\frac{x-X_n}{h_n}\right)$  un estimateur de  $f(x)r(x)$  et  $h_n^{-1}K\left(\frac{x-X_n}{h_n}\right)$  un estimateur de  $f(x)$ , il a obtenu

$$r_n(x) = \left(1 - \frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right)\right)r_{n-1}(x) + \frac{1}{nh_n}Y_nK\left(\frac{x-X_n}{h_n}\right).$$

Révész [86] a étudié les propriétés asymptotiques de cet estimateur lorsque la fenêtre  $(h_n)$  est choisie égale à  $(n^{-a})$  avec  $a \in ]\frac{1}{2}, 1[$ .

Sous la condition  $f(x) > \frac{1-a}{2}$ , il a montré que

$$\sqrt{nh_n}[r_n(x) - r(x)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{f(x)\text{Var}[Y|X=x]}{2f(x) - (1-a)}\right).$$

De plus, pour  $I$  un intervalle borné de  $\mathbb{R}$ . Révész [86] a établi que, sous la condition

$$\inf_{x \in I} f(x) > \frac{1-a}{2},$$

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{nh_n}}{(\ln n)^2} \sup_{x \in I} |r_n(x) - r(x)| = 0 \quad p.s.$$

L'estimateur proposé par Révész a deux gros inconvénients. Le premier est que la vitesse de convergence de  $r_n$  est plus petite que  $n^{1/4}$ , alors que la vitesse optimale de l'estimateur de Nadaraya-Watson vaut  $n^{2/5}$ .

Le deuxième est que les hypothèses sont  $f(x) > \frac{1-a}{2}$  pour la convergence ponctuelle et  $\inf_{x \in I} f(x) > \frac{1-a}{2}$  pour la convergence uniforme alors que les hypothèses usuelles pour l'estimation d'une fonction de régression sont  $f(x) > 0$  et  $\inf_{x \in I} f(x) > 0$ .

Ainsi, Mokkadem et al. [67] ont construit un estimateur plus performant que celui de Révész.

Dans un premier temps, ils ont généralisé l'estimateur de Révész en autorisant d'autres pas que le pas  $(n^{-1})$ . Cet estimateur est donné par

$$r_n^{GR}(x) = \left(1 - \frac{\gamma_n}{h_n} K\left(\frac{x - X_n}{h_n}\right)\right) r_{n-1}^{GR}(x) + \frac{\gamma_n}{h_n} Y_n K\left(\frac{x - X_n}{h_n}\right). \quad (1.2.3)$$

L'étude du comportement asymptotique de l'estimateur  $r_n^{GR}$  défini par l'équation (1.2.3) est faite sous les hypothèses suivantes :

(H1)  $K : \mathbb{R} \rightarrow \mathbb{R}$  est la fonction noyau, continue, bornée et vérifiant les propriétés de la Définition 1.1.1.

(H2) i)  $(\gamma_n) \in \mathcal{GS}(-\alpha)$  avec  $\alpha \in ]\frac{3}{4}, 1]$ ; de plus  $\lim_{n \rightarrow +\infty} (n\gamma_n)^{-1}$  existe.

ii)  $(h_n) \in \mathcal{GS}(-a)$  avec  $a \in ]\frac{1-\alpha}{4}, \frac{\alpha}{3}]$ .

(H3) i)  $g(s, t)$  est deux fois continuellement différentiable par rapport à la première variable.

ii)  $\forall q \in \{0, 1, 2\}$ ,  $s \mapsto \int_{\mathbb{R}} t^q g(s, t) dt$  est une fonction bornée et continue au point  $s = x \in \mathbb{R}$ .

Pour  $q \in \{2, 3\}$ ,  $s \mapsto \int_{\mathbb{R}} |t|^q g(s, t) dt$  est une fonction bornée.

iii) Pour  $q \in \{0, 1\}$ ,  $\int_{\mathbb{R}} |t|^q \left| \frac{\partial g}{\partial x} g(x, t) \right| dt < +\infty$  et la fonction  $s \mapsto \int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial x^2} g(s, t) dt$  est bornée et continue au point  $s = x$ .

Posons  $\xi = \lim_{n \rightarrow +\infty} (n\gamma_n)^{-1}$  et pour  $f(x) > 0$

$$m^{(2)}(x) = \frac{1}{2f(x)} \left[ \int_{\mathbb{R}} t \frac{\partial g}{\partial x} g(x, t) dt - r(x) \int_{\mathbb{R}} t \frac{\partial^2 g}{\partial x^2} g(x, t) dt \right] \sigma_K^2.$$

On donne la vitesse ponctuelle de convergence en loi de l'estimateur de Révész généralisé  $r_n^{GR}$  donné par (1.2.3).

**Théorème 1.2.4.** (Mokkadem et al. [67])

Supposons que les hypothèses (H1)-(H3) soient vérifiées pour  $x \in \mathbb{R}$  tel que  $f(x) > 0$ .

1. S'il existe  $c \geq 0$  tel que  $\gamma_n^{-1}h_n^5 \rightarrow c$ , et si  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{1-a}{2f(x)}$ , alors

$$\sqrt{\gamma_n^{-1}h_n} [r_n^{GR}(x) - r(x)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left( \frac{c^{1/2}m^{(2)}(x)}{f(x) - 2a\xi}, \frac{\text{Var}[Y|X=x]f(x)}{2f(x) - (1-a)\xi} \int_{\mathbb{R}} K^2(u)du \right).$$

2. Si  $\gamma_n^{-1}h_n^5 \rightarrow +\infty$ , et si  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{2a}{f(x)}$ , alors

$$\frac{1}{h_n^2} [r_n^{GR}(x) - r(x)] \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{f(x)m^{(2)}(x)}{f(x) - 2a\xi}.$$

Les conditions sur le pas  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{1-a}{2f(x)}$  (première partie du théorème) et  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{2a}{f(x)}$  (deuxième partie du théorème) remplacent les conditions sur la densité marginale  $f$  utilisées pour l'estimateur de Révész. On a également, pour un pas  $(\gamma_n) = (n^{-1})$ , les fenêtres  $((h_n) \in \mathcal{GS}(-a)$  avec  $a \in ]0, \frac{1}{3}[$ ) conduisent à des vitesses meilleures que celles de Révész.

On donne maintenant deux résultats : la vitesse de convergence presque sûre ponctuelle de l'estimateur de Révész généralisé et une majoration de sa vitesse de convergence presque sûre uniforme.

**Théorème 1.2.5.** (Mokkadem et al. [67])

Supposons que les hypothèses (H1)-(H3) soient vérifiées au point  $x \in \mathbb{R}$  tel que  $f(x) > 0$ . On pose  $s_n = \sum_{k=1}^n \gamma_k$ .

1. S'il existe  $c_1 \geq 0$  tel que  $\gamma_n^{-1}h_n^5/(\ln s_n) \xrightarrow[n \rightarrow +\infty]{} c_1$ , et si  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{1-a}{2f(x)}$ , alors, avec probabilité un, la suite

$$\left( \sqrt{\frac{\gamma_n^{-1}h_n}{2 \ln s_n}} [r_n^{GR}(x) - r(x)] \right)$$

est relativement compacte et l'ensemble de ses valeurs d'adhérence est l'intervalle

$$\left[ \sqrt{\frac{c_1}{2}} \frac{f(x)m^{(2)}(x)}{f(x) - 2a\xi} - \sqrt{\frac{\text{Var}[Y|X=x]f(x) \int_{\mathbb{R}} K^2(u)du}{2f(x) - (1-a)\xi}}, \right. \\ \left. \sqrt{\frac{c_1}{2}} \frac{f(x)m^{(2)}(x)}{f(x) - 2a\xi} + \sqrt{\frac{\text{Var}[Y|X=x]f(x) \int_{\mathbb{R}} K^2(u)du}{2f(x) - (1-a)\xi}} \right].$$

2. Si  $\gamma_n^{-1}h_n^5/(\ln s_n) \xrightarrow{n \rightarrow +\infty} +\infty$ , et si  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{2a}{f(x)}$ , alors

$$\frac{1}{h_n^2} [r_n^{GR}(x) - r(x)] \xrightarrow[n \rightarrow +\infty]{p.s.} \frac{f(x)m^{(2)}(x)}{f(x) - 2a\xi}.$$

Pour établir une majoration de la vitesse de convergence uniforme de  $r_n^{GR}$ , on ajoute l'hypothèse suivante.

- (H4) i)  $K$  est une fonction lipschitzienne.  
 ii) Il existe  $t^* > 0$  tel que  $\mathbb{E}[\exp(t^*|Y|)] < +\infty$ .  
 iii)  $a \in ]1 - \alpha, \alpha - \frac{2}{3}[$ .  
 iv) Pour  $q \in \{0, 1\}$ , la fonction  $x \mapsto \int_{\mathbb{R}} |t|^q |\frac{\partial g}{\partial x} g(x, t)| dt$  est bornée sur l'ensemble  $\{x, f(x) > 0\}$ .

**Théorème 1.2.6.** (Mokkadem et al. [67])

Soit  $I$  un intervalle borné de  $\mathbb{R}$  sur lequel  $\varphi = \inf_{x \in I} f(x) > 0$ . Supposons que les hypothèses (H1)-(H4) soient vérifiées pour tout  $x \in I$ .

1. Si la suite  $\gamma_n^{-1}h_n^5/(\ln n)^2$  est bornée et si  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{1-a}{2\varphi}$ , alors

$$\sup_{x \in I} |r_n^{GR}(x) - r(x)| = O\left(\sqrt{\gamma_n h_n^{-1} \ln n}\right) p.s.$$

2. Si  $\gamma_n^{-1}h_n^5/(\ln n)^2 \rightarrow +\infty$  et si  $\lim_{n \rightarrow +\infty} (n\gamma_n) > \frac{2a}{\varphi}$ , alors

$$\sup_{x \in I} |r_n^{GR}(x) - f(x)| = O(h_n^2) p.s.$$

Slaoui [106, 107] a établi les principes de grandes déviations et de déviations modérées pour l'estimateur  $r_n^{GR}$ . Concernant le choix pratique de fenêtre, le lecteur peut voir le travail de Slaoui [108] qui a utilisé la méthode de Plug-in.

L'avantage des estimateurs récursifs par rapport aux estimateurs non récursifs est le gain en temps de calcul. En effet, la récursivité facilite la mise à jour de l'estimateur lorsque on passe d'un échantillon de taille  $n$  à un échantillon de taille  $n + 1$ .

Malgré leur intérêt apporté au domaine de la statistique, les estimateurs à noyau, à la fois récurrents et non récurrents, connaissent quelques inconvénients parmi lesquels se situent les problèmes du support de la fonction à estimer lorsque celui-ci n'est plus  $\mathbb{R}$ . Ce phénomène, dont on parlera dans le paragraphe suivant, est appelé l'effet du bord.

### 1.2.3 Problèmes de support

Malgré leur importance, les estimateurs à noyau ont quelques inconvénients le plus connu est le problème de support. Dans le cas où la fonction à estimer aurait un support borné au moins d'un côté, l'utilisation d'un noyau symétrique est inconvenable puisque cela cause du biais au bord. Lorsque le lissage se fait aux points du bord, ce type de noyau, conçu pour l'estimation des fonctions régulières, introduit des poids à l'extérieur du support de la fonction à estimer qui devient discontinue près de ces points. Cela rend l'estimateur non consistant.

Pour expliquer ce phénomène, prenant le cas d'estimation d'une densité de probabilité  $f$  de support  $[0, +\infty[$ . Ici le problème se pose près de l'origine. Loin du bord, qui signifie donc pour tout  $x > h = h_n$ , on rappelle l'expression de l'espérance (respectivement de la variance) donnée par (1.1.6) (respectivement (1.1.7)) de l'estimateur  $\widehat{f}_n$  introduit par (1.1.1).

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

et

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

Maintenant, près de l'origine et en écrivant  $x = ph$ , on trouve

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

et

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



En particulier,  $\mathbb{E} [\widehat{f}_n(0)] \simeq f(0)/2$ , qui est la moitié de la vraie valeur. Cet effet se trouve aussi dans le cas d'estimation d'une fonction de distribution ou d'une fonction de régression.

Plusieurs solutions à ce problème ont été proposées dans la littérature. Schuster [98], Silverman [101] et Cline et Hart [17] ont utilisé la méthode dite des données reflétées ("reflection data method") pour une fonction de densité  $f$  définie sur un intervalle du type  $[0, b]$  avec  $f'(0) = f'(b) = 0$ . Ensuite, Gasser et Müller [31], Gasser et al. [32], Jones [45], Zhang et Karunamuni [128] Karunamuni et Alberts [49] et Karunamuni et Sriram [50] ont suggéré l'utilisation de la méthode des noyaux du bord qui consiste à modifier le noyau près du bord. Le problème de ces nouveaux noyaux est qu'ils prennent des valeurs négatives. Pour corriger cette déficience, certaines solutions ont été proposés. Pour plus de détails, on se réfère aux travaux de Jones [45], Jones et Foster [46], Gasser et Müller [31], Karunamuni et Sriram [50] ainsi que Karunamuni et Zhang [51].

Toujours dans le cadre de résolution du problème d'effet du bord, Cowling et Hall [20] ont proposé la méthode des pseudo-données ("pseudo-data method"). Ils ont généré des observations supplémentaires par la méthode dite les "trois points pour mieux régner" et ils les ont combinées avec les observations d'origine pour former un nouveau estimateur à noyau. Dans le même contexte, Marron et Ruppert [62] ont proposé la méthode de transformation ("transformation method"). L'idée de cette méthode est constituée de trois étapes. Tout d'abord, trouver une transformation  $T$  de telle sorte que la première dérivée de la densité des variables  $Y_i = T(X_i)$  soit approximativement nulle aux bords. Puis, construire un estimateur à noyau de la réflexion pour les  $Y_i$ . Enfin, par la formule de changement des variables, obtenir une estimation de  $f$ .

Dans le même sujet, d'autres auteurs ont proposé l'utilisation de noyaux dont le support coïncide avec celui de la fonction à estimer. Ces noyaux utilisés sont généralement asymétriques et peuvent changer de forme selon la position du point d'estimation ce qui permet de résoudre l'effet du bord.

Dans ce cadre, on cite le travail de Chen [14, 15] avec les noyaux beta et gamma pour estimer les densités à support respectivement  $[0, 1]$  et  $[0, +\infty[$  et Scaillet [94] avec les noyaux inverses gaussien et sa réciproque pour les densités à support  $]0, +\infty[$ .

Dans le paragraphe suivant, on présente la méthode introduite par Vitale [119] qui a construit des estimateurs basé sur les polynômes de Bernstein et le théorème de Weierstrass pour approximer une fonction continue à support dans  $[0, 1]$ .

Cette méthode a été largement discutée dans plusieurs travaux comme Tenbusch [115], Ghosa [33], Kakizawa [47, 48], Igarashi and Kakizawa [42], Rao [84], Leblanc [55], Leblanc [56, 57], Babu et al. [2], Babu et Chaubey [3] and Jmaei et al. [43]. On se réfère à cette méthode aussi dans les chapitres qui suivent pour construire nos estimateurs.

### 1.3 Estimateurs fonctionnels utilisant les polynômes de Bernstein

Les premiers pas des méthodes d'estimation non paramétrique basés sur les polynômes de Bernstein ont été fait en 1912 quand Bernstein a cherché une démonstration constructive et probabiliste du théorème classique de Weierstrass pour l'approximation des fonctions continues sur des intervalles de type  $[a, b]$ .

Dans ce cadre, Bernstein [5] a introduit une famille de polynômes, qui portera son nom par la suite, dont la définition est la suivante.

**Définition 1.3.1.** (*Polynômes de Bernstein*)

Pour  $m \in \mathbb{N}$  et  $0 \leq k \leq m$ , on appelle polynôme de Bernstein le polynôme

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

Ces polynômes possèdent plusieurs propriétés intéressantes dans le domaine de probabilités et statistiques. On en cite quelques-unes les plus connues.

**Proposition 1.3.1.**

Les polynômes de Bernstein possèdent les propriétés suivantes :

(i) *Partition de l'unité :*

$$\sum_{k=0}^m b_k(m, x) = 1, \quad \forall x \in [0, 1].$$

(ii) *Positivité :*

$$\forall k \in \{0, \dots, m\}, \quad b_k(m, x) \geq 0,$$

et  $b_k(m, x)$  ne s'annule qu'en 0 et en 1.

(iii) *Symétrie :*

$$\forall k \in \{0, \dots, m\}, \quad b_k(m, x) = b_{m-k}(m, 1-x).$$

(iv) *Formule de récurrence :* Pour  $m > 0$ ,

$$b_k(m, x) = \begin{cases} (1-x)b_k(m-1, x) & \text{si } k = 0 \\ (1-x)b_k(m-1, x) + xb_{k-1}(m-1, x) & \forall k \in \{1, \dots, m-1\} \\ xb_{k-1}(m-1, x) & \text{si } k = m. \end{cases}$$

D'un point de vue probabiliste, on peut voir le polynôme  $b_k(m, p)$  comme la probabilité  $\mathbb{P}(X = k)$ , où  $X$  est une variable aléatoire suivant une loi binomiale de paramètre  $(m, p)$ . C'est l'approche que Bernstein a fait dans sa démonstration du théorème de Weierstrass qui s'annonce autrement comme suit.

**Théorème 1.3.1.** (Bernstein [5])

Soit  $f : [0, 1] \rightarrow \mathbb{R}$  une fonction continue. On définit le polynôme de Bernstein associé à  $f$  d'ordre  $m \in \mathbb{N}$  :

$$\begin{aligned} \forall x \in [0, 1], \quad B_m(f)(x) &= \sum_{k=0}^m f\left(\frac{k}{m}\right) b_k(m, x), \\ &= \sum_{k=0}^m f\left(\frac{k}{m}\right) \binom{m}{k} x^k (1-x)^{m-k}. \end{aligned}$$

Alors, on a

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

*En particulier, toute fonction continue sur  $[0, 1]$  est limite uniforme d'une suite de polynômes de Bernstein.*

D'autres supports différents de l'intervalle  $[0, 1]$  sont possibles en faisant des transformations appropriées. Pour plus de détails sur les polynômes de Bernstein, le lecteur peut consulter le livre de Lorentz [59].

Vitale [119] a utilisé ce théorème pour construire un estimateur de la fonction de distribution dont on parlera dans le paragraphe suivant.

### 1.3.1 Estimateur d'une distribution

Soient  $X_1, \dots, X_n$  des variables aléatoires i.i.d. ayant une fonction de distribution inconnue  $F$  et sa densité associée  $f$  de support  $[0, 1]$ . Vitale [119] a défini l'estimateur de  $F$  comme suit

$$\tilde{F}_{n,m}(x) = \sum_{k=0}^m F_n \left( \frac{k}{m} \right) b_k(m, x), \quad (1.3.1)$$

où  $F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq t\}}$  est la fonction de distribution empirique. On remarque que  $\tilde{F}_{n,m}(0) = 0 = F(0)$  et  $\tilde{F}_{n,m}(1) = 1 = F(1)$ .

Cet estimateur défini en (1.3.1) a été en outre étudié sous une différente forme par Babu et al. [2] et Leblanc [55, 56]. En effet, Babu et al. [2] ont montré que  $\tilde{F}_{n,m}$  vérifie les propriétés d'une fonction de distribution à savoir :

$\tilde{F}_{n,m}$  est continue sur  $[0, 1]$ ,  $0 \leq \tilde{F}_{n,m} \leq 1$  pour  $x \in [0, 1]$  et pour la monotonie il suffit d'écrire cet estimateur comme

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

avec

$$f_n(0) = 0, \quad f_n\left(\frac{k}{m}\right) = F_n\left(\frac{k}{m}\right), \quad k = 1, \dots, m \quad (1.3.2)$$

et

$$B_k(m, x) = \sum_{j=k}^m b_k(m, x) = m \binom{m-1}{k-1} \int_0^x t^{k-1} (1-t)^{m-k} dt. \quad (1.3.3)$$

Pour  $k \in \{0, \dots, m\}$ ,  $f_n\left(\frac{k}{m}\right)$  est positive d'après (1.3.2) et  $B_k(m, x)$  est une fonction croissante par rapport à  $x$  d'après (1.3.3). Ainsi, la fonction  $\tilde{F}_{n,m}$  est croissante sur  $[0, 1]$  et elle se qualifie comme un estimateur de la distribution  $F$ .

On donne dans la proposition suivante l'espérance et la variance de l'estimateur  $\tilde{F}_{n,m}$ .

**Proposition 1.3.2.** (Leblanc [56])

On suppose que  $F$  soit une fonction continue et admette deux dérivées bornées sur  $[0, 1]$ . Pour  $x \in [0, 1]$ , on a

$$\mathbb{E} \left[ \tilde{F}_{n,m}(x) \right] = F(x) + m^{-1}b(x) + o(m^{-1}),$$

avec  $b(x) = \frac{x(1-x)f'(x)}{2}$ . On a aussi

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

avec  $\sigma^2(x) = F(x)[1 - F(x)]$ , et  $V(x) = f(x) \left[ \frac{2x(1-x)}{\pi} \right]^{1/2}$ .

Pour plus de détails, le lecteur peut voir le travail de Leblanc [56].

Notez que le résultat précédent implique que  $\tilde{F}_{n,m}$  a un biais uniforme dans l'intervalle unité d'où l'absence du biais de bords. D'autre part, de sa définition donnée en (1.3.1), il semble naturel de poser  $h = m^{-1}$  comme fenêtre de l'estimateur de Bernstein  $\tilde{F}_{n,m}$ . Alors, le biais de  $\tilde{F}_{n,m}$  devient  $O(m^{-1}) = O(h)$  qui est plus grand que le biais obtenu avec les estimateurs à noyaux ayant généralement un biais de l'ordre de  $O(h^2)$ .

Une autre conséquence du résultat précédent est que  $\tilde{F}_{n,m}$  domine asymptotiquement la fonction de distribution empirique à chaque  $x \in ]0, 1[$  en termes de l'erreur quadratique moyenne  $MSE$ . En effet, on a

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

En revanche, il est bien connu que

$$MSE [F_n(x)] = Var [F_n(x)] = n^{-1} \sigma^2(x),$$

de sorte que  $\tilde{F}_{n,m}$  et  $F_n$  sont équivalents en terme de MSE jusqu'au premier ordre. Cependant, en considérant aussi des termes d'ordre supérieur, il s'avère que  $\tilde{F}_{n,m}$  admet une  $MSE$  plus petite que  $F_n$  lorsque l'ordre  $m$  est bien choisi. Les deux estimateurs atteignent une  $MSE$  égale à zéro pour  $x = 0, 1$ .

On donne maintenant l'erreur quadratique intégrée ( $MISE$ ) de l'estimateur  $\tilde{F}_{n,m}$

$$MISE \left( \tilde{F}_{n,m} \right) = n^{-1} C_1 - m^{-1/2} n^{-1} C_2 + m^{-2} C_3 + o(m^{-1/2} n^{-1}) + o(m^{-2}),$$

avec  $C_1 = \int_0^1 \sigma^2(x) dx$ ,  $C_2 = \int_0^1 V(x) dx$ , et  $C_3 = \int_0^1 b^2(x) dx$ .

Notez que les constantes  $C_1$ ,  $C_2$  et  $C_3$  sont toutes strictement positives, sauf dans le cas trivial où  $f$  est la densité uniforme, auquel cas  $C_3 = 0$ . Le choix optimale de l'ordre  $m$  qui minimise la  $MISE$  est donné par

$$m_{opt} = \left[ \frac{4C_3}{C_2} \right]^{2/3} n^{2/3},$$

et dans ce cas, on a

$$MISE \left( \tilde{F}_{n,m_{opt}} \right) = n^{-1} C_1 - n^{-4/3} \frac{3}{4} \left[ \frac{C_2^{4/3}}{4^{1/3} C_3^{1/3}} \right] + o(n^{-4/3}).$$

Des résultats similaires à ceux-ci ont été obtenus pour les estimateurs classiques à noyau par Jones [44], entre autres.

Notez cependant que l'approche de plug-in suggérée par Altman et Léger [1] et la méthode de validation croisée de Bowman et al. [10] pourraient être certainement adapter dans le choix pratique de  $m$ .

L'estimateur  $\tilde{F}_{n,m}$  est fortement consistant. En effet on a le résultat suivant :

**Théorème 1.3.2.** (*Babu et al. [2]*)

Soit  $F$  une fonction de distribution continue sur l'intervalle  $[0, 1]$ . Si  $m, n \rightarrow +\infty$ , alors

$$\left\| \tilde{F}_{n,m} - F \right\| = \sup_{x \in [0,1]} \left| \tilde{F}_{n,m}(x) - F(x) \right| \xrightarrow[n \rightarrow +\infty]{} 0, \quad p.s.$$

Le théorème suivant donne la distance entre l'estimateur  $\tilde{F}_{n,m}$  et la fonction de distribution empirique  $F_n$ .

**Théorème 1.3.3.** (*Babu et al. [2]*)

Soit  $F$  une fonction de distribution continue, différentiable sur l'intervalle  $[0, 1]$  et de densité  $f$ . Si  $f$  est lipschitzienne d'ordre 1, alors pour  $n^{2/3} \leq m \leq (n/\ln n)^2$  on a

$$\left\| \tilde{F}_{n,m} - F_n \right\| = O \left( (n^{-1} \ln n)^{1/2} (m^{-1} \ln m)^{1/4} \right).$$

Pour la démonstration de ces deux théorèmes, on se réfère au travail du Babu et al. [2].

Leblanc [56] a établi la normalité asymptotique de l'estimateur Bernstein  $\tilde{F}_{n,m}$  pour chaque  $x$  dans l'intervalle unité. Ce résultat est donné par le théorème suivant

**Théorème 1.3.4.** (*Leblanc [56]*)

On suppose que  $F$  soit une fonction continue et admettant deux dérivées bornées sur  $[0, 1]$ . Pour  $x \in ]0, 1[$ , tel que  $0 < F(x) < 1$ , on a

$$n^{1/2} \left[ \tilde{F}_{n,m}(x) - B_m(x) \right] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(x)),$$

avec  $B_m(x) = \mathbb{E} \left[ \tilde{F}_{n,m}(x) \right] = \sum_{k=0}^m F \left( \frac{k}{m} \right) b_k(m, x).$

Notez que, selon un choix approprié de fenêtre, un résultat similaire à ce théorème a été obtenu par Watson et Leadbetter [121] pour les estimateurs classiques à noyau de la fonction de distribution. Maintenant, comme l'intérêt est principalement dans la façon dont  $\tilde{F}_{n,m}$  se comporte par rapport à  $F$ , en remarquant que

$$n^{1/2} \left[ \tilde{F}_{n,m}(x) - F(x) \right] = n^{1/2} \left[ \tilde{F}_{n,m}(x) - B_m(x) \right] + m^{-1}n^{1/2}b(x) + o(m^{-1}n^{1/2}),$$

on obtient le corollaire suivant

**Corollaire 1.3.1.** (*Leblanc [56]*)

*On suppose que  $F$  soit une fonction continue et admettant deux dérivées bornées sur  $[0, 1]$ . Pour  $x \in ]0, 1[$ , tel que  $0 < F(x) < 1$ , on a*

1. Si  $m^{-1}n^{1/2} \xrightarrow[n \rightarrow +\infty]{} c$  pour une constante  $c \geq 0$ ,

$$n^{1/2} \left[ \tilde{F}_{n,m}(x) - F(x) \right] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(cb(x), \sigma^2(x)).$$

2. Si  $m^{-1}n^{1/2} \xrightarrow[n \rightarrow +\infty]{} +\infty$ ,

$$m^{-1} \left[ \tilde{F}_{n,m}(x) - F(x) \right] \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} b(x).$$

Ce résultat établit essentiellement, que lorsque l'ordre  $m$  de l'estimateur de Vitale est choisi assez grand (de sorte que le biais devient négligeable) la distribution asymptotique de  $\tilde{F}_{n,m}$  est la même que celle de la distribution empirique  $F_n$ .

Partant de cet estimateur  $\tilde{F}_{n,m}$ , Vitale a construit par une dérivation l'estimateur  $\tilde{f}_{n,m}$  qui sera étudié dans le paragraphe suivant avec l'estimateur de Leblanc [55] qui réduit le biais.



### 1.3.2 Estimateurs d'une densité

Pour obtenir l'estimateur de Bernstein de la densité à partir de l'estimateur de distribution, on utilise la formule de Babu et al. [2]

$$\tilde{F}_{n,m}(x) = \sum_{k=0}^m f_n\left(\frac{k}{m}\right) B_k(m, x), \quad (1.3.4)$$

avec  $f_n\left(\frac{k}{m}\right)$  et  $B_k(m, x)$  sont donnés respectivement par (1.3.2) et (1.3.3).

En dérivant par rapport à  $x$  l'équation (1.3.4), on obtient

$$\begin{aligned} \tilde{f}_{n,m}(x) &= \sum_{k=1}^m f_n\left(\frac{k}{m}\right) \frac{d}{dx} B_k(m, x) \\ &= m \sum_{k=1}^{m-1} f_n\left(\frac{k+1}{m}\right) b_k(m-1, x) \\ &= m \sum_{k=1}^{m-1} \left[ F_n\left(\frac{k+1}{m}\right) - F_n\left(\frac{k}{m}\right) \right] b_k(m-1, x). \end{aligned} \quad (1.3.5)$$

Cet estimateur  $\tilde{f}_{n,m}$  donné par (1.3.5) a été introduit par Vitale [119] et étudié après par [2]. Par ailleurs, Bouezmarni and Rolin [8] ont considéré le cas des densités non bornées et Tenbusch [115] et Babu et Chaubey [3] ont fait la généralisation de cet estimateur dans le cas multidimensionnel.

Kakizawa [47] et Rao [84] ont défini d'autres Bernstein-estimateurs basés sur les noyaux du bord et utilisant les polynômes de Bernstein généralisés.

Enfin, Petrone [77] a introduit une approche entièrement Bayésienne de l'estimation non paramétrique pour une densité définie sur un intervalle compact grâce à l'utilisation de polynômes de Bernstein. L'estimateur obtenu est une moyenne pondérée de différents estimateurs  $\tilde{f}_{n,m}$  pour des différentes valeurs de  $m$  et repose sur la famille de Bernstein-Dirichlet introduit par Petrone [78] et a été étudiée par Ghosa [33] et Petrone et Wasserman [79].

Kruijer et van der Vaart [53] ont considéré une version modifiée de l'estimateur de Petrone.

On donne par la suite les propriétés de l'estimateur  $\tilde{f}_{n,m}$  défini par (1.3.5). Pour plus de détails, le lecteur peut voir les travaux de Leblanc [56] et Babu et al. [2].

Les résultats concernant l'espérance et la variance de l'estimateur  $\tilde{f}_{n,m}$  sont donnés dans la proposition suivante .

**Proposition 1.3.3.** (Leblanc [56])

Soit  $f$  une densité de probabilité. On suppose que  $f$  soit une fonction continue et admettant deux dérivées bornées sur  $[0, 1]$ . Pour  $x \in [0, 1]$ , on a

$$\mathbb{E}[\tilde{f}_{n,m}(x)] = f(x) + m^{-1}\Delta_1(x) + o(m^{-1}),$$

avec

$$\Delta_1(x) = \frac{1}{2} [(1 - 2x)f'(x) + x(1 - x)f''(x)].$$

D'autre part, on a

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

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

On note que le biais de l'estimateur  $\tilde{f}_{n,m}$  est uniforme sur l'intervalle  $[0, 1]$  d'où l'absence du biais de bords. D'autre part, en posant  $h = m^{-1}$  la fenêtre de l'estimateur de Vitale pour la densité, on obtient un biais de l'ordre de  $O(h)$  pour l'estimateur  $\tilde{f}_{n,m}$  qui est plus grand que le biais obtenu par l'estimateur à noyau classique  $\hat{f}_n$ .

Par ailleurs, pour la variance de  $\tilde{f}_{n,m}$ , elle est de l'ordre de  $O(h^{-1/2}/n)$  pour  $x \in ]0, 1[$  par contre la variance est de l'ordre de  $O(h^{-1}/n)$  pour l'estimateur à noyau classique  $\hat{f}_n$ . Ainsi, on obtient une réduction de la variance. On remarque aussi que la variance est de l'ordre de  $O(m^{1/2}/n)$  à l'intérieur de l'intervalle  $[0, 1]$  mais elle est d'ordre plus grand égale à  $O(m/n)$  aux frontières. En effet, en examinant l'expression de la variance de  $\tilde{f}_{n,m}$  pour  $x \in ]0, 1[$ , on constate que  $\psi(x) \rightarrow +\infty$  lorsque  $x \rightarrow 0$  et  $x \rightarrow 1$  ceci explique l'augmentation de la variance près des bords. Mais le comportement de  $\psi$  aux voisinages de frontières n'a aucun effet sur la *MISE* de l'estimateur  $\tilde{f}_{n,m}$ . Ceci est vu dans le résultat suivant, pour lequel on pose

$$K_1 = \int_0^1 f(x)\psi(x)(x)dx, \quad K_2 = \int_0^1 \Delta_1^2(x)dx.$$

**Proposition 1.3.4.** (Leblanc [56])

Si  $f$  est une fonction continue et admettant deux dérivées bornées sur  $[0, 1]$ , alors

$$MISE\left(\tilde{f}_{n,m}\right) = \frac{m^{1/2}}{n}K_1 + m^{-2}K_2 + o\left(\frac{m^{1/2}}{n}\right) + o\left(m^{-2}\right).$$

Ainsi le choix optimal de  $m$  qui minimise la  $MISE$  est

$$m_{opt} = \left[\frac{4K_2}{K_1}\right]^{2/5} n^{2/5},$$

et dans ce cas, on a

$$MISE\left(\tilde{f}_{n,m_{opt}}\right) = \frac{5}{4}4^{1/5}K_1^{4/5}K_2^{1/5}n^{-4/5} + o\left(n^{-4/5}\right).$$

On note que le résultat de la proposition précédente n'est pas obtenue directement à partir du biais et de la variance puisque l'erreur de cette dernière n'est pas uniforme en  $x$ . Plutôt, l'expression de  $MISE$  de l'estimateur  $\tilde{f}_{n,m}$  est obtenue en utilisant le théorème de Fubini et en intégrant le biais et la variance dont l'expression est donnée par

$$Var\left[\tilde{f}_{n,m}(x)\right] = \frac{m^2}{n}\left[A_m(x) - f_m^2(x)\right],$$

avec

$$\begin{aligned} A_m(x) &= \sum_{k=0}^m \left[ F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right) \right] b_k^2(m, x), \\ f_m(x) &= \mathbb{E}\left[\tilde{f}_{n,m}(x)\right] = \sum_{k=0}^m \left[ F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right) \right] b_k(m, x). \end{aligned}$$

De plus, l'estimateur  $\tilde{f}_{n,m}$  atteint le taux optimal de convergence en  $MISE$  pour les estimateurs de densité qui est de l'ordre de  $n^{-4/5}$ . Comparé avec les estimateurs à noyau classiques,  $\tilde{f}_{n,m}$  atteint cet ordre avec un biais plus grand et une variance plus petite. Enfin, on note que la fenêtre optimale pour  $\tilde{f}_{n,m}$  est  $h \simeq n^{-2/5}$  au lieu de  $h \simeq n^{-1/5}$  pour les estimateurs à noyau classiques.

On rappelle dans la suite, quelques résultats concernant la convergence de l'estimateur  $\tilde{f}_{n,m}$  donnés par Babu et al. [2].

**Théorème 1.3.5.** (Babu et al. [2])

Pour  $2 \leq m \leq (n/\ln n)$ , on a

$$\left\| \tilde{f}_{n,m} - f \right\| = \sup_{x \in [0,1]} \left| \tilde{f}_{n,m}(x) - f(x) \right| = O \left( m^{1/2} \left( \frac{\ln n}{n} \right)^{1/2} \right) + O(\|F'_m - f\|),$$

avec  $F'_m$  est la dérivée de la fonction  $F_m(x) = \sum_{k=0}^m F \left( \frac{k}{m} \right) b_k(m, x)$ .

Ainsi, si  $m = o(n/\ln n)$  on a

$$\left\| \tilde{f}_{n,m} - f \right\| \xrightarrow[n \rightarrow +\infty]{} 0 \quad p.s.$$

On donne, par la suite, deux théorèmes sur la normalité asymptotique de  $\tilde{f}_{n,m}$ .

**Théorème 1.3.6.** (Babu et al. [2])

Pour  $x \in [0, 1]$  tel que  $f(x) > 0$  et lorsque  $2 \leq m \leq (n/\ln n)$  et  $n^{2/3}m \xrightarrow[n \rightarrow +\infty]{} 0$ , on a

$$n^{1/2}m^{-1/4} \left[ \tilde{f}_{n,m}(x) - f(x) \right] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, f(x)\psi(x)).$$

**Théorème 1.3.7.** (Babu et al. [2])

Supposons que la densité  $f$  admette deux dérivées continues sur  $[0, 1]$ . Pour  $x \in [0, 1]$

tel que  $f(x) > 0$  et si  $n^{-1/5}m^{1/2} \xrightarrow[n \rightarrow +\infty]{} c$  avec  $c > 0$ , on a

$$n^{2/5} \left[ \tilde{f}_{n,m}(x) - f(x) \right] - c^{-2}\Delta_1(x) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, cf(x)\psi(x)).$$

Comme on a mentionné, l'estimateur  $\tilde{f}_{n,m}$  a un biais plus grand que les estimateurs classiques de la densité. En effet, en utilisant la méthode de réduction du biais, ce problème peut être corrigé sans augmenter la variance.

Cette méthode a été utilisée par [82] dans le cadre d'estimation de la densité spectrale qui fait le lien avec le travail de Schucany et al. [95] et Schucany et Sommers [96] sur la réduction du biais des estimateurs.

Leblanc [55] a utilisé cette même méthode pour introduire un nouveau estimateur qui a un biais plus petit. Cet estimateur est donné par

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,m/2}(x) = 2\tilde{f}_{n,m}(x) - \tilde{f}_{n,m/2}(x). \quad (1.3.6)$$

L'étude des caractéristiques de  $\tilde{f}_{n,m,m/2}$  se fait sous l'hypothèse suivante :

(H) La densité  $f$  est continue et bornée admettant quatre dérivées continues et bornées sur  $[0, 1]$ .

La proposition suivante donne l'espérance et la variance de l'estimateur  $\tilde{f}_{n,m,m/2}$ .

**Proposition 1.3.5.** (Leblanc [55])

Pour  $x \in [0, 1]$ , on a

$$\mathbb{E} \left[ \tilde{f}_{n,m,m/2}(x) \right] = f(x) - 2m^{-2}\Delta_2(x) + o(m^{-2}),$$

avec

$$\Delta_2(x) = \frac{1}{6}(1 - 6x(1 - x))f''(x) + \frac{5}{12}x(1 - x)(1 - 2x)f^{(3)}(x) + \frac{1}{8}x^2(1 - x)^2f^{(4)}(x).$$

Aussi,

$$\text{Var}[\tilde{f}_{n,m,m/2}(x)] = \begin{cases} \left( \frac{1}{\sqrt{2}} + 4 \left( 1 - \sqrt{\frac{2}{3}} \right) \right) \frac{m^{1/2}}{n} f(x)\psi(x) + o_x \left( \frac{m^{1/2}}{n} \right) & \text{si } x \in ]0, 1[, \\ \frac{5}{2} \frac{m}{n} f(x) + o_x \left( \frac{m}{n} \right) & \text{si } x = 0, 1. \end{cases}$$

On constate la réduction du biais de l'estimateur  $\tilde{f}_{n,m,m/2}$  pour devenir de l'ordre de  $O(m^{-2})$  au lieu de  $O(m^{-1})$  pour l'estimateur de Vitale et ceci est uniformément sur l'intervalle  $[0, 1]$  qui implique l'absence du biais du bord.

D'autre part, pour  $x \in ]0, 1[$ , on a

$$\text{Var}[\tilde{f}_{n,m,m/2}(x)] \simeq 1.441 \text{Var}[\tilde{f}_{n,m}(x)] + o_x(1),$$

alors la correction du biais en utilisant l'équation (1.3.6) au lieu de l'estimateur de Vitale entraîne une légère augmentation de la variance. Aussi, en posant  $h = m^{-1}$  la fenêtre de l'estimateur  $\tilde{f}_{n,m,m/2}$ , on obtient un biais de l'ordre de  $O(h^2)$  et une variance qui reste de l'ordre de  $O(h^{-1/2}/n)$ . Sous l'hypothèse que  $f$  soit quatre fois continument différentiable sur  $[0, 1]$ , les estimateurs classiques basés sur des noyaux d'ordre supérieurs peuvent atteindre un biais d'ordre  $O(h^4)$  sauf peut être aux bords et une variance d'ordre  $O(h^{-1}/n)$ . Ainsi, malgré la correction du biais, l'estimateur  $\tilde{f}_{n,m,m/2}$  a encore un biais plus grand et une variance plus petite que les estimateurs à noyau classiques.

Pour obtenir la  $MISE$  de l'estimateur  $\tilde{f}_{n,m,m/2}$ , on rappelle les constantes utilisées pour l'estimateur de Vitale  $\tilde{f}_{n,m}$

$$K_1 = \int_0^1 f(x)\psi(x)(x)dx, \quad K_2 = \int_0^1 \Delta_1^2(x)dx,$$

et on définit

$$K_3 = \frac{1}{\sqrt{2}} + 4 \left( 1 - \sqrt{\frac{2}{3}} \right), \quad K_4 = \int_0^1 \Delta_2^2(x)dx.$$

**Proposition 1.3.6.** (Leblanc [55])

Supposons que la densité  $f$  soit quatre fois continument différentiable sur  $[0, 1]$ , on a

$$MISE \left( \tilde{f}_{n,m,m/2} \right) = \frac{m^{1/2}}{n} K_1 K_3 + 4m^{-4} K_4 + o(m^{-4}).$$

Ainsi, l'ordre optimal qui minimise la  $MISE$  de  $\tilde{f}_{n,m,m/2}$  est égale

$$m_{opt} = \left[ \frac{32K_4}{K_1 K_3} \right]^{2/9} n^{2/9},$$

pour lequel on a

$$MISE \left( \tilde{f}_{n,m_{opt},m_{opt}/2} \right) = \frac{9(32K_1^8 K_4)^{1/9}}{8n^{8/9}} + o(n^{-8/9}).$$

Ce résultat implique que l'estimateur  $\tilde{f}_{n,m,m/2}$  atteint un taux optimal de convergence en terme de  $MISE$  pour les densités qui sont quatre fois continument différentiable. Ce taux de convergence est obtenu avec une fenêtre  $h \simeq n^{-2/9}$  au lieu de  $h \simeq n^{-1/9}$  qui est le choix optimal pour les estimateurs à noyau classiques.

Igarashi and Kakizawa [42] ont généralisé l'estimateur de Leblanc  $\tilde{f}_{n,m,m/2}$  (1.3.6) pour devenir

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,m/b}(x) = \frac{b}{b-1} \tilde{f}_{n,m}(x) - \frac{1}{b-1} \tilde{f}_{n,m/b}(x), \quad (1.3.7)$$

avec  $b = 2, 3, \dots$ ,  $\tilde{f}_{n,m}$  et  $\tilde{f}_{n,m/b}$  sont les estimateurs de Vitale (1.3.5) avec l'ordre  $m$  and  $m/b$  respectivement.

La proposition suivante donne l'espérance, la variance et la  $MISE$  de l'estimateur  $\tilde{f}_{n,m,m/b}$  (1.3.7)

**Proposition 1.3.7.** (Igarashi and Kakizawa [42])

Supposons que la densité  $f$  soit continue, bornée et admettant quatre dérivées continues et bornées sur  $[0, 1]$ , on a

$$\mathbb{E}[\tilde{f}_{n,m,m/b}(x)] - f(x) = -\frac{b}{m^2}\Delta_2(x) + o(m^{-2}), \forall x \in [0, 1].$$

D'autre part, on a

$$\text{Var}[\tilde{f}_{n,m,m/b}(x)] = \begin{cases} \lambda_1(b)\frac{m^{1/2}}{n}f(x)\psi(x) + o_x\left(\frac{m^{1/2}}{n}\right) & \text{si } x \in ]0, 1[, \\ \lambda_2(b)\frac{m}{n}f(x) + o_x\left(\frac{m}{n}\right) & \text{si } x = 0, 1, \end{cases}$$

avec

$$\begin{aligned} \lambda_1(b) &= \frac{1}{(1-b)^2} \left\{ b^2 + b^{-1/2} - 2b \left( \frac{2}{b+1} \right)^{1/2} \right\}, \\ \lambda_2(b) &= \frac{1}{(1-b)^2} \{ b^2 + b^{-1} - 2 \}, \end{aligned}$$

Par ailleurs, on a

$$\text{MISE}(\tilde{f}_{n,m,m/b}) = \lambda_1(b)K_1\frac{m^{1/2}}{n} + \frac{b^2K_2}{m^4} + o(m^{1/2}n) + o(m^{-4}).$$

Ainsi, l'ordre optimal qui minimise la MISE de  $\tilde{f}_{n,m,m/b}$  est égale à

$$\left( \left[ \frac{b^2}{\lambda_1(b)} \frac{8K_2}{K_1} \right]^{2/9} n^{2/9} \right),$$

pour lequel, on a

$$\text{MISE}(\tilde{f}_{n,m,m/b}) = \frac{9}{8} (b\lambda_1^4(b))^{2/9} (8K_1^8K_2)^{1/9} n^{-8/9} + o(n^{-8/9}).$$

Ce dernier résultat implique que  $b = 2$  est le choix optimal en terme de MISE de l'estimateur  $\tilde{f}_{n,m,m/b}$ , puisque la fonction  $b \rightarrow (b\lambda_1^4(b))^{2/9}$  est croissante.

Comme on a mentionné, cette méthode permet la réduction du biais mais les estimateurs obtenus ne sont pas positifs. Pour rectifier ce problème, Terrell et Scott [116] ont développé une méthode multiplicative pour la correction du biais.

Cette méthode a été adaptée par Hirukawa [41] pour l'estimateur de noyau beta introduit par Chen [14] et par Igarashi and Kakizawa [42] dans le cadre d'estimation par les polynômes de Bernstein pour obtenir l'estimateur suivant

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,b,\varepsilon}(x) = \left\{ \tilde{f}_{n,m}(x) \right\}^{b/(b-1)} \left\{ \tilde{f}_{n,m/b}(x) + \varepsilon \right\}^{-1/b-1}, \quad (1.3.8)$$

pour  $\varepsilon = \varepsilon(m) > 0$ , qui tend vers zéro lorsque  $m$  tend vers l'infini .

La proposition suivante donne l'espérance, la variance et la *MISE* de l'estimateur  $\tilde{f}_{n,m,b,\varepsilon}$  (1.3.8)

**Proposition 1.3.8.** (*Igarashi and Kakizawa [42]*)

Supposons que la densité  $f$  soit continue, bornée et admettant quatre dérivées continues et bornées sur  $[0, 1]$ . Pour  $x \in [0, 1]$  tel que  $f(x) > 0$  avec  $m = O(n^\eta)$  et  $\varepsilon \approx m^\tau$  où  $\eta \in ]0, 1[$  et  $\tau > 2$ , on a

$$\begin{aligned} \mathbb{E}[\tilde{f}_{n,m,b,\varepsilon}(x)] - f(x) &= -\frac{b}{m^2} \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) \right\} \\ &\quad + O\left( \text{Var}[\tilde{f}_{n,m}(x)] + \text{Var}[\tilde{f}_{n,m/b}(x)] \right) + o(m^{-2}). \end{aligned}$$

D'autre part, on a

$$\text{Var}[\tilde{f}_{n,m,b,\varepsilon}(x)] = \text{Var}[\tilde{f}_{n,m,m/b}(x)] + o\left( \text{Var}[\tilde{f}_{n,m}(x)] + \text{Var}[\tilde{f}_{n,m/b}(x)] + m^{-4} \right),$$

Par ailleurs, on a

$$\text{MISE}(\tilde{f}_{n,m,b,\varepsilon}) = \lambda_1(b)K_1 \frac{m^{1/2}}{n} + \frac{b^2 K_5}{m^4} + o(m^{1/2}n) + o(m^{-4}),$$

$$\text{avec } K_5 = \int_0^1 \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) \right\}^2 dx.$$

Pour minimiser la *MISE* de  $\tilde{f}_{n,m,b,\varepsilon}$ , l'ordre  $(m_n)$  doit être égale à

$$\left( \left[ \frac{b^2}{\lambda_1(b)} \frac{8K_5}{K_1} \right]^{2/9} n^{2/9} \right),$$

pour lequel, on a

$$\text{MISE}(\tilde{f}_{n,m,b,\varepsilon}) = \frac{9}{8} (b\lambda_1^4(b))^{2/9} (8K_1^8 K_5)^{1/9} n^{-8/9} + o(n^{-8/9}).$$

On note que l'estimateur  $\tilde{f}_{n,m,b,\varepsilon}$  conserve la positivité mais ne constitue pas une densité. En fait, l'estimateur  $\tilde{f}_{n,m,b,\varepsilon}$  ne s'intègre généralement pas à l'unité. Pour résoudre ce problème, Igarashi and Kakizawa [42] ont proposé l'estimateur normalisé suivant

$$\forall x \in [0, 1], \quad \tilde{f}_{n,m,b,\varepsilon}^N(x) = \frac{\tilde{f}_{n,m,b,\varepsilon}(x)}{\int_0^1 \tilde{f}_{n,m,b,\varepsilon}(y) dy}. \quad (1.3.9)$$



La proposition suivante donne l'espérance, la variance et la *MISE* de l'estimateur  $\tilde{f}_{n,m,\varepsilon}^N$  (1.3.9).

**Proposition 1.3.9.** (*Igarashi and Kakizawa [42]*)

Supposons que la densité  $f$  soit continue, bornée et admettant quatre dérivées continues et bornées sur  $[0, 1]$ . Pour  $x \in [0, 1]$  tel que  $f(x) > 0$ , et avec  $m = O(n^\eta)$  et  $\varepsilon \approx m^\tau$  où  $\eta \in (0, 1)$  et  $\tau > 2$ , on a

$$\begin{aligned} \mathbb{E}[\tilde{f}_{n,m,b,\varepsilon}^N(x)] - f(x) &= -\frac{b}{m^2} \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) + f(x) \int_0^1 \frac{\Delta_1^2(y)}{2f(y)} dy \right\} \\ &+ O\left( \text{Var}[\tilde{f}_{n,m}(x)] + \text{Var}[\tilde{f}_{n,m/b}(x)] + n^{-1}m^{1/2} \right) \\ &+ o(m^{-2}). \end{aligned}$$

D'autre part, on a

$$\begin{aligned} \text{Var}[\tilde{f}_{n,m,b,\varepsilon}^N(x)] &= \text{Var}[\tilde{f}_{n,m,m/b}(x)] \\ &+ o\left( \text{Var}[\tilde{f}_{n,m}(x)] + \text{Var}[\tilde{f}_{n,m/b}(x)] + n^{-1}m^{1/2} + m^{-4} \right), \end{aligned}$$

Par ailleurs, on a

$$\text{MISE}(\tilde{f}_{n,m,b,\varepsilon}^N) = \lambda_1(b)K_1 \frac{m^{1/2}}{n} + \frac{b^2 K_6}{m^4} + o(m^{1/2}n) + o(m^{-4}),$$

L'ordre optimal pour minimiser la *MISE* de  $\tilde{f}_{n,m,a,\varepsilon}^N$  est égale à

$$\left( \left[ \frac{b^2}{\lambda_1(b)} \frac{8K_6}{K_1} \right]^{2/9} n^{2/9} \right),$$

pour lequel, on a

$$\text{MISE}(\tilde{f}_{n,m,b,\varepsilon}^N) = \frac{9}{8} (b\lambda_1^4(b))^{2/9} (8K_1^8 K_6)^{1/9} n^{-8/9} + o(n^{-8/9}).$$

Pour le choix pratique de l'ordre  $m$  le lecteur peut voir les travaux de Leblanc [55], Kakizawa [47, 48] et Igarashi and Kakizawa [42] qui ont utilisé la méthode de validation croisée.

Dans le troisième chapitre, on va comparer les estimateurs  $\tilde{f}_{n,m,m/2}$  (1.3.6),  $\tilde{f}_{n,m,m/b}$  (1.3.7),  $\tilde{f}_{n,m,b,\varepsilon}$  (1.3.8) et  $\tilde{f}_{n,m,b,\varepsilon}^N$  (1.3.9) avec notre estimateur récursif pour une densité.

# Chapitre 2

## Recursive distribution estimator using Bernstein polynomials

### 2.1 Introduction

There have been several methods for smooth estimation of density and distribution functions, from the introduction of kernel method of smoothing by Rosenblatt [88] and the advances made by Parzen [76]. For smoothing techniques for curve estimation, we refer the reader to Silverman [101] and Härdle [40]. However, these methods consider the support of the distribution as the whole real line.

When we have a random variable  $X$  with distribution  $F$  supported on a compact interval  $[a, b]$ ;  $a < b$ , it is simple to transform  $X$  into  $Y$ , a random variable with support  $[0, 1]$  through the transformation  $Y = (X - a)/(b - a)$ . Transformations such as  $Y = X/(1 + X)$  and  $Y = 1/2 + \pi^{-1} \arctan(X)$  can be used also to cover the cases of random variables  $X$  with support  $\mathbb{R}_+$  and  $\mathbb{R}$ , respectively. Once the random variable  $X$  has been transformed into  $Y$ ; we can apply Bernstein polynomials to approximate the distribution function of  $Y$  on the unit interval.

Let us now consider  $X_1, X_2, \dots, X_n$  a sequence of i.i.d random variables having a common unknown distribution  $F$  with associated density  $f$  supported on  $[0, 1]$ .

Using Robbins-Monro's (1951) recursive scheme, we set

$$(i) F_0(x) \in \mathbb{R},$$

$$(ii) \text{ for all } n \geq 1, F_n(x) = F_{n-1}(x) + \gamma_n W_n,$$

where  $(\gamma_n)$  is a nonrandom positive sequence tending to zero as  $n \rightarrow \infty$ , called the stepsize of this algorithm. To define  $W_n$ , we follow Vitale [119] (see also Babu et al. [2]) and introduce the Bernstein polynomial of order  $m > 0$  (assuming that  $m = m_n$  depends on  $n$ ),

$$b_k(m, x) = \binom{m}{k} x^k (1-x)^{m-k}$$

and set

$$W_n = \sum_{k=0}^m \mathbb{I} \left\{ X_n \leq \frac{k}{m} \right\} b_k(m, x) - F_{n-1}(x).$$

So the recursive estimator of the distribution  $F$  at the point  $x$  can be defined as

$$F_n(x) = (1 - \gamma_n)F_{n-1}(x) + \gamma_n \sum_{k=0}^m \mathbb{I} \left\{ X_n \leq \frac{k}{m} \right\} b_k(m, x). \quad (2.1.1)$$

Throughout this chapter, we suppose that  $F_0(x) = 0$  and we let  $\Pi_n = \prod_{j=1}^n (1 - \gamma_j)$  and  $Z_n(x) = \sum_{k=0}^m \mathbb{I} \left\{ X_n \leq \frac{k}{m} \right\} b_k(m, x)$ . Then, it follows from (2.1.1) that one can estimate  $F$  recursively at the point  $x$  by

$$F_n(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k Z_k(x). \quad (2.1.2)$$

The aim of this chapter is to study the properties of the recursive distribution estimator (2.1.2), as a competitor for Vitale's (1975) distribution estimator

$$\tilde{F}_n(x) = \sum_{k=0}^m \hat{F}_n \left( \frac{k}{m} \right) b_k(m, x), \quad (2.1.3)$$

where  $\hat{F}_n$  is the empirical distribution.

The Bernstein polynomial density and distribution estimators have been widely discussed in several frameworks. See, for instance, the original work of Vitale [119] and extensions given by Tenbusch [115], Babu et al. [2], Kakizawa [47], Rao [84] and Babu et Chaubey [3].

For our recursive estimator  $F_n$  defined by (2.1.2), we compute the bias and the variance. It turns out that they heavily depends on the choice of the stepsize  $(\gamma_n)$ . Moreover, we give the optimal order  $(m_n)$  which minimizes the mean integrated squared error ( $MISE$ ). Further, we show that using the stepsize  $(\gamma_n) = ([2/3+\varepsilon]n^{-1})$  with  $\varepsilon > 0$  (very close to zero) and the best implemented order  $(m_n)$ , the proposed estimator  $F_n$  can dominate Vitale's estimator  $\tilde{F}_n$  in terms of the  $MISE$ . Finally, we confirm our theoretical results by a simulation study.

The remainder of this chapter is organized as follows. In Section 2.2, we list our notations and assumptions. In Section 2.3, we state the main theoretical results regarding bias, variance,  $MISE$  and pointwise convergence of our recursive estimator introduced in (2.1.2). Section 2.4 is devoted to some numerical studies : first, a simulation study is presented in Subsection 2.4.1 and, then, an application to a real dataset is described in Subsection 2.4.2. Finally, we discuss our conclusion in Section 2.5 and detail the proofs of our theoretical results in Section 2.6.

## 2.2 Assumptions and Notations

**Definition 2.2.1.** *Let  $\gamma \in \mathbb{R}$  and  $(v_n)_{n \geq 1}$  be a nonrandom positive sequence.*

*We say that  $(v_n) \in \mathcal{GS}(\gamma)$  if*

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

This condition was introduced by Galambos et Seneta [30] to define regularly varying sequences (see also Bojanic and Seneta [7]). 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 obtain the behavior of the estimator defined in (2.1.2), we make the following assumptions :

(A1)  $F$  admits a continuous second derivative on  $[0, 1]$ .

(A2)  $(\gamma_n) \in \mathcal{GS}(-\alpha)$ ,  $\alpha \in (\frac{1}{2}, 1]$ .

(A3)  $(m_n) \in \mathcal{GS}(a)$ ,  $a \in (0, 1)$ .

(A4)  $\lim_{n \rightarrow \infty} (n\gamma_n) \in (\min(a, (2\alpha + a)/4), \infty)$ .

Assumption (A1) was standard for the bias and variance of the Bernstein polynomial distribution estimator (see Babu et al. [2] and Leblanc [56]).

Assumption (A2) on the stepsize was used in the recursive framework for density estimation (see Mokkadem et al. [66] and Slaoui [103, 104]).

Assumption (A3) on  $(m_n)$  is introduced similarly to the assumption on the bandwidth used for the recursive kernel distribution estimator (see Slaoui [105]).

Assumption (A4) on the limit of  $(n\gamma_n)$  as  $n$  goes to infinity is usual in the framework of stochastic approximation algorithms.

Throughout this chapter we will use the following notations :

$$\begin{aligned} \Pi_n &= \prod_{j=1}^n (1 - \gamma_j), & Z_n(x) &= \sum_{k=0}^m \mathbb{I} \left\{ X_n \leq \frac{k}{m} \right\} b_k(m, x), & \xi &= \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}, \\ b(x) &= \frac{x(1-x)f'(x)}{2}, & \sigma^2(x) &= F(x)[1-F(x)], & V(x) &= f(x) \left[ \frac{2x(1-x)}{\pi} \right]^{\frac{1}{2}}, \\ C_1 &= \int_0^1 \sigma^2(x) dx, & C_2 &= \int_0^1 V(x) dx, & C_3 &= \int_0^1 b^2(x) dx. \end{aligned}$$

## 2.3 Main Results

Our first result is the following proposition which gives the bias and the variance of  $F_n$ .

**Proposition 2.3.1** (Bias and variance of  $F_n$ ).

Let Assumptions (A1) – (A4) hold. We denote by  $o_x$  the pointwise bound in  $x$  (i.e., the error is not uniform in  $x \in [0, 1]$ ).

1. If  $0 < a \leq \frac{2}{3}\alpha$ , then

$$\mathbb{E}[F_n(x)] - F(x) = m_n^{-1} \frac{1}{1 - a\xi} b(x) + o(m_n^{-1}). \quad (2.3.1)$$

If  $\frac{2}{3}\alpha < a < 1$ , then

$$\mathbb{E}[F_n(x)] - F(x) = o\left(\sqrt{\gamma_n m_n^{-1/2}}\right). \quad (2.3.2)$$

2. If  $\frac{2}{3}\alpha \leq a < 1$ , then

$$\begin{aligned} \text{Var}[F_n(x)] &= \gamma_n \frac{1}{2 - \alpha\xi} \sigma^2(x) - \gamma_n m_n^{-1/2} \frac{2}{4 - (2\alpha + a)\xi} V(x) \\ &\quad + o_x(\gamma_n m_n^{-1/2}). \end{aligned} \quad (2.3.3)$$

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

$$\text{Var}[F_n(x)] = \gamma_n \frac{1}{2 - \alpha\xi} \sigma^2(x) + o_x(\gamma_n).$$

If  $0 < a < \frac{1}{2}\alpha$ , then

$$\text{Var}[F_n(x)] = o_x(m_n^{-2}). \quad (2.3.4)$$

3. If  $\lim_{n \rightarrow \infty} (n\gamma_n) > \max(a, (2\alpha + a)/4)$ , then (2.3.1) and (2.3.3) hold simultaneously.

The following proposition gives the *MISE* of the estimator  $F_n$ .

**Proposition 2.3.2** (MISE of  $F_n$ ).

Let Assumptions (A1)-(A4) hold.

1. If  $0 < a < \frac{1}{2}\alpha$ , then

$$MISE(F_n) = m_n^{-2} \frac{1}{(1 - a\xi)^2} C_3 + o(m_n^{-2}).$$

2. If  $\frac{1}{2}\alpha \leq a < \frac{2}{3}\alpha$ , then

$$MISE(F_n) = m_n^{-2} \frac{1}{(1 - a\xi)^2} C_3 + \gamma_n \frac{1}{2 - \alpha\xi} C_1 + o(m_n^{-2} + \gamma_n).$$

3. If  $a = \frac{2}{3}\alpha$ , then

$$\begin{aligned} MISE(F_n) &= m_n^{-2} \frac{1}{(1 - a\xi)^2} C_3 + \gamma_n \frac{1}{2 - \alpha\xi} C_1 - \gamma_n m_n^{-1/2} \frac{2}{4 - (2\alpha + a)\xi} C_2 \\ &\quad + o(m_n^{-2} + \gamma_n m_n^{-1/2}). \end{aligned}$$

4. If  $\frac{2}{3}\alpha < a < 1$ , then

$$MISE(F_n) = \gamma_n \frac{1}{2 - \alpha\xi} C_1 - \gamma_n m_n^{-1/2} \frac{2}{4 - (2\alpha + a)\xi} C_2 + o(\gamma_n m_n^{-1/2}).$$

The following result (Corollary 2.3.1) is a consequence of the previous proposition which gives the optimal order ( $m_n$ ) of the estimator  $F_n$  defined in (2.1.2) and the corresponding *MISE* in the case when  $(\gamma_n)$  is chosen in  $\mathcal{GS}(-1)$ ,  $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$  with  $\gamma_0 \in (2/3, \infty)$ .

**Corollary 2.3.1.**

Let Assumptions (A1)-(A4) hold. To minimize the *MISE* of  $F_n$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$ ,  $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$ ,  $(m_n)$  must be equal to

$$\left( 2^{2/3} (\gamma_0 - 2/3)^{-2/3} \left\{ \frac{4C_3}{C_2} \right\}^{2/3} n^{2/3} \right), \quad (2.3.5)$$

and then

$$\begin{aligned} MISE(F_n) &= n^{-1} \frac{\gamma_0^2}{2\gamma_0 - 1} C_1 - \frac{3}{4} \frac{1}{2^{4/3}} \frac{\gamma_0^2}{(\gamma_0 - 2/3)^{2/3}} \frac{C_2^{4/3}}{4^{1/3} C_3^{1/3}} n^{-4/3} \\ &\quad + o(n^{-4/3}). \end{aligned} \quad (2.3.6)$$

Now, from Theorems 1 and 3 in Leblanc [56] we recall that under Assumption (A1), the bias, variance and *MISE* of Vitale's distribution estimator (2.1.3) are given by

$$\begin{aligned}\mathbb{E} \left[ \tilde{F}_n(x) \right] - F(x) &= m_n^{-1}b(x) + o(m_n^{-1}), \\ \text{Var} \left[ \tilde{F}_n(x) \right] &= n^{-1}\sigma^2(x) - m_n^{-1/2}n^{-1}V(x) + o(m_n^{-1/2}n^{-1}), \\ \text{MISE} \left( \tilde{F}_n \right) &= n^{-1}C_1 - m_n^{-1/2}n^{-1}C_2 + m_n^{-2}C_3 + o(m_n^{-1/2}n^{-1}) + o(m_n^{-2}).\end{aligned}$$

Besides, to minimize the *MISE* of  $\tilde{F}_n$ ,  $(m_n)$  must be equal to

$$\left( \left\{ \frac{4C_3}{C_2} \right\}^{2/3} n^{2/3} \right), \quad (2.3.7)$$

and then

$$\text{MISE}(\tilde{F}_n) = n^{-1}C_1 - \frac{3}{4} \frac{C_2^{4/3}}{4^{1/3}C_3^{1/3}} n^{-4/3} + o(n^{-4/3}). \quad (2.3.8)$$

The following remark establishes that, for a special choice of the stepsize  $(\gamma_n)$ , the proposed distribution estimator  $F_n$  can dominate Vitale's estimator  $\tilde{F}_n$  in terms of the *MISE* performance.

**Remark 2.3.1.**

Under the assumptions (A1)-(A4), we consider the estimator  $F_n$  defined by (2.1.2) with  $(m_n)$  equal to (2.3.5) and the stepsize  $(\gamma_n) = (\gamma_0 n^{-1})$  such that  $\gamma_0 = \frac{2}{3} + \varepsilon$  with  $\varepsilon > 0$  (very close to zero) and the estimator  $\tilde{F}_n$  defined by (2.1.3) with  $(m_n)$  equal to (2.3.7). Then, ignoring the  $o(n^{-3/4})$  terms in (2.3.6) and (2.3.8), the condition for  $\text{MISE}(F_n) - \text{MISE}(\tilde{F}_n) \leq 0$ , ensures that

$$\begin{aligned}n &\leq \frac{3^3}{2^{10}} \frac{C_2^4}{4C_3C_1^3} \frac{(2\gamma_0 - 1)^3}{(\gamma_0 - 1)^6} (\gamma_0 - 2/3)^{-2} \left( \gamma_0 - 2^{2/3} (\gamma_0 - 2/3)^{1/3} \right)^3 \left( \gamma_0 + 2^{2/3} (\gamma_0 - 2/3)^{1/3} \right)^3 \\ &\simeq \frac{1}{2^4} \frac{C_2^4}{4C_3C_1^3} (1 + 33\varepsilon)\varepsilon^{-2}.\end{aligned}$$

Note that since the optimal *MISE* formulas (2.3.6) and (2.3.8), ignoring the  $o(n^{-3/4})$ -terms, must be positive, the lower bound of  $n$  is given by

$$\begin{aligned}n &\geq \left( \frac{3}{4} \right)^3 \frac{C_2^4}{4C_3C_1^3} \max \left\{ \frac{(2\gamma_0 - 1)^3}{2^4 (\gamma_0 - 2/3)^2}, 1 \right\} \\ &\simeq \left( \frac{3}{4} \right)^3 \frac{C_2^4}{4C_3C_1^3} \max \left\{ \frac{1}{2^4} \frac{1}{3^3} (1 + 18\varepsilon)\varepsilon^{-2}, 1 \right\}.\end{aligned}$$



Then,  $n_\varepsilon^{Lower} = \lfloor (\frac{3}{4})^3 \frac{C_2^4}{4C_3C_1^3} \max \{ \frac{1}{2^4} \frac{1}{3^3} (1 + 18\varepsilon)\varepsilon^{-2}, 1 \} \rfloor$  and  
 $n_\varepsilon^{Upper} = \lfloor \frac{1}{2^4} \frac{C_2^4}{4C_3C_1^3} (1 + 33\varepsilon)\varepsilon^{-2} \rfloor$ .

**Remark 2.3.2.**

Under the assumptions (A1)-(A4), it follows from (2.3.6) and (2.3.8), that for  $n$  large enough, the MISE of Vitale's estimator  $\tilde{F}_n$  is smaller than the MISE of our recursive estimator  $F_n$ . On the other hand, we can easily see from (2.3.6) and (2.3.8)

that  $\lim_{n \rightarrow \infty} \frac{MISE(\hat{F}_n)}{MISE(F_n)} = \frac{2\gamma_0 - 1}{\gamma_0^2} \leq 1$  (the equality holds if  $\gamma_0 = 1$ ).

Finally, we show in the following proposition the asymptotic normality of the recursive estimator (2.1.2).

**Théorème 2.3.1** (Asymptotic normality).

Let Assumption (A1)-(A4) hold.

1. If  $\gamma_n^{-1/2} m_n^{-1} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2} (F_n(x) - F(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( \frac{c}{1 - a\xi} b(x), \frac{1}{2 - a\xi} \sigma^2(x) \right).$$

2. If  $\gamma_n^{-1/2} m_n^{-1} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n (F_n(x) - F(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{b(x)}{1 - a\xi}, \tag{2.3.9}$$

where  $\xrightarrow[n \rightarrow +\infty]{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the Gaussian distribution and  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}}$  the convergence in probability.

## 2.4 Numerical studies

### 2.4.1 Simulations

The aim of this subsection is to compare Vitale's estimator (2.1.3) with the proposal (2.1.2), through a simulation study.

When applying (2.1.2) one needs to choose two quantities :

- The stepsize  $(\gamma_n) = (\gamma_0 n^{-1})$ , where  $\gamma_0 = 2/3 + c$ , with  $c \in (0, 1/3]$ . The extreme case  $(\gamma_n) = (n^{-1})$ , represents the case which ensures that asymptotically the recursive estimator (2.1.2) is equivalent to Vitale's estimator (2.1.3) in terms of the *MISE* (see Remark 2.3.2).
- The order  $(m_n)$  is chosen to be equal to (2.3.5).

When applying (2.1.3) one needs to choose the order  $(m_n)$  to be equal to (2.3.7).

In our simulation study, we consider six sample sizes,  $n = 30$ ,  $n = 50$ ,  $n = 100$  (small) and  $n = 200$ ,  $n = 500$ ,  $n = 950$  (large), and the following five distributions functions :

- a) the beta distribution  $\mathcal{B}(2, 1)$
- b) the beta distribution  $\mathcal{B}(10, 10)$
- c) the beta mixture distribution  $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$
- d) the truncated normal distribution  $\mathcal{N}_{[0,1]}(0.5, 0.25)$
- e) the truncated normal mixture  $0.25\mathcal{N}_{[0,1]}(5, 1) + 0.75\mathcal{N}_{[0,1]}(-5, 1)$ .

For each distribution function and sample size  $n$ , we approximate the average *ISE* of the estimator using  $N = 500$  trials of sample size  $n$ ;  $\overline{ISE} = \frac{1}{N} \sum_{k=1}^N ISE[\hat{g}^{[k]}]$ , where  $\hat{g}^{[k]}$  is the estimator computed from the  $k$ th sample, and  $ISE[\hat{g}] = \int_0^1 \{\hat{g}(x) - F(x)\}^2 dx$ .

From Tables 2.1 and 2.2, we conclude that :

- For the case  $c$ ) of the beta mixture, the average  $ISE$  of Vitale's distribution estimator (2.1.3) is smaller than that of the proposed recursive distribution estimator (2.1.2). However, in all the other cases, using an appropriate choice of the stepsize  $(\gamma_n)$ , the average  $ISE$  of the distribution estimator (2.1.2) is smaller than that of Vitale's estimator (2.1.3) even when the sample size is very large.
- The average  $ISE$  decreases as the sample size increases.
- For small sample sizes, we need to use the stepsize  $(\gamma_n) = ([2/3 + 0.05] n^{-1})$  and for large sample sizes, we need to use the stepsize  $(\gamma_n) = ([2/3 + 0.02] n^{-1})$ .

For qualitative comparison between the Vitale's distribution estimator  $\tilde{F}_n$  given in (2.1.3) and our proposed distribution estimator  $F_n$  (2.1.2). Figures 2.1, 2.2, 2.3 and 2.4 show, respectively, distribution estimates plotted for 500 simulated samples from  $\mathcal{B}(2, 1)$ ,  $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$ ,  $\mathcal{N}_{[0,1]}(0.5, 0.25)$  and  $0.25\mathcal{N}_{[0,1]}(5, 1) + 0.75\mathcal{N}_{[0,1]}(-5, 1)$  distribution of sizes  $n = 50$  (left panel) and  $n = 100$  (right panel).

From Figures 2.1, 2.2, 2.3 and 2.4, we conclude that :

- Our proposal (2.1.2) is closer to the true distribution function than Vitale's estimator (2.1.3).
- When the sample size increases, we get closer estimation of the true distribution.

## 2.4.2 Real dataset

The cross-validation procedure is a usual way of selecting the smoothing parameter. Sarda [93] proposed to use

$$CV(m) = \sum_{i=1}^n (\hat{F}_n(x_i) - F_{-i}(x_i))^2.$$

In this case, the  $CV$ -function measures the difference between the empirical distribution function  $\hat{F}_n$  and the leave-one-out version of our recursive distribution estimator,  $\hat{F}_{-i}$ , that uses the sample except for  $x_i$ .

We analyze the suicide data given in Table 2.1 of Silverman [101]. These data set consists of durations (in days) of psychiatric treatment for 86 patients used as controls in a study of suicide risks. It is clear in this setup that the distribution to be estimated is defined only for  $x > 0$ . For convenience, we also assume that the maximum treatment duration is 800 days (the data are such that  $\min(x_i) = 1$  and  $\max(x_i) = 737$ ) and analyse the original data rescaled to the unit interval.

We display the recursive estimator of the underlying distribution  $F$  of treatment durations along with the empirical distribution. We graph in the left panel of Figure 2.5, Vitale's distribution estimator (2.1.3) with the choice of  $m = 77$  obtained using the least-squares cross-validation method, and in the right panel, the proposed estimator  $F_n$  introduced in (2.1.2) using  $(\gamma_n)$  equal to  $([2/3 + 0.05]n^{-1})$  and  $m_k = \lfloor (16k)^{2/3} \rfloor$  (the data-driven optimal choice of  $m$  based on least-squares cross-validation for the distribution estimation problem).

## 2.5 Conclusion

In this chapter, we propose a recursive estimator of the cumulative distribution function based on a stochastic algorithm derived from the Robbins-Monro's scheme, using Bernstein polynomials. We first study its theoretical behavior. Then, we conduct a simulation study and analyse a real data application on suicide risks. For all the cases except for the case (c), the *ISE* of the our proposed estimator (2.1.2) with an appropriate choice of the stepsize  $(\gamma_n)$  and the order  $(m_n)$  is smaller than that of Vitale's estimator (2.1.3). In addition, a major advantage of our proposal is that its update, when new sample points are available, require less computational cost than Vitale's estimator.

In conclusion, using the proposed recursive estimator  $F_n$  we can obtain better results than those given by Vitale's distribution estimator. Hence, we plan to work on the automatic choice of the order  $(m_n)$  through a plug-in method and compare it with the work given in Slaoui [105].

## 2.6 Proofs

Let us first state the following technical lemma.

**Lemma 2.6.1.** (*Mokkadem et al. [66]*) *Let  $(v_n) \in \mathcal{GS}(v^*)$ ,  $(\gamma_n) \in \mathcal{GS}(-\alpha)$ , and  $l > 0$  such that  $l - v^*\xi > 0$ . We have*

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

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

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

We use this lemma repeatedly, throughout the proofs. Its application requires assumption (A4). Now, since  $\lim_{n \rightarrow \infty} (n\gamma_n) < \infty$  only if  $\alpha = 1$ , the condition  $\lim_{n \rightarrow \infty} (n\gamma_n) \in (\min(a, (2\alpha + a)/4), \infty]$  in (A4), which appears throughout our proofs, is equivalent to the condition  $\lim_{n \rightarrow \infty} (n\gamma_n) \in (\min(a, (2 + a)/4), \infty]$ . Similarly, since  $\xi \neq 0$  only if  $\alpha = 1$ , we can consider  $\alpha = 1$  in all the results given in this chapter.

### 2.6.1 Proof of Proposition 2.3.1

In view of (2.1.2), we have

$$\mathbb{E}[F_n(x)] - F(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k (\mathbb{E}[Z_k(x)] - F(x)). \quad (2.6.1)$$

It is easy to see that

$$\mathbb{E}[Z_n(x)] - F(x) = b(x)m_n^{-1} + o(m_n^{-1}) \quad (\text{see, e.g., Theorem 1 in Leblanc [56]}).$$

Substituting this result into (2.6.1) leads to

$$\mathbb{E}[F_n(x)] - F(x) = b(x) \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k \left[ m_k^{-1} (1 + o(1)) \right].$$

In the case  $0 < a \leq \frac{2}{3}\alpha$ , we have  $\sqrt{\gamma_n m_n^{-1/2}} = o(m_n^{-1})$ , since  $a \leq (2\alpha + a)/4$ .

Lemma 2.6.1 gives

$$\mathbb{E}[Z_n(x)] - F(x) = \frac{1}{1 - a\xi} b(x) m_n^{-1} + o(m_n^{-1}).$$

In the case  $\frac{2}{3}\alpha < a < 1$ , we have  $m_n^{-1} = o\left(\sqrt{\gamma_n m_n^{-1/2}}\right)$ , since  $a > (2\alpha + a)/4$ .

Lemma 2.6.1 gives

$$\mathbb{E}[F_n(x)] - F(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k o\left(\sqrt{\gamma_k m_k^{-1/2}}\right) = o\left(\sqrt{\gamma_n m_n^{-1/2}}\right).$$

Next, we consider

$$\text{Var}[F_n(x)] = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{Var}[Z_k(x)].$$

We know from Theorem 1 in Leblanc [56] that

$$\text{Var}[Z_n(x)] = \sigma^2(x) - V(x) m_n^{-1/2} + o_x(m_n^{-1/2}).$$

It follows that

$$\text{Var}[F_n(x)] = \sigma^2(x) \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 - V(x) \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k^{-1/2} (1 + o_x(1)).$$

In the case  $\frac{2}{3}\alpha \leq a < 1$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > (2\alpha + a)/4$ . Lemma 2.6.1 gives

$$\text{Var}[F_n(x)] = \gamma_n \frac{1}{2 - \alpha\xi} \sigma^2(x) - \gamma_n m_n^{-1/2} \frac{2}{4 - (2\alpha + a)\xi} V(x) + o_x(\gamma_n m_n^{-1/2}).$$

In the case  $\frac{1}{2}\alpha \leq a \leq \frac{2}{3}\alpha$ , we have  $\gamma_n m_n^{-1/2} = o(m_n^{-2})$  and  $\lim_{n \rightarrow \infty} (n\gamma_n) > \frac{1}{2}\alpha$ . Lemma 2.6.1 gives

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

In the case  $0 < a < \frac{1}{2}\alpha$ , we have  $\gamma_n = o(m_n^{-2})$ , and  $\lim_{n \rightarrow \infty} (n\gamma_n) > a$ . Lemma 2.6.1 gives

$$\text{Var}[F_n(x)] = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k o_x(m_k^{-2}) = o_x(m_n^{-2}).$$

## 2.6.2 Proof of Proposition 2.3.2

Using Proposition 2.3.1, we have

- In the case  $0 < a \leq \frac{2}{3}\alpha$ , we have

$$\begin{aligned} & \int_0^1 \left\{ \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k [\mathbb{E}(Z_k(x)) - F(x)] \right\}^2 dx \\ &= \int_0^1 \left[ m_n^{-1} \frac{1}{1-a\xi} b(x) + o(m_n^{-1}) \right]^2 dx \\ &= m_n^{-2} \frac{1}{(1-a\xi)^2} \int_0^1 b^2(x) dx + o(m_n^{-2}). \end{aligned} \quad (2.6.2)$$

- In the case  $\frac{2}{3}\alpha < a < 1$ , we have

$$\int_0^1 \left\{ \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k [\mathbb{E}(Z_k(x)) - F(x)] \right\}^2 dx = o(\gamma_n m_n^{-1/2}). \quad (2.6.3)$$

On the other hand, we note that

$$\int_0^1 \text{Var}[F_n(x)] dx = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \int_0^1 \text{Var}[Z_k(x)] dx.$$

We know (see the proof of Theorem 3 in Leblanc [56]) that

$$\int_0^1 \text{Var}[Z_n(x)] dx = \int_0^1 \sigma^2(x) dx - \int_0^1 V(x) dx m_n^{-1/2} (1 + o(1)).$$

It follows that

$$\begin{aligned} \int_0^1 \text{Var}[F_n(x)] dx &= \int_0^1 \sigma^2(x) dx \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \\ &\quad - \int_0^1 V(x) dx \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k^{-1/2} (1 + o(1)). \end{aligned}$$

The same arguments as in Subsection 2.6.1 for proving the convergence of the variance  $\text{Var}[F_n(x)]$ , together with (2.6.2) and (2.6.3), yield Proposition 2.3.2 (the details are omitted).

## 2.6.3 Proof of Corollary 2.3.1

Set

$$K_1(\xi) = \frac{C_1}{2 - \alpha\xi}, \quad K_2(\xi) = \frac{2C_2}{4 - (2\alpha + a)\xi}, \quad K_3(\xi) = \frac{C_3}{(1 - a\xi)^2}.$$

It follows from Proposition 2.3.2 that

$$MISE(F_n) = \begin{cases} K_3(\xi)m_n^{-2} [1 + o(1)] & \text{if } 0 < a < \frac{1}{2}\alpha, \\ K_1(\xi)\gamma_n + K_3(\xi)m_n^{-2} [1 + o(1)] & \text{if } \frac{1}{2}\alpha \leq a < \frac{2}{3}\alpha, \\ K_1(\xi)\gamma_n - m_n^{-2} [K_2(\xi)\gamma_n m_n^{3/2} - K_3(\xi) + o(1)] & \text{if } a = \frac{2}{3}\alpha, \\ K_1(\xi)\gamma_n - \gamma_n m_n^{-1/2} [K_2(\xi) + o(1)] & \text{if } \frac{2}{3}\alpha < a < 1. \end{cases} \quad (2.6.4)$$

By (A2) we have  $\frac{1}{2} < \alpha \leq 1$ . If  $0 < a < \frac{1}{2}\alpha$ ,  $(K_3(\xi)m_n^{-2}) \in \mathcal{GS}(-2a)$  with  $-2a > -\alpha$ , and in the other cases  $(MISE(F_n)) \in \mathcal{GS}(-\alpha)$ . Then the parameter  $a$  must be chosen as  $\frac{1}{2}\alpha \leq a < 1$ . Then we focus our comparison on

$(MISE(F_n)) \in \mathcal{GS}(-\alpha) - K_1(\xi)\gamma_n$  with  $\frac{1}{2}\alpha \leq a < 1$ . We have

$$MISE(F_n) - K_1(\xi)\gamma_n = \begin{cases} K_3(\xi)m_n^{-2} + o(m_n^{-2}) & \text{if } a \in [\frac{1}{2}\alpha, \frac{2}{3}\alpha[, \\ K_3(\xi)m_n^{-2} - K_2(\xi)\gamma_n m_n^{-1/2} + o(m_n^{-2}) & \text{if } a = \frac{2}{3}\alpha, \\ -K_2(\xi)\gamma_n m_n^{-1/2} + o(\gamma_n m_n^{-1/2}) & \text{if } a \in ]\frac{2}{3}\alpha, 1[. \end{cases}$$

If  $a = \frac{2}{3}\alpha$ ,  $(K_3(\xi)m_n^{-2} - K_2(\xi)\gamma_n m_n^{-1/2}) \in \mathcal{GS}(-\frac{4}{3}\alpha)$  with  $\frac{4}{3}\alpha = 2a$ .

If  $\frac{1}{2}\alpha \leq a < \frac{2}{3}\alpha$ ,  $(K_3(\xi)m_n^{-2}) \in \mathcal{GS}(-2a)$  with  $-2a > -\frac{4}{3}\alpha$ , and, if  $\frac{2}{3}\alpha < a < 1$ ,  $(K_2(\xi)\gamma_n m_n^{-1/2}) \in \mathcal{GS}(-(2\alpha + a)/2)$  with  $-(2\alpha + a)/2 > -2a$ . It follows that, for a given  $\alpha$ , to minimize the  $MISE(F_n) - K_1(\xi)\gamma_n$ , the parameter  $a$  must be chosen equal to  $\frac{2}{3}\alpha$ . Moreover, in view of (2.6.4) the parameter  $\alpha$  must be equal to 1. We conclude that to minimize the  $MISE(F_n)$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$  and the order  $(m_n)$  in  $\mathcal{GS}(2/3)$ .

Since the function  $x \mapsto K_3(\xi)x^{-2} - \gamma_n K_2(\xi)x^{-1/2}$  attains its minimum at the point  $\left[\frac{4K_3(\xi)}{\gamma_n K_2(\xi)}\right]^{2/3}$ , to minimize  $MISE(F_n) - K_1(\xi)\gamma_n$ , the order  $(m_n)$  must be equal to  $\left(\frac{4K_3(\xi)}{K_2(\xi)}\gamma_n^{-1}\right)^{2/3}$ . For such a choice, the  $MISE$  of  $F_n$  becomes

$$MISE(F_n) = \gamma_n K_1(\xi) - \frac{3}{4} \frac{1}{2^{2/3}} \gamma_n^{4/3} K_2(\xi)^{4/3} K_3(\xi)^{-1/3} [1 + o(1)].$$

Now, we assume that  $(\gamma_n) = (\gamma_0 n^{-1})$  for some  $\gamma_0 \in (2/3, \infty)$ . In this case, note that  $\xi = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1} = \gamma_0^{-1}$ , where  $\gamma_0 > 2/3$  ensures that  $K_2(\xi) > 0$ . Then, (2.3.5) (hence (2.3.6)) is a consequence of the standard trade-off argument.



## 2.6.4 Proof of Theorem 2.3.1

To prove Theorem 2.3.1, we will use the fact that if  $\frac{1}{2}\alpha \leq a < 1$ , then

$$\gamma_n^{-1/2}(F_n(x) - \mathbb{E}[F_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{2 - \alpha\xi}\sigma^2(x)\right), \quad (2.6.5)$$

which will be proved later. In the case  $\frac{1}{2}\alpha \leq a \leq \frac{2}{3}\alpha$ , we have

$$\begin{aligned} \gamma_n^{-1/2}(F_n(x) - F(x)) &= \gamma_n^{-\frac{1}{2}}(F_n(x) - \mathbb{E}[F_n(x)]) + \gamma_n^{-1/2}(\mathbb{E}[F_n(x)] - F(x)) \\ &= \gamma_n^{-1/2}(F_n(x) - \mathbb{E}[F_n(x)]) + \gamma_n^{-1/2}m_n^{-1}\frac{1}{1 - \alpha\xi}[b(x) + o(1)], \end{aligned}$$

so if  $\gamma_n^{-1/2}m_n^{-1} \xrightarrow[n \rightarrow +\infty]{} c$ , for some  $c > 0$ , Part 1 of Theorem 2.3.1 follows from (2.6.5).

In the case  $\frac{2}{3}\alpha < a < 1$  (note that  $\gamma_n^{-1/2}m_n^{-1} \xrightarrow[n \rightarrow +\infty]{} 0$ ), we have, using (2.3.2),

$$\gamma_n^{-1/2}(F_n(x) - F(x)) = \gamma_n^{-\frac{1}{2}}(F_n(x) - \mathbb{E}[F_n(x)]) + o\left(\sqrt{m_n^{-1/2}}\right),$$

where the first term is asymptotically normal (use (2.6.5)) and the second term is  $o(1)$ . In the case  $0 < a < \frac{1}{2}\alpha$ , (2.3.4) implies that  $m_n(F_n(x) - \mathbb{E}[F_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$ , and an application of (2.3.1) gives Part 2 of Theorem 2.3.1. On the other hand, (2.3.9) follows from (2.6.5) and

$$\begin{aligned} m_n(F_n(x) - F(x)) &= m_n(F_n(x) - \mathbb{E}[F_n(x)]) + m_n(\mathbb{E}[F_n(x)] - F(x)) \\ &= [m_n\gamma_n^{1/2}] \gamma_n^{-1/2}(F_n(x) - \mathbb{E}[F_n(x)]) + \frac{b(x)}{1 - \alpha\xi}[1 + o(1)], \end{aligned}$$

since, in this case,  $m_n\gamma_n^{1/2} \xrightarrow[n \rightarrow +\infty]{} 0$ .

We now prove (2.6.5). We have

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

We set

$$Y_k(x) = \Pi_k^{-1} \gamma_k (Z_k(x) - \mathbb{E}[Z_k(x)]).$$

The application of Lemma 2.6.1 gives

$$v_n^2 = \sum_{k=1}^n \text{Var}[Y_k(x)] = \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{var}[Z_k(x)] = \frac{\gamma_n}{\Pi_n^2} \left[ \frac{1}{2 - \alpha\xi} \sigma^2(x) + o(1) \right].$$

On the other hand, for all  $p > 0$ , we have  $\mathbb{E}[|Z_k(x)|^{2+p}] = O(1)$ , and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > \frac{1}{2}\alpha$ , there exists a  $p > 0$  such that  $\lim_{n \rightarrow \infty} (n\gamma_n) > \frac{1+p}{2+p}\alpha$ , so Lemma 2.6.1 gives

$$\sum_{k=1}^n \mathbb{E}[|Y_k(x)|^{2+p}] = O\left(\sum_{k=1}^n \Pi_k^{-2-p} \gamma_k^{2+p} \mathbb{E}[|Z_k(x)|^{2+p}]\right) = O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}}\right)$$

hence

$$\frac{1}{v_n^{2+p}} \sum_{k=1}^n \mathbb{E}[|Y_k(x)|^{2+p}] = O(\gamma_n^{p/2}) = o(1).$$

Then the convergence in (2.6.5) follows from the application of Lyapounov's theorem.

Case	$[n_{0.05}^{Lower}, n_{0.05}^{Upper}]$	$n$	Vitale's estimator	Recursive estimator			
				$(\gamma_n) = ([2/3 + \varepsilon]n^{-1})$			$(\gamma_n) = (n^{-1})$
				$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.2$	
a)	[68, 2020]	30	0.001092	<b>0.000234</b>	0.001269	0.001630	0.001677
		50	0.000978	<b>0.000613</b>	0.001102	0.001253	0.001266
		100	0.000660	<b>0.000569</b>	0.000740	0.000779	0.000777
b)	[41, 1197]	30	0.000750	<b>0.000474</b>	0.000853	0.000969	0.000980
		50	0.000573	<b>0.000469</b>	0.000645	0.000689	0.000689
		100	0.000355	<b>0.000337</b>	0.000396	0.000404	0.000401
c)	[4, 92]	30	<b>0.004486</b>	0.004794	0.004976	0.004847	0.004775
		50	<b>0.002847</b>	0.003109	0.003154	0.003044	0.002993
		100	<b>0.001509</b>	0.001684	0.001670	0.001597	0.001567
d)	[59, 1744]	30	0.001190	<b>0.000447</b>	0.001361	0.001671	0.001707
		50	0.000992	<b>0.000693</b>	0.001123	0.001249	0.001258
		100	0.000653	<b>0.000582</b>	0.000731	0.000761	0.000758
e)	[29, 861]	30	0.001489	<b>0.001146</b>	0.001680	0.001824	0.001831
		50	0.001079	<b>0.000964</b>	0.001206	0.001254	0.001250
		100	0.000641	<b>0.000635</b>	0.000714	0.000716	0.000709

TABLE 2.1 – Average *ISE* (over  $N = 500$  trials) of Vitale's estimator  $\tilde{F}_n$  and the proposal  $F_n$  for the case  $(\gamma_n) = ([2/3 + \varepsilon]n^{-1})$ , with  $\varepsilon = 0.05, 0.1, 0.2$ , and  $(\gamma_n) = (n^{-1})$ .

Case	$[n_{0.02}^{Lower}, n_{0.02}^{Upper}]$	$n$	Vitale's estimator	Recursive estimator			
				$(\gamma_n) = ([2/3 + \varepsilon]n^{-1})$			$(\gamma_n) = (n^{-1})$
				$\varepsilon = 0.02$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	
a)	[304, 7909]	200	0.000399	<b>0.000163</b>	0.000389	0.000445	0.000446
		500	0.000187	<b>0.000136</b>	0.000197	0.000208	0.000201
		950	0.000106	<b>0.000092</b>	0.000116	0.000118	0.000112
b)	[183, 4747]	200	0.000204	<b>0.000125</b>	0.000209	0.000227	0.000223
		500	0.000092	<b>0.000078</b>	0.000100	0.000103	0.000098
		950	0.000052	<b>0.000049</b>	0.000057	0.000057	0.000054
c)	[14, 360]	200	<b>0.000789</b>	0.000828	0.000893	0.000872	0.000812
		500	<b>0.000329</b>	0.000366	0.000378	0.000363	0.000336
		950	<b>0.000177</b>	0.000202	0.000205	0.000195	0.000180
d)	[263, 6830]	200	0.0000389	<b>0.000184</b>	0.000385	0.000433	0.000430
		500	0.0000180	<b>0.000138</b>	0.000191	0.000200	0.000193
		950	0.0000102	<b>0.000090</b>	0.000111	0.000113	0.000107
e)	[130, 3373]	200	0.000361	<b>0.000256</b>	0.000378	0.000400	0.000388
		500	0.000160	<b>0.000144</b>	0.000176	0.000178	0.000168
		950	0.000089	<b>0.000087</b>	0.000099	0.000098	0.000092

TABLE 2.2 – Average  $ISE$  (over  $N = 500$  trials) of Vitale's estimator  $\tilde{F}_n$  and the proposal  $F_n$  for the case  $(\gamma_n) = ([2/3 + \varepsilon]n^{-1})$ , with  $\varepsilon = 0.02, 0.05, 0.1$ , and  $(\gamma_n) = (n^{-1})$ .

FIGURE 2.1 – Qualitative comparison between the estimator  $\tilde{F}_n$  defined in (2.1.3) (dashed blue line) and the proposed distribution estimator (2.1.2) with stepsize  $(\gamma_n) = ([2/3 + 0.05]n^{-1})$  (solid red line), for 500 samples respectively of size 50 (left panel) and of size 100 (right panel) for the beta distribution  $\mathcal{B}(2, 1)$ .

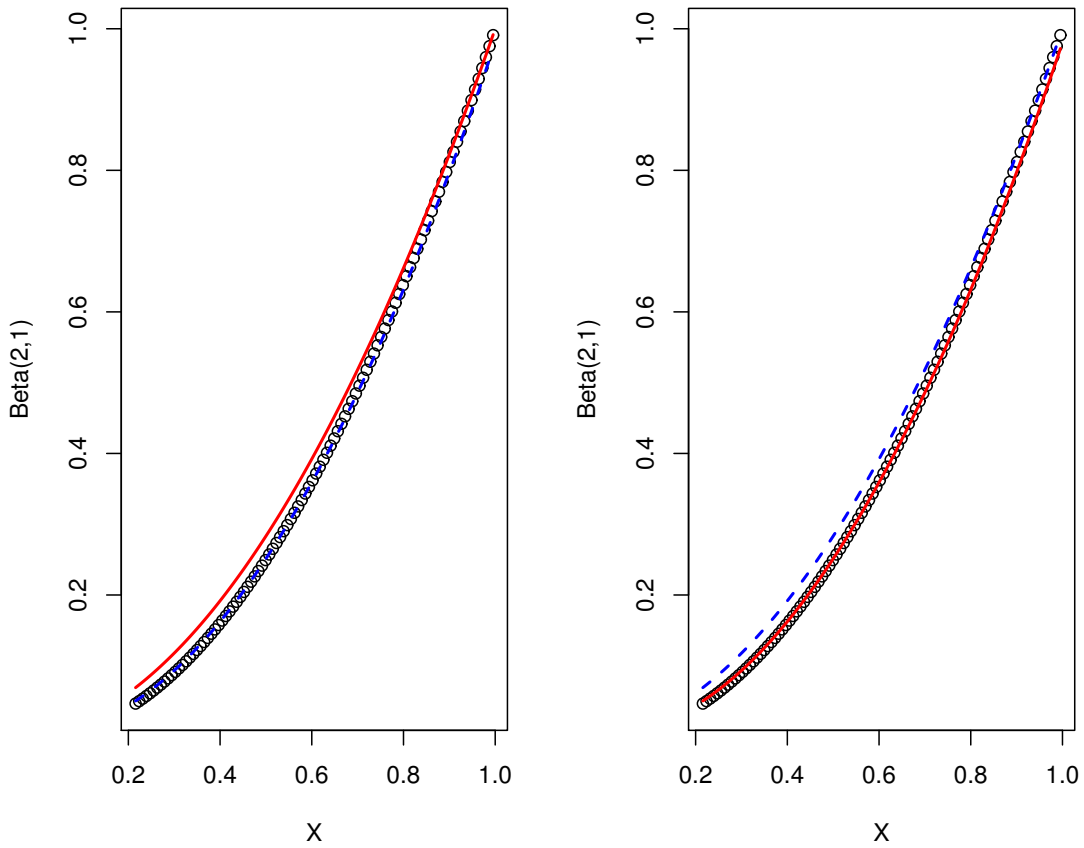


FIGURE 2.2 – Qualitative comparison between the estimator  $\tilde{F}_n$  defined in (2.1.3) (dashed blue line) and the proposed distribution estimator (2.1.2) with stepsize  $(\gamma_n) = ([2/3 + 0.02]n^{-1})$  (solid red line), for 500 samples respectively of size 50 (left panel) and of size 100 (right panel) for the beta mixture distribution  $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$ .

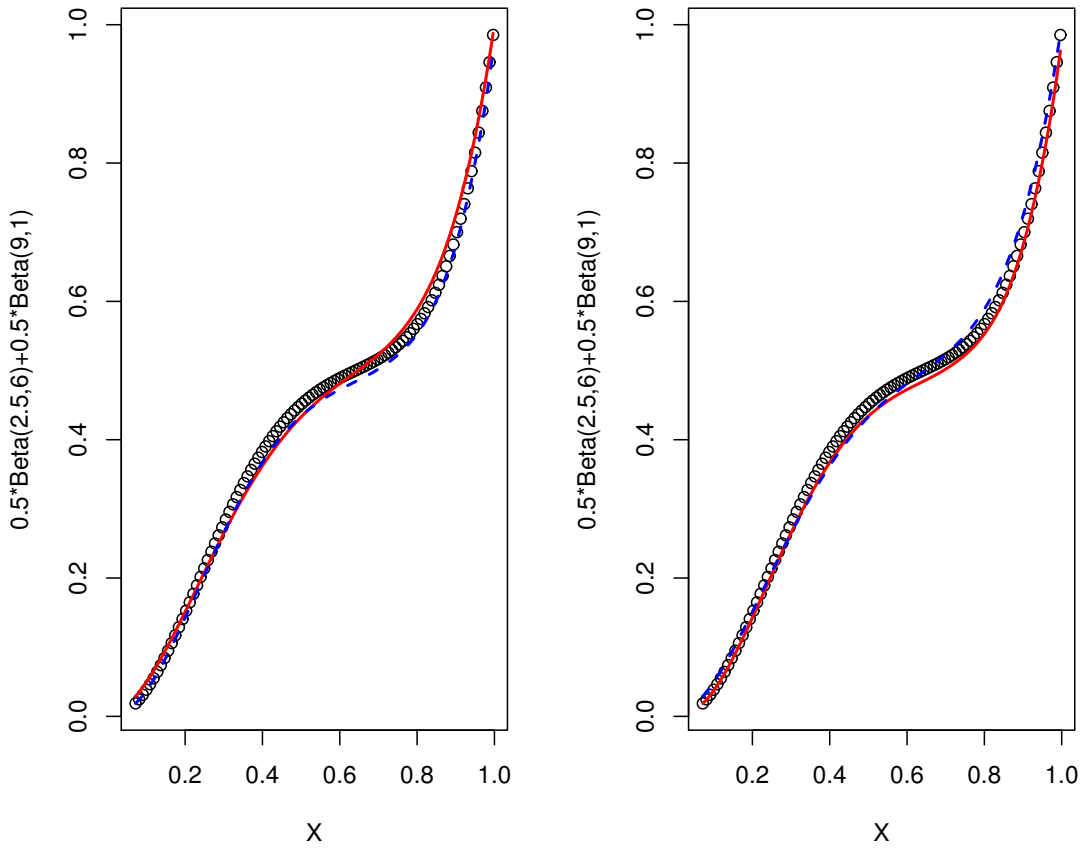


FIGURE 2.3 – Qualitative comparison between the estimator  $\tilde{F}_n$  defined in (2.1.3) (dashed blue line) and the proposed distribution estimator (2.1.2) with stepsize  $(\gamma_n) = ([2/3 + 0.05]n^{-1})$  (solid red line), for 500 samples respectively of size 50 (left panel) and of size 100 (right panel) for the truncated normal distribution  $\mathcal{N}_{[0,1]}(0.5, 0.25)$ .

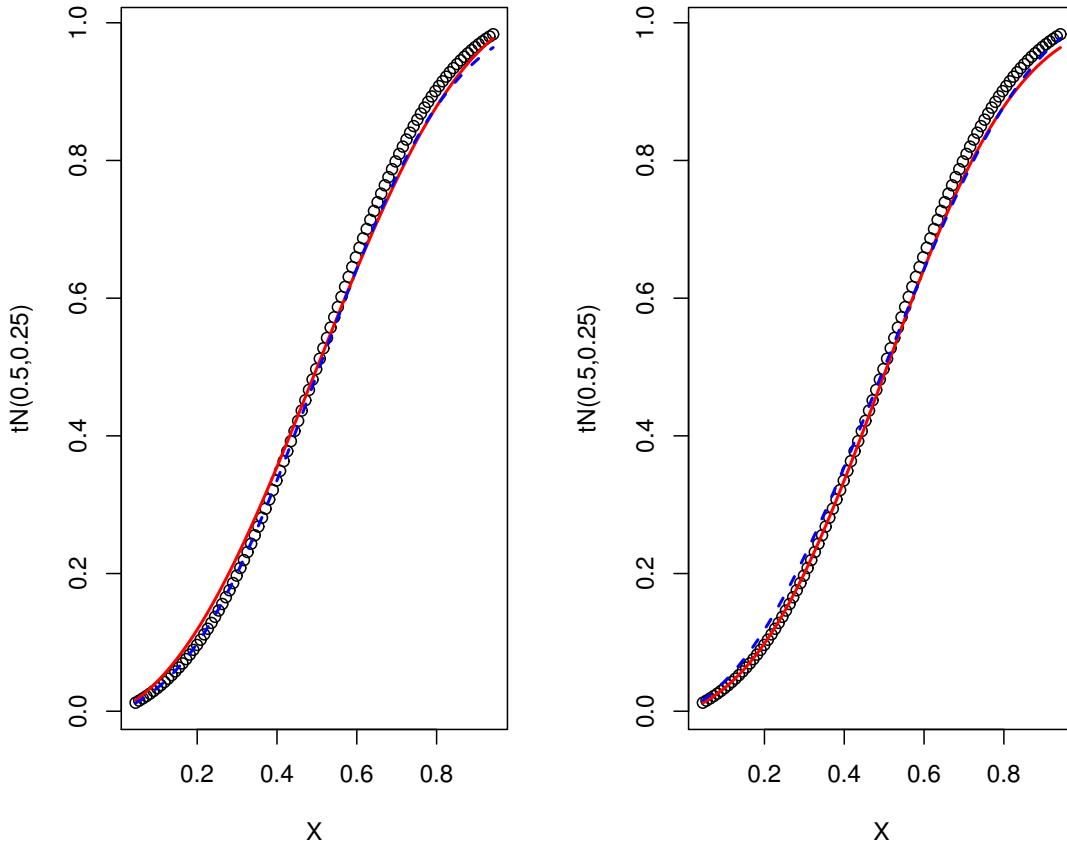


FIGURE 2.4 – Qualitative comparison between the estimator  $\tilde{F}_n$  defined in (2.1.3) (dashed red line) and the proposed distribution estimator (2.1.2) with stepsize  $(\gamma_n) = ([2/3 + 0.05]n^{-1})$  (solid blue line), for 500 samples respectively of size 50 (left panel) and of size 100 (right panel) for the truncated normal mixture distribution  $0.25\mathcal{N}_{[0,1]}(5, 1) + 0.75\mathcal{N}_{[0,1]}(-5, 1)$ .

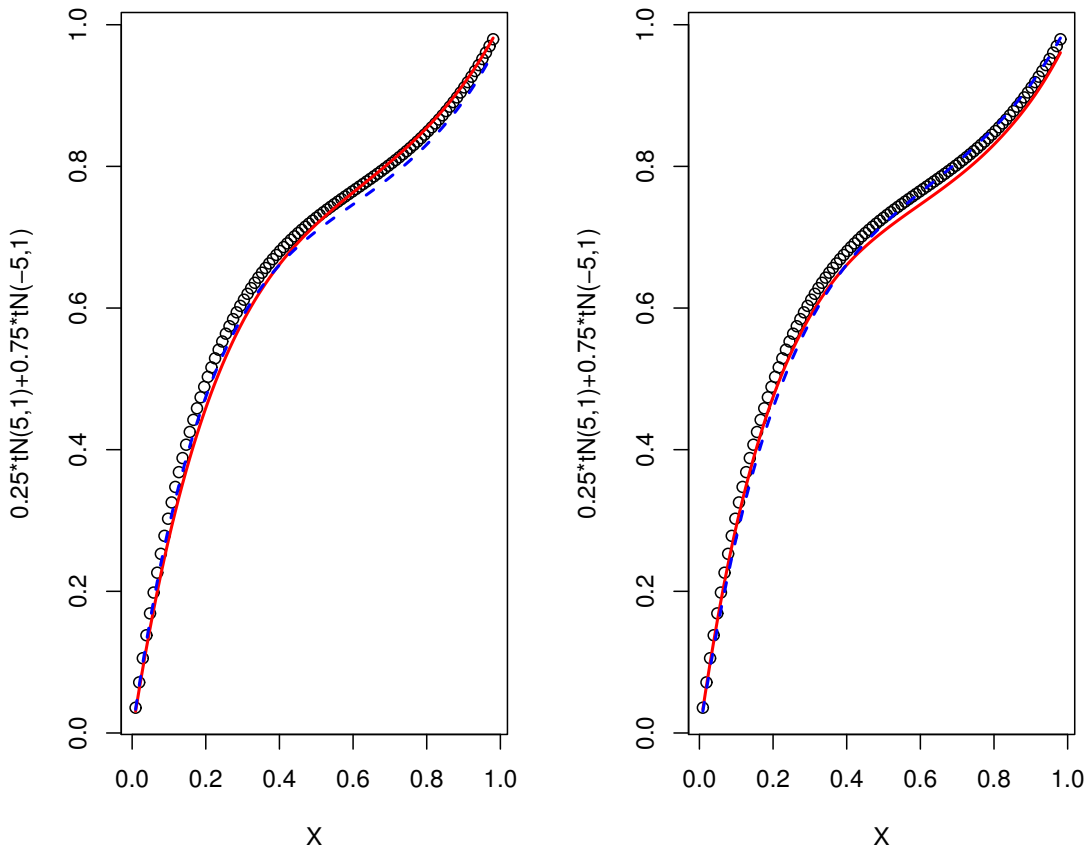
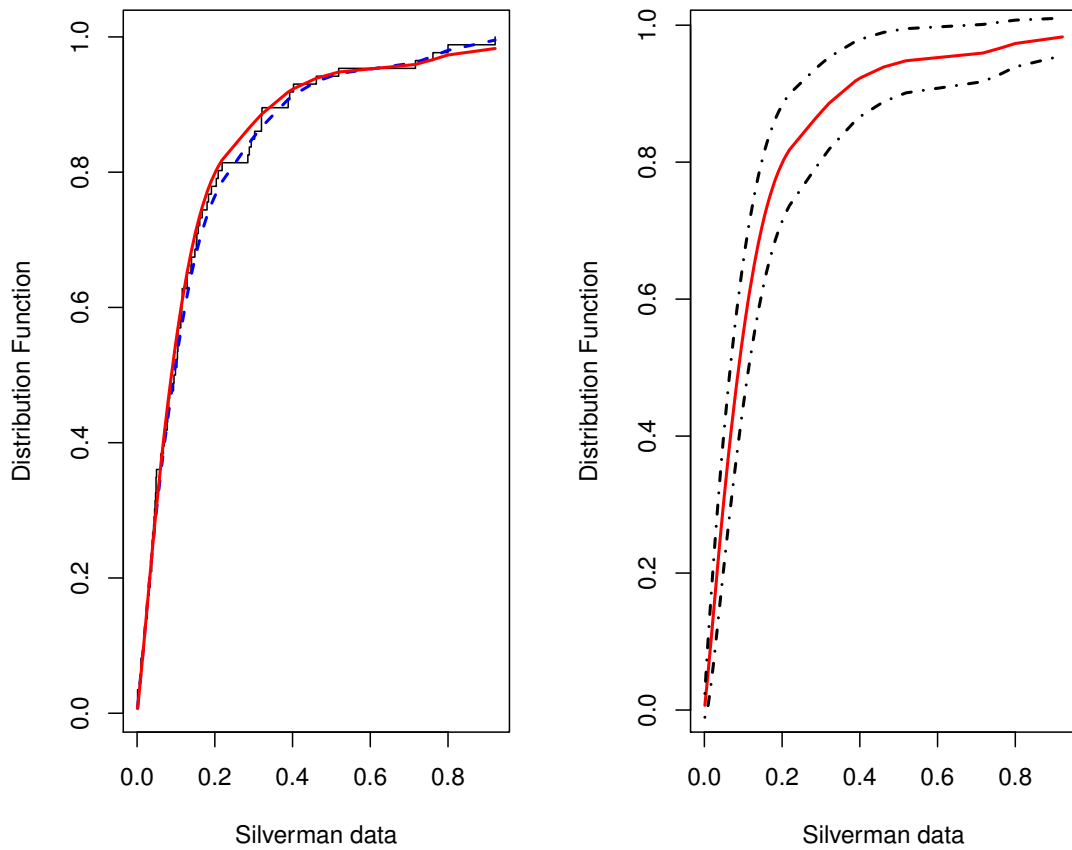




FIGURE 2.5 – The left panel shows a qualitative comparison between the empirical distribution function, Vitale’s distribution estimator (2.1.3) (dashed red line) and our proposed recursive estimator  $F_n$  (2.1.2) (solid blue line). The right panel shows 95% pointwise confidence intervals of  $F_n$  using the stepsize  $(\gamma_n) = ([2/3 + 0.05]n^{-1})$  and the suicide data given in Silverman [101].



# Chapitre 3

## Recursive density estimator using Bernstein polynomials

### 3.1 Introduction

There has been a considerable development of methods for smooth estimation of density and distribution functions, following the introduction of several kernel smoothing by Rosenblatt [88] and the further advances made on kernel method by Parzen [76]. The reader is referred to an excellent text by Härdle [40] for several kernel smoothing techniques. However, these methods have difficulties at and near boundaries when curve estimation is attempted over a region with boundaries. Moreover, it is well known in nonparametric kernel density estimation that the bias of the standard kernel density estimator

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

is of a larger order near the boundary than that in the interior. 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, so we are dealing with positive data.

For convenience, take  $K$  to be a symmetric probability density (for instance, normal kernels). Away from the boundary, which thus means at any  $x > h$ , the usual asymptotic mean and variance expressions apply. Suppose  $f$  has two continuous derivatives everywhere, and that as  $n \rightarrow \infty$ ,  $h = h_n \rightarrow 0$  and  $nh \rightarrow 0$ . Then,

$$\mathbb{E} \left[ \widehat{f}(x) \right] \simeq f(x) + \frac{1}{2} h^2 f''(x) \int x^2 K(x) dx,$$

and

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

Near the boundary, things are different. Write  $x = ph$ , we have

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

and

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

This bias phenomena is called boundary bias.

Many authors have suggested methods for reducing this phenomena such as data reflection (Schuster [98]), boundary kernels (Müller [68], Müller et Prewitt [69] and Müller et Wand [70]), the local linear estimator ([58] and [45]), the use of beta and gamma kernels (Chen [14, 15]).

For smooth estimate of a density function with finite known support, Vitale's method (Vitale [119]) based on the Bernstein polynomials, illustrated bellow, also has been investigated in the literature (Ghosa [33], Babu et al. [2], Kakizawa [47], Rao [84]) and more recently by Leblanc [55] and Igarashi and Kakizawa [42].

The idea comes from the Weierstrass's approximation theorem that for any continuous function  $u$  on the interval  $[0, 1]$ , we have

$$\sum_{k=0}^m u \left( \frac{k}{m} \right) b_k(m, x) \rightarrow u(x),$$

uniformly in  $x \in [0, 1]$ , where  $b_k(m, x) = \binom{m}{k} x^k (1-x)^{m-k}$  is the Bernstein polynomial of order  $m$ .

In the context of distribution function  $F$  with support  $[0, 1]$ , Vitale [119] proposed an estimator

$$\tilde{F}_n(x) = \sum_{k=0}^m F_n\left(\frac{k}{m}\right) b_k(m, x),$$

where  $F_n$  is the empirical distribution based on a random sample  $X_1, X_2, \dots, X_n$ .

Hence, an estimator for the density  $f$  is given by

$$\begin{aligned} \tilde{f}_n(x) &= \frac{d}{dx} \tilde{F}_n(x) \\ &= m \sum_{k=0}^{m-1} \left[ F_n\left(\frac{k+1}{m}\right) - F_n\left(\frac{k}{m}\right) \right] b_k(m-1, x). \end{aligned} \quad (3.1.1)$$

In this chapter, we propose a recursive method to estimate an unknown density function  $f$ . The advantage of recursive estimators is that their update from a sample of size  $n$  to one of size  $n+1$ , requires considerably less computations. This property is particularly important, since the number of points at which the function is estimated is usually very large.

Let  $X_1, X_2, \dots, X_n$  be a sequence of i.i.d random variables having a common unknown distribution  $F$  with associated density  $f$  supported on  $[0, 1]$ .

Using Robbins-Monro's scheme (Robbins et Monro [87]), we set

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

$$(ii) \quad f_n(x) = (1 - \gamma_n)f_{n-1}(x) + \gamma_n Z_n \text{ for all } n \in \mathbb{N}^*,$$

where  $(\gamma_n)$  is a sequence of real numbers, called a stepsize. We follow Leblanc [55] to define  $Z_n(x) = 2T_{n,m}(x) - T_{n,m/2}(x)$ , where

$$\begin{aligned} T_{n,m}(x) &= m \sum_{k=0}^{m-1} \left( \mathbb{I}_{\{X_n \leq \frac{k+1}{m}\}} - \mathbb{I}_{\{X_n \leq \frac{k}{m}\}} \right) b_k(m-1, x) \\ &= m \sum_{k=0}^{m-1} \mathbb{I} \left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\} b_k(m-1, x) \\ &= m b_{k_n}(m-1, x). \end{aligned}$$

Here, we write  $k_n = [mX_n]$ , where  $[x]$  denotes the largest integer smaller than  $x$ .

We suppose that  $f_0(x) = 0$  and we let  $\Pi_n = \prod_{j=1}^n (1 - \gamma_j)$ .

Then, one can estimate  $f$  recursively at the point  $x$  by

$$f_n(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k Z_k(x). \quad (3.1.2)$$

Our first aim in this chapter, is to compute the bias, the variance, the mean squared error ( $MSE$ ) and the mean integrated squared error ( $MISE$ ) of our proposed recursive estimators. It turns out that they heavily depends on the choice of the stepsize  $(\gamma_n)$ . Moreover, we give the optimal order  $(m_n)$  which minimizes the  $MSE$  and the  $MISE$  of the proposed recursive estimators. Further, we show that using the stepsize  $(\gamma_n) = (n^{-1})$  and the optimal order  $(m_n)$ , the proposed estimator  $f_n$  can dominate Vitale's estimator  $\tilde{f}_n$  in terms of  $MISE$ . Finally, we confirm our theoretical results by a simulation study.

The remainder of this chapter is organized as follows. In Section 3.2, we list our notations and assumptions. In Section 3.3, we state the main theoretical results regarding bias, variance,  $MSE$  and  $MISE$ . Section 3.4 is devoted to some numerical studies : first, a simulation study is presented in Subsection 3.4.1 and, then, an application to a real dataset is described in Subsection 3.4.2. Finally, we discuss our conclusion in Section 3.5. The proofs are given in the Section 3.6.

## 3.2 Assumptions and Notations

**Definition 3.2.1.** *Let  $\gamma \in \mathbb{R}$  and  $(v_n)_{n \geq 1}$  be a nonrandom positive sequence.*

*We say that  $(v_n) \in \mathcal{GS}(\gamma)$  if*

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

This condition was introduced by Galambos et Seneta [30] to define regularly varying sequences (see also Bojanic and Seneta [7]). 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 study the estimator (3.1.2), we make the following assumptions :

(A1)  $f$  admits a continuous fourth-order derivative  $f^{(4)}$  on  $[0, 1]$ .

(A2)  $(\gamma_n) \in \mathcal{GS}(-\alpha)$ ,  $\alpha \in (\frac{1}{2}, 1]$ .

(A3)  $(m_n) \in \mathcal{GS}(a)$ ,  $a \in (0, 1)$ .

(A4)  $\lim_{n \rightarrow \infty} (n\gamma_n) \in (\min(2a, (2\alpha - a)/4), \infty)$ .

Assumption (A1) is standard (see Leblanc [55]). Assumption (A2) on the stepsize was used in the recursive framework for density estimation (see Mokkadem et al. [66] and Slaoui [103, 104]). Assumption (A3) on  $(m_n)$  was introduced similarly to the assumption on the bandwidth used for the recursive kernel distribution estimator (see Slaoui [105]). Assumption (A4) on the limit of  $(n\gamma_n)$  as  $n$  goes to infinity is usual in the framework of stochastic approximation algorithms.

Throughout this chapter we will use the following notations :

$$\begin{aligned} \Delta_1(x) &= \frac{1}{2} [(1 - 2x)f'(x) + x(1 - x)f''(x)], \quad \psi(x) = (4\pi x(1 - x))^{-1/2}, \quad \xi = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}, \\ \Delta_2(x) &= \frac{1}{6}(1 - 6x(1 - x))f''(x) + \frac{5}{12}x(1 - x)(1 - 2x)f^{(3)}(x) + \frac{1}{8}x^2(1 - x)^2f^{(4)}(x), \\ C_1 &= \int_0^1 f(x)\psi(x)(x)dx, \quad C_2 = \int_0^1 \Delta_2^2(x)dx, \quad C_3 = \frac{1}{\sqrt{2}} + 4 \left(1 - \sqrt{\frac{2}{3}}\right), \\ C_4 &= \int_0^1 \Delta_1^2(x)dx, \quad C_5 = \int_0^1 \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) \right\}^2 dx, \\ C_6 &= \int_0^1 \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) + f(x) \int_0^1 \frac{\Delta_1^2(y)}{2f(y)} dy \right\}^2 dx. \end{aligned}$$

### 3.3 Main Results

We begin by studying the bias, the variance and the  $MSE$  of the proposed recursive estimators using Bernstein estimator first in the boundary region and then at the edges.

#### 3.3.1 Within the interval $[0, 1]$

The following proposition gives the bias, variance, and  $MSE$  of the proposed recursive estimator  $f_n(x)$  for  $x \in (0, 1)$ .

**Proposition 3.3.1.**

Let Assumptions (A1) – (A4) hold. For  $x \in (0, 1)$ , we have

1. If  $0 < a \leq \frac{2}{9}\alpha$ , then

$$\mathbb{E}[f_n(x)] - f(x) = -m_n^{-2} \frac{2}{1 - 2a\xi} \Delta_2(x) + o(m_n^{-2}). \quad (3.3.1)$$

- If  $\frac{2}{9}\alpha < a < 1$ , then

$$\mathbb{E}[f_n(x)] - f(x) = o\left(\sqrt{\gamma_n m_n^{1/2}}\right). \quad (3.3.2)$$

2. If  $\frac{2}{9}\alpha \leq a < 1$ , then

$$\text{Var}[f_n(x)] = C_3 \gamma_n m_n^{1/2} \frac{2}{4 - (2\alpha - a)\xi} f(x)\psi(x) + o(\gamma_n m_n^{1/2}). \quad (3.3.3)$$

- If  $0 < a < \frac{2}{9}\alpha$ , then

$$\text{Var}[f_n(x)] = o(m_n^{-4}). \quad (3.3.4)$$

3. If  $\lim_{n \rightarrow \infty} (n\gamma_n) > \max(2a, (2\alpha - a)/4)$ , that implies  $\lim_{n \rightarrow \infty} (n\gamma_n) > 2a$  and  $\lim_{n \rightarrow \infty} (n\gamma_n) > (2\alpha - a)/4$  so  $1 - 2a\xi > 0$  and  $4 - (2\alpha - a)\xi > 0$  then (3.3.1) and (3.3.3) hold simultaneously.

4. If  $0 < a < \frac{2}{9}\alpha$ , then

$$MSE [f_n(x)] = \Delta_2^2(x)m_n^{-4} \frac{4}{(1-2a\xi)^2} + o(m_n^{-4}).$$

If  $a = \frac{2}{9}\alpha$ , then

$$\begin{aligned} MSE [f_n(x)] &= \Delta_2^2(x)m_n^{-4} \frac{4}{(1-2a\xi)^2} + C_3 f(x)\psi(x)\gamma_n m_n^{1/2} \frac{2}{4-(2\alpha-a)\xi} \\ &\quad + o(m_n^{-4} + \gamma_n m_n^{1/2}). \end{aligned}$$

If  $\frac{2}{9}\alpha < a < 1$ , then

$$MSE [f_n(x)] = C_3 f(x)\psi(x)\gamma_n m_n^{1/2} \frac{2}{4-(2\alpha-a)\xi} + o(\gamma_n m_n^{1/2}).$$

To minimize the  $MSE$  of  $f_n(x)$ , for  $x \in (0, 1)$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$  and  $(m_n)$  must be in  $\mathcal{GS}(2/9)$  such that

$$\left( 2^{2/9} \left( 1 - \frac{4}{9}\xi \right)^{-2/9} \left[ \frac{32\Delta_2^2(x)}{C_3 f(x)\psi(x)} \right]^{2/9} \gamma_n^{-2/9} \right),$$

and then

$$MSE [f_n(x)] = \frac{9(32C_3^8)^{1/9} (\Delta_2(x))^{2/9} (f(x)\psi(x))^{8/9}}{8 \cdot 2^{2/9} (1 - \frac{4}{9}\xi)^{10/9}} \gamma_n^{\frac{8}{9}} + o\left(\gamma_n^{\frac{8}{9}}\right).$$

In the case when  $(\gamma_n) = (\gamma_0 n^{-1})$ , we have  $\xi = \gamma_0^{-1}$ , then, we obtain

$$m_n = 2^{2/9} (\gamma_0 - 4/9)^{-2/9} \left[ \frac{32\Delta_2^2(x)}{C_3 f(x)\psi(x)} \right]^{2/9} n^{2/9},$$

and

$$MSE [f_n(x)] = \frac{9(32C_3^8)^{1/9} (\Delta_2(x))^{2/9} (f(x)\psi(x))^{8/9}}{8} \frac{\gamma_0^2}{2^{8/9} (\gamma_0 - 4/9)^{10/9}} n^{-8/9} + o(n^{-8/9}).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator  $f_n$  defined in (3.1.2), for  $x \in (0, 1)$ .



**Theorem 3.3.1** (Weak pointwise convergence rate).

Let Assumption (A1) – (A4) hold. For  $x \in (0, 1)$ , we have

1. If  $\gamma_n^{-1/2} m_n^{-9/4} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2} m_n^{-1/4} (f_n(x) - f(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( -\frac{2c}{1 - 2a\xi} \Delta_2(x), C_3 \frac{2}{4 - (2\alpha - a)\xi} f(x) \psi(x) \right).$$

2. If  $\gamma_n^{-1/2} m_n^{-9/4} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n^{-2} (f_n(x) - f(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} -\frac{2}{1 - 2a\xi} \Delta_2(x),$$

where  $\xrightarrow[n \rightarrow +\infty]{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the Gaussian-distribution and  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}}$  the convergence in probability.

### 3.3.2 The edges of the interval $[0, 1]$

For the cases  $x = 0, 1$ , we need an additional assumption :

$$(A'4) \quad \lim_{n \rightarrow \infty} (n\gamma_n) \in (\min(2a, (\alpha - a)/2), \infty).$$

The following proposition gives bias, variance and  $MSE$  of  $f_n(x)$ , for  $x = 0, 1$ .

**Proposition 3.3.2.**

Let Assumptions (A1) – (A3) and (A'4) hold, for  $x = 0, 1$ , we have

1. If  $0 < a \leq \frac{\alpha}{5}$ , then

$$\mathbb{E}[f_n(x)] - f(x) = -m_n^{-2} \frac{2}{1 - 2a\xi} \Delta_2(x) + o(m_n^{-2}). \quad (3.3.5)$$

If  $\frac{\alpha}{5} < a < 1$ , then

$$\mathbb{E}[f_n(x)] - f(x) = o(\sqrt{\gamma_n m_n}). \quad (3.3.6)$$

2. If  $\frac{\alpha}{5} \leq a < 1$ , then

$$\text{Var}[f_n(x)] = \frac{5}{2} \gamma_n m_n \frac{1}{2 - (\alpha - a)\xi} f(x) + o(\gamma_n m_n). \quad (3.3.7)$$

If  $0 < a < \frac{\alpha}{5}$ , then

$$\text{Var}[f_n(x)] = o(m_n^{-4}). \quad (3.3.8)$$

3. If  $\lim_{n \rightarrow \infty} (n\gamma_n) > \max(2a, (\alpha - a)/2)$ , that implies  $\lim_{n \rightarrow \infty} (n\gamma_n) > 2a$  and  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a)/2$  so  $1 - 2a\xi > 0$  and  $2 - (\alpha - a)\xi > 0$ , then, (3.3.5) and (3.3.7) hold simultaneously.

4. If  $0 < a < \frac{\alpha}{5}$ , then

$$MSE [f_n(x)] = \Delta_2^2(x)m_n^{-4} \frac{4}{(1 - 2a\xi)^2} + o(m_n^{-4}).$$

If  $a = \frac{\alpha}{5}$ , then

$$MSE [f_n(x)] = \Delta_2^2(x)m_n^{-4} \frac{4}{(1 - 2a\xi)^2} + \frac{5}{2}f(x)\gamma_n m_n \frac{1}{2 - (\alpha - a)\xi} + o(m_n^{-4} + \gamma_n m_n).$$

If  $\frac{\alpha}{5} < a < 1$ , then

$$MSE [f_n(x)] = \frac{5}{2}f(x)\gamma_n m_n \frac{1}{2 - (\alpha - a)\xi} + o(\gamma_n m_n).$$

To minimize the  $MSE$  of  $f_n(x)$ , for  $x = 0, 1$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$  and  $(m_n)$  must be in  $\mathcal{GS}(1/5)$  such that

$$\left( 2^{1/5} \left( 1 - \frac{2}{5}\xi \right)^{-1/5} \left[ \frac{32\Delta_2^2(x)}{5f(x)} \right]^{1/5} \gamma_n^{-1/5} \right),$$

and then

$$MSE [f_n(x)] = \frac{5^{8/5} 32^{1/5} (\Delta_2(x))^{2/5} (f(x))^{4/5}}{8 \cdot 2^{4/5} \left( 1 - \frac{2}{5}\xi \right)^{6/5}} \gamma_n^{4/5} + o(\gamma_n^{4/5}).$$

Moreover, in the case when  $(\gamma_n) = (\gamma_0 n^{-1})$ , we have  $\xi = \gamma_0^{-1}$ , and then

$$\left( 2^{1/5} (\gamma_0 - 2/5)^{-1/5} \left[ \frac{32\Delta_2^2(x)}{5f(x)} \right]^{1/5} n^{1/5} \right),$$

and the corresponding  $MSE$

$$MSE [f_n(x)] = \frac{5^{8/5} 32^{1/5} (\Delta_2(x))^{2/5} (f(x))^{4/5}}{8} \frac{\gamma_0^2}{2^{4/5} (\gamma_0 - 2/5)^{6/5}} n^{-4/5} + o(n^{-4/5}).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator  $f_n$  defined in (3.1.2), for  $x = 0, 1$ .

**Theorem 3.3.2** (Weak pointwise convergence rate).

Let Assumption (A1) – (A3) and (A'4) hold. For  $x = 0, 1$ , we have

1. If  $\gamma_n^{-1/2} m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} c$ , for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2} m_n^{-1/2} (f_n(x) - f(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( -\frac{2c}{1 - 2a\xi} \Delta_2(x), \frac{5}{2} \frac{1}{2 - (\alpha - a)\xi} f(x) \right).$$

2. If  $\gamma_n^{-1/2} m_n^{-9/4} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n^{-2} (f_n(x) - f(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} -\frac{2}{1 - 2a\xi} \Delta_2(x).$$

where  $\xrightarrow[n \rightarrow +\infty]{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the Gaussian-distribution and  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}}$  the convergence in probability.

### 3.3.3 MISE of $f_n$

The following proposition gives the *MISE* of the proposed recursive estimator  $f_n$ .

**Proposition 3.3.3.**

Let Assumptions (A1) – (A4) hold. We have

1. If  $0 < a < \frac{2}{9}\alpha$ , then

$$MISE(f_n) = C_2 m_n^{-4} \frac{4}{(1 - 2a\xi)^2} + o(m_n^{-4}).$$

2. If  $a = \frac{2}{9}\alpha$ , then

$$MISE(f_n) = C_2 m_n^{-4} \frac{4}{(1 - 2a\xi)^2} + C_1 C_3 \gamma_n m_n^{1/2} \frac{2}{4 - (2\alpha - a)\xi} + o(m_n^{-4} + \gamma_n m_n^{1/2}).$$

3. If  $\frac{2}{9}\alpha < a < 1$ , then

$$MISE(f_n) = C_1 C_3 \gamma_n m_n^{1/2} \frac{2}{4 - (2\alpha - a)\xi} + o(\gamma_n m_n^{1/2}).$$

The following result is a consequence of the previous proposition which gives the optimal order ( $m_n$ ) of the estimator (3.1.2) and the corresponding *MISE*.

**Corollary 3.3.1.**

Let Assumptions (A1) – (A4) hold. To minimize the *MISE* of  $f_n$ , the stepsize ( $\gamma_n$ ) must be chosen in  $\mathcal{GS}(-1)$  and ( $m_n$ ) must be in  $\mathcal{GS}(2/9)$  such that

$$\left( 2^{2/9} \left( 1 - \frac{4}{9}\xi \right)^{-2/9} \left[ \frac{32C_2}{C_1C_3} \right]^{2/9} \gamma_n^{-2/9} \right),$$

and then

$$MISE(f_n) = \frac{9(32C_1^8C_3^8C_2)^{1/9}}{8} \frac{1}{2^{8/9} \left( 1 - \frac{4}{9}\xi \right)^{10/9}} \gamma_n^{\frac{8}{9}} + o\left(\gamma_n^{\frac{8}{9}}\right).$$

Moreover, in the case when  $(\gamma_n) = (\gamma_0 n^{-1})$ , we have  $\xi = \gamma_0^{-1}$ , and then

$$m_n = 2^{2/9} (\gamma_0 - 4/9)^{-2/9} \left[ \frac{32C_2}{C_1C_3} \right]^{2/9} n^{2/9}, \quad (3.3.9)$$

and the corresponding *MISE*

$$MISE(f_n) = \frac{9(32C_1^8C_3^8C_2)^{1/9}}{8} \frac{\gamma_0^2}{2^{8/9} (\gamma_0 - 4/9)^{10/9}} n^{-8/9} + o(n^{-8/9}). \quad (3.3.10)$$

Note that the minimum of  $\frac{\gamma_0^2}{(\gamma_0 - 4/9)^{10/9}}$  is reached for  $\gamma_0 = 1$ .

Furthermore, to minimize the variance of  $f_n$ , we should choose  $\gamma_0 = 1 - \frac{a}{2}$ , with  $a = 2/9$  for  $x \in (0, 1)$  and  $\gamma_0 = 1 - a$  with  $a = 1/5$  for  $x = 0, 1$ . Therefore, in Section 3.4, we will consider as a stepsize  $(\gamma_n) = (n^{-1})$ ,  $(\gamma_n) = (\frac{8}{9}n^{-1})$  and  $(\gamma_n) = (\frac{4}{5}n^{-1})$ .

Now, let us recall the bias, variance and *MISE* of Vitale's Bernstein density estimator (3.1.1). Under the classical assumption on the density, that  $f$  is continuous and admits two continuous and bounded derivatives, for  $x \in [0, 1]$ , we have

$$\mathbb{E}[\tilde{f}_n(x)] - f(x) = \frac{\Delta_1(x)}{m} + o(m^{-1}), \text{ uniformly in } x \in [0, 1].$$

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

Also, we have

$$MISE(\tilde{f}_n) = \frac{m^{1/2}C_1}{n} + \frac{C_4}{m^2} + o(m^{1/2}n^{-1}) + o(m^{-2}).$$

To minimize the  $MISE$  of  $\tilde{f}_n$ , the order  $(m_n)$  must equal to

$$\left( \left[ \frac{4C_4}{C_1} \right]^{2/5} n^{2/5} \right), \quad (3.3.11)$$

and the corresponding  $MISE$

$$MISE(\tilde{f}_n) = \frac{5(C_1^4 C_4)^{1/5}}{4} n^{-4/5} + o(n^{-4/5}).$$

If we let  $h = m^{-1}$  be the bandwidth of the Vitale's estimator  $\tilde{f}_n$ , we notice that the bias of  $\tilde{f}_n$  is  $O(h^{-1})$  which is more than the bias typically obtained with typical kernel density estimators which generally of  $O(h^2)$  except near the boundaries.

To reduce this bias, Leblanc [55] suggested a new Bernstein estimator using the method of bias correction. This methodology was adopted by Politis et Romano [82] in the context of spectral density estimation and is linked with the work of Schucany et al. [95] and Schucany et Sommers [96] on bias reduction in estimation.

The bias-corrected Bernstein density estimator is given by

$$\tilde{f}_{n,m,m/2}(x) = 2\tilde{f}_{n,m}(x) - \tilde{f}_{n,m/2}(x), \quad x \in [0, 1] \quad (3.3.12)$$

where  $\tilde{f}_{n,m}$  and  $\tilde{f}_{n,m/2}$  are the Bernstein density estimators introduced by Vitale with order  $m$  and  $m/2$  respectively which defined in (3.1.1). Let us now recall the characteristics of the estimator  $\tilde{f}_{n,m,m/2}$ . Under the Assumption (A1), we have

$$\mathbb{E}[\tilde{f}_{n,m,m/2}(x)] - f(x) = -2\frac{\Delta_2(x)}{m^2} + o(m^{-2}), \quad \text{uniformly in } x \in [0, 1].$$

$$\text{Var}[\tilde{f}_{n,m,m/2}(x)] = \begin{cases} C_3 \frac{m^{1/2}}{n} f(x)\psi(x) + o_x\left(\frac{m^{1/2}}{n}\right) & \text{for } x \in (0, 1), \\ \frac{5}{2} \frac{m}{n} f(x) + o_x\left(\frac{m}{n}\right) & \text{for } x = 0, 1. \end{cases}$$

and then

$$MISE(\tilde{f}_{n,m,m/2}) = \frac{C_1 C_3 m^{1/2}}{n} + \frac{4C_2}{m^4} + o(m^{1/2}n^{-1}) + o(m^{-4}).$$

To minimize the *MISE* of  $\tilde{f}_{n,m,m/2}$ , the order  $(m_n)$  must equal to

$$\left( \left[ \frac{32C_2}{C_1C_3} \right]^{2/9} n^{2/9} \right), \quad (3.3.13)$$

and then

$$MISE \left( \tilde{f}_{n,m,m/2} \right) = \frac{9(32C_1^8C_2C_3^8)^{1/9}}{8} n^{-8/9} + o(n^{-8/9}).$$

Igarashi and Kakizawa [42] have generalized the estimator proposed by Leblanc  $\tilde{f}_{n,m,m/2}$  and defined in (3.3.12)

$$\tilde{f}_{n,m,m/b}(x) = \frac{b}{b-1} \tilde{f}_{n,m}(x) - \frac{1}{b-1} \tilde{f}_{n,m/b}(x), \quad x \in [0, 1] \quad (3.3.14)$$

where  $b = 2, 3, \dots$  and  $\tilde{f}_{n,m}$  and  $\tilde{f}_{n,m/b}$  are the Vitale's density estimators defined in (3.1.1) with order  $m$  and  $m/b$  respectively. Under the Assumption (A1), we have

$$\mathbb{E} \left[ \tilde{f}_{n,m,m/b}(x) \right] - f(x) = -\frac{b}{m^2} \Delta_2(x) + o(m^{-2}), \quad \text{uniformly in } x \in [0, 1].$$

$$Var[\tilde{f}_{n,m,m/b}(x)] = \begin{cases} \lambda_1(b) \frac{m^{1/2}}{n} f(x) \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right) & \text{for } x \in (0, 1), \\ \lambda_2(b) \frac{m}{n} f(x) + o_x \left( \frac{m}{n} \right) & \text{for } x = 0, 1, \end{cases}$$

$$\lambda_1(b) = \frac{1}{(1-b)^2} \left\{ b^2 + b^{-1/2} - 2b \left( \frac{2}{b+1} \right)^{1/2} \right\},$$

$$\lambda_2(b) = \frac{1}{(1-b)^2} \{ b^2 + b^{-1} - 2 \}.$$

Also, we have

$$MISE \left( \tilde{f}_{n,m,m/b} \right) = \lambda_1(b) C_1 \frac{m^{1/2}}{n} + \frac{b^2 C_2}{m^4} + o(m^{1/2} n^{-1}) + o(m^{-4}).$$

To minimize the *MISE* of  $\tilde{f}_{n,m,m/b}$ , the order  $(m_n)$  must equal to

$$\left( \left[ \frac{b^2}{\lambda_1(b)} \frac{8C_2}{C_1} \right]^{2/9} n^{2/9} \right), \quad (3.3.15)$$

and then

$$MISE \left( \tilde{f}_{n,m,m/b} \right) = \frac{9}{8} (b\lambda_1^4(b))^{2/9} (8C_1^8C_2)^{1/9} n^{-8/9} + o(n^{-8/9}). \quad (3.3.16)$$

The equation (3.3.16) indicates that  $b = 2$  is the best choice in terms of the *MISE* for the estimator  $\tilde{f}_{n,m,m/b}$ , since the factor  $(b\lambda_1^4(b))^{2/9}$  is increasing in  $b = 2, 3, \dots$

This method of bias correction reduces the bias of Bernstein estimator from  $O(m^{-1})$  to  $O(m^{-2})$ , but it loses nonnegativity. As an additive bias correction to the logarithm of estimator, Terrell et Scott [116] originally developed a multiplicative bias correction that enables keeping the nonnegativity. This method was adopted by Hirukawa [41] for the beta kernel estimator introduced by Chen [14]. Igarashi and Kakizawa [42] also applied this method for the Bernstein estimator to obtain

$$\tilde{f}_{n,m,b,\varepsilon}(x) = \left\{ \tilde{f}_{n,m}(x) \right\}^{b/(b-1)} \left\{ \tilde{f}_{n,m/b}(x) + \varepsilon \right\}^{-1/b-1}, \quad x \in [0, 1], \quad (3.3.17)$$

for some  $\varepsilon = \varepsilon(m) > 0$ , converging to zero at a suitable rate. Under the Assumption (A1), for  $x \in [0, 1]$  such as  $f(x) > 0$ , with  $m = O(n^\eta)$  and  $\varepsilon \approx m^\tau$  where  $\eta \in (0, 1)$  and  $\tau > 2$  we have

$$\begin{aligned} \mathbb{E}[\tilde{f}_{n,m,b,\varepsilon}(x)] - f(x) &= -\frac{b}{m^2} \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) \right\} \\ &\quad + O\left( \text{Var}[\tilde{f}_{n,m}(x)] + \text{Var}[\tilde{f}_{n,m/b}(x)] \right) + o(m^{-2}). \end{aligned}$$

$$\text{Var}[\tilde{f}_{n,m,b,\varepsilon}(x)] = \text{Var}[\tilde{f}_{n,m,m/b}(x)] + o\left( \text{Var}[\tilde{f}_{n,m}(x)] + \text{Var}[\tilde{f}_{n,m/b}(x)] + m^{-4} \right).$$

Also, we have

$$\text{MISE}\left(\tilde{f}_{n,m,b,\varepsilon}\right) = \lambda_1(b)C_1 \frac{m^{1/2}}{n} + \frac{b^2 C_5}{m^4} + o(m^{1/2}n) + o(m^{-4}),$$

To minimize the *MISE* of  $\tilde{f}_{n,m,b,\varepsilon}$ , the order  $(m_n)$  must equal to

$$\left( \left[ \frac{b^2}{\lambda_1(b)} \frac{8C_5}{C_1} \right]^{2/9} n^{2/9} \right), \quad (3.3.18)$$

and the corresponding *MISE*

$$\text{MISE}\left(\tilde{f}_{n,m,b,\varepsilon}\right) = \frac{9}{8} (b\lambda_1^4(b))^{2/9} (8C_1^8 C_5)^{1/9} n^{-8/9} + o(n^{-8/9}).$$

Note that the estimator  $\tilde{f}_{n,m,b,\varepsilon}$  retains nonnegativity, but it is not a genuine density. In fact,  $\tilde{f}_{n,m,b,\varepsilon}$  does not generally integrate to unity. To solve this problem, Igarashi and Kakizawa [42] proposed the normalized bias corrected Bernstein estimator

$$\tilde{f}_{n,m,b,\varepsilon}^N(x) = \frac{\tilde{f}_{n,m,b,\varepsilon}(x)}{\int_0^1 \tilde{f}_{n,m,b,\varepsilon}(y) dy}, \quad x \in [0, 1]. \quad (3.3.19)$$

Under the Assumption (A1), for  $x \in [0, 1]$  such as  $f(x) > 0$ , with  $m = O(n^\eta)$  and  $\varepsilon \approx m^\tau$  where  $\eta \in (0, 1)$  and  $\tau > 2$  we have

$$\begin{aligned} \mathbb{E}[\tilde{f}_{n,m,b,\varepsilon}^N(x)] - f(x) &= -\frac{b}{m^2} \left\{ -\frac{\Delta_1^2(x)}{2f(x)} + \Delta_2(x) + f(x) \int_0^1 \frac{\Delta_1^2(y)}{2f(y)} dy \right\} \\ &\quad + O\left(Var[\tilde{f}_{n,m}(x)] + Var[\tilde{f}_{n,m/b}(x)] + n^{-1}m^{1/2}\right) \\ &\quad + o(m^{-2}). \end{aligned}$$

$$\begin{aligned} Var[\tilde{f}_{n,m,b,\varepsilon}^N(x)] &= Var[\tilde{f}_{n,m,m/b}(x)] \\ &\quad + o\left(Var[\tilde{f}_{n,m}(x)] + Var[\tilde{f}_{n,m/b}(x)] + n^{-1}m^{1/2} + m^{-4}\right), \end{aligned}$$

and then, we have

$$MISE\left(\tilde{f}_{n,m,b,\varepsilon}^N\right) = \lambda_1(b)C_1 \frac{m^{1/2}}{n} + \frac{b^2C_6}{m^4} + o(m^{1/2}n^{-1}) + o(m^{-4}).$$

To minimize the *MISE* of  $\tilde{f}_{n,m,b,\varepsilon}^N$ , the order  $(m_n)$  must equal to

$$\left( \left[ \frac{b^2}{\lambda_1(b)} \frac{8C_6}{C_1} \right]^{2/9} n^{2/9} \right), \quad (3.3.20)$$

and the corresponding *MISE*

$$MISE\left(\tilde{f}_{n,m,b,\varepsilon}^N\right) = \frac{9}{8} (b\lambda_1^4(b))^{2/9} (8C_1^8C_6)^{1/9} n^{-8/9} + o(n^{-8/9}).$$



## 3.4 Applications

In using the Bernstein polynomials, we must consider a density on  $[0, 1]$ . To do so, we need to make some suitable transformations in different cases (we list below) :

1. Suppose that  $X$  is concentrated on a finite support  $[a, b]$ . Then we work with the sample values  $Y_1, \dots, Y_n$  where  $Y_i = \frac{X_i - a}{b - a}$ . Denoting  $g_n(x)$  the estimated density function of  $Y_1, \dots, Y_n$ , we compute the estimated density  $f_n$  of  $X$

$$f_n(x) = \frac{1}{b - a} g_n \left( \frac{x - a}{b - a} \right).$$

2. For the densities functions concentrated on the interval  $(-\infty, +\infty)$ , we can use the transformed sample  $Y_i = \frac{1}{2} + \frac{1}{\pi} \arctan(X_i)$  which transforms the range to the interval  $(0, 1)$ . Hence

$$f_n(x) = \frac{1}{\pi(1 + x^2)} g_n \left( \frac{1}{2} + \frac{1}{\pi} \arctan(x) \right).$$

3. For the support  $[0, \infty)$ , we can use the transformed sample  $Y_i = \frac{X_i}{X_i + 1}$  which transforms the range to the interval  $(0, 1)$ . Hence

$$f_n(x) = \frac{1}{(1 + x)^2} g_n \left( \frac{x}{1 + x} \right).$$

The aim of this paragraph is to compare the performance of Vitale's estimator  $\tilde{f}_n$  defined in (3.1.1), Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  given in (3.3.12), the generalized estimator  $\tilde{f}_{n,m,m/b}$ , defined in (3.3.14), the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,b,\varepsilon}$  given in (3.3.17) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  given in (3.3.19) with that of the proposed estimator (3.1.2).

- (1) When applying  $f_n$ , one needs to choose two quantities :

- The stepsize  $(\gamma_n) = (n^{-1})$  or  $(\gamma_n) = (\frac{8}{9}n^{-1})$ , or  $(\gamma_n) = (\frac{4}{5}n^{-1})$ .
- The order  $(m_n)$  is chosen to be equal to (3.3.9).

- (2) When applying  $\tilde{f}_n$ , one needs to choose the order  $(m_n)$  to be equal to (3.3.11).

- (3) When applying  $\tilde{f}_{n,m,m/2}$ , one needs to choose the order  $(m_n)$  to be equal to (3.3.13).

- (4) When applying  $\tilde{f}_{n,m,m/b}$ , one needs to choose the order  $(m_n)$  to be equal to (3.3.15) and  $b = 3, 4$ .
- (5) When applying  $\tilde{f}_{n,m,b,\varepsilon}$ , one needs to choose the order  $(m_n)$  to be equal to (3.3.18),  $b = 2, 3, 4$  and  $\varepsilon = 0.00001$ .
- (6) When applying  $\tilde{f}_{n,m,b,\varepsilon}^N$ , one needs to choose the order  $(m_n)$  to be equal to (3.3.20),  $b = 2, 3, 4$  and  $\varepsilon = 0.00001$ .

### 3.4.1 Simulations

We consider ten density functions, as follows :

(a) the beta density  $\mathcal{B}(3, 5)$ ,  $f(x) = \frac{x^2(1-x)^4}{B(3,5)}$ ,

(b) the beta density  $\mathcal{B}(1, 6)$ ,  $f(x) = \frac{(1-x)^5}{B(1,6)}$

(c) the beta density  $\mathcal{B}(3, 1)$ ,  $f(x) = \frac{x^2}{B(3,1)}$

(d) the beta mixture density  $1/2\mathcal{B}(3, 9) + 1/2\mathcal{B}(9, 3)$ ,  $f(x) = 0.5\frac{x^2(1-x)^8}{B(3,9)} + 0.5\frac{x^8(1-x)^2}{B(9,3)}$ ,

(e) the beta mixture density  $1/2\mathcal{B}(3, 1) + 1/2\mathcal{B}(10, 10)$ ,  $f(x) = 0.5\frac{x^2}{B(3,1)} + 0.5\frac{x^9(1-x)^9}{B(10,10)}$ ,

(f) the beta mixture density  $1/2\mathcal{B}(1, 6) + 1/2\mathcal{B}(3, 5)$ ,  $f(x) = 0.5\frac{(1-x)^5}{B(1,6)} + 0.5\frac{x^2(1-x)^4}{B(3,5)}$ ,

(g) the beta mixture density  $1/2\mathcal{B}(2, 1) + 1/2\mathcal{B}(1, 4)$ ,  $f(x) = 0.5\frac{x}{B(2,1)} + 0.5\frac{(1-x)^3}{B(1,4)}$ ,

(h) the truncated exponential density  $\mathcal{E}_{[0,1]}(1/0.8)$ ,  $f(x) = \frac{\exp(-x/0.8)}{0.8\{1-\exp(-1/0.8)\}}$ ,

(i) the truncated normal density  $\mathcal{N}_{[0,1]}(0, 1)$ ,  $f(x) = \frac{\exp(-x^2/2)}{\int_0^1 \exp(-t^2/2) dt}$

(j) the truncated normal mixture density  $1/4\mathcal{N}_{[0,1]}(2, 1) + 3/4\mathcal{N}_{[0,1]}(-3, 1)$ ,

$$f(x) = 0.25 \frac{\exp(-(x-2)^2/2)}{\int_0^1 \exp(-(t-2)^2/2) dt} + 0.75 \frac{\exp(-(x+3)^2/2)}{\int_0^1 \exp(-(t+3)^2/2) dt}.$$

For each density function and sample of size  $n$ , we approximate the average integrated squared error (*ISE*) of the estimator using  $N = 500$  trials of sample size  $n$ ;

$\overline{ISE} = \frac{1}{N} \sum_{k=1}^N ISE[\hat{f}_k]$ , where  $\hat{f}_k$  is the estimator computed from the  $k$ th sample,

and,  $ISE[\hat{f}] = \int_0^1 \left\{ \hat{f}(x) - f(x) \right\}^2 dx$ .

From Tables 3.1 and 3.2 we conclude that :

- In all the cases, the average  $ISE$  of our density estimator (3.1.2) is smaller than that of Vitale's estimator (3.1.1), except the cases (e) and (f) of the Beta mixture and the cases (a), (d), (g) and (h) for the small size  $n = 50$  and in the case (i) for the size  $n = 50$  and  $n = 200$ .
- In all the cases, the average  $ISE$  of our recursive density estimator (3.1.2) is slightly larger than that of Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  given in (3.3.12).
- In all the cases, the average  $ISE$  of our density estimator (3.1.2) is smaller than that of the generalized estimator  $\tilde{f}_{n,m,m/b}$ , with  $b = 4$  (see (3.3.14)).
- In all the cases, the average  $ISE$  of our density estimator (3.1.2) is smaller than that of the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,b,\varepsilon}$  given in (3.3.17) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  given in (3.3.19), except the cases (g) and (h).
- The average  $ISE$  of the generalized estimator  $\tilde{f}_{n,m,m/b}$ , (3.3.14) increase when  $b$  increase so the optimal choice is  $b = 2$  which corresponds to Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  given in (3.3.12).
- The average  $ISE$  of the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,b,\varepsilon}$  given in (3.3.17) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) increase when  $b$  increase
- The average  $ISE$  of the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,b,\varepsilon}$  given in (3.3.17) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  given in (3.3.19) are smaller than that of the estimator  $\tilde{f}_{n,m,m/b}$  defined in (3.3.14), in the cases (g) and (h) and are larger in the other cases.
- The average  $ISE$  of the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,b,\varepsilon}$  (3.3.17) is larger than that of the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  given in (3.3.19) in the cases (a), (g), (h), (i) and (j) and is smaller in the other cases.
- The average  $ISE$  decreases as the sample size increases.

For qualitative comparison between the Vitale's estimator  $\tilde{f}_n$  given in (3.1.1), Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  given in (3.3.12), the generalized estimator  $\tilde{f}_{n,m,m/b}$  given in ((3.3.14)) with  $b = 2, 3$ , the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17), the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) and the proposed estimator (3.1.2). Figure 3.1, shows density estimates plotted for 500 simulated samples from  $\mathcal{B}(3, 5)$  density of sizes  $n = 50$  (left panel) and  $n = 250$  (right panel).

From figure 3.1, we conclude that :

- Our density estimator (3.1.2) using the stepsize  $(\gamma_n) = (n^{-1})$  is closer to the true density function than that of Vitale's estimator  $\tilde{f}_n$  given in (3.1.1), the generalized estimator  $\tilde{f}_{n,m,m/4}$  given in (3.3.14), Kakizwa's estimator  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) with  $b = 2$  and  $\varepsilon = 0.00001$
- Within the interval  $[0, 1]$ , our density estimator (3.1.2) using the stepsize  $(\gamma_n) = (n^{-1})$  is closer to the true density function than that of Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12).
- When the sample size increases, we get closer estimation of the true density function.

### 3.4.2 Real dataset

In any practical situation, to estimate an unknown density function  $f$ , it is essential to specify the order  $m$  to be used for the estimator. One way to do this is by using least squares cross-validation (*LSCV*) method to obtain a data-driven choice of  $m$ .

Let us recall that the *LSCV* method is based on minimizing the integrated square error between the estimated density function  $\hat{f}$  and the true density function  $f$

$$\int_0^1 \left( \hat{f}(x) - f(x) \right)^2 dx = \int_0^1 \hat{f}^2(x) dx - 2 \int_0^1 \hat{f}(x) f(x) dx + \int_0^1 f^2(x) dx.$$

From this, Silverman [101, page 48-49] derived the score function

$$LSCV_{\hat{f}}(m) = \int_0^1 \hat{f}^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i), \quad (3.4.1)$$

where  $\hat{f}_{-i}$  is the density estimate without the data point  $X_i$ .

The smoothing parameter is chosen by minimizing  $LSCV(m)$  ( $m = \arg \min_m LSCV(m)$ ).

For our recursive estimator (3.1.2), we set  $(\gamma_n) = (n^{-1})$ , and the order  $m$  is chosen to minimize the following criterion :

$$LSCV_{f_n}(m) = \int_0^1 f_n^2(x) dx - \frac{2}{n} \sum_{i=1}^n f_{n,-i}(X_i).$$

We define integer sequences  $p_i = [m_i X_i]$  and  $q_i = [m_i X_i / 2]$ , so that  $X_i \in \left(\frac{p_i}{m_i}, \frac{p_i+1}{m_i}\right]$  and  $X_i \in \left(\frac{2q_i}{m_i}, \frac{2(q_i+1)}{m_i}\right]$ . Then we obtain

$$f_{n,-i}(x) = \frac{1}{n-1} \left[ n f_n(x) - \sum_{i=1}^n \left\{ 2m_i b_{p_i}(m_i - 1, x) - \frac{m_i}{2} b_{q_i} \left( \frac{m_i}{2} - 1, x \right) \right\} \right],$$

so we conclude

$$LSCV_{f_n}(m) = \int_0^1 f_n^2(x) dx - \frac{2}{n-1} \left[ \sum_{i=1}^n f_n(X_i) - \frac{1}{n} \sum_{i=1}^n \left\{ 2m_i b_{p_i}(m_i - 1, X_i) - \frac{m_i}{2} b_{q_i} \left( \frac{m_i}{2} - 1, X_i \right) \right\} \right].$$

Note that, the  $LSCV$  function for Vitale's estimator (3.1.1), is written as

$$LSCV_{\tilde{f}_n}(m) = \int_0^1 \left\{ \tilde{f}_n(x) \right\}^2 dx - \frac{2}{n} \sum_{i=1}^n \tilde{f}_{n,-i}(X_i).$$

We define integer sequences  $k_i = [m X_i]$ , so that  $X_i \in \left(\frac{k_i}{m}, \frac{k_i+1}{m}\right]$ , then we have

$$\tilde{f}_{n,-i}(x) = \frac{1}{n-1} \left[ \tilde{f}_n(x) - m b_{k_i}(m-1, x) \right],$$

so we conclude

$$LSCV_{\tilde{f}_n}(m) = \int_0^1 \left\{ \tilde{f}_n(x) \right\}^2 dx - \frac{2}{n-1} \left[ \sum_{i=1}^n \tilde{f}_n(X_i) - \frac{m}{n} \sum_{i=1}^n b_{k_i}(m-1, X_i) \right].$$

The *LSCV* function for the estimator  $\tilde{f}_{n,m,m/b}$  defined in (3.3.14) with  $b = 2, 3, 4, \dots$ , is written as

$$LSCV_{\tilde{f}_{n,m,m/b}}(m) = \int_0^1 \left\{ \tilde{f}_{n,m,m/b}(x) \right\}^2 dx - \frac{2}{n} \sum_{i=1}^n \tilde{f}_{n,m,m/b,-i}(X_i).$$

We define integer sequences  $k_i = [mX_i]$  and  $r_i = [mX_i/b]$ , so that  $X_i \in \left(\frac{k_i}{m}, \frac{k_i+1}{m}\right]$  and  $X_i \in \left(\frac{br_i}{m}, \frac{b(r_i+1)}{m}\right]$ . Then we obtain

$$\begin{aligned} LSCV_{\tilde{f}_{n,m,m/b}}(m) &= \int_0^1 \left\{ \tilde{f}_{n,m,m/b}(x) \right\}^2 dx - \frac{2}{n-1} \left[ \sum_{i=1}^n \tilde{f}_{n,m,m/b}(X_i) \right. \\ &\quad \left. - \frac{1}{n} \sum_{i=1}^n \left\{ \frac{b}{b-1} mb_{k_i}(m-1, X_i) - \frac{1}{b-1} \frac{m}{b} b_{r_i} \left( \frac{m}{b} - 1, X_i \right) \right\} \right]. \end{aligned}$$

Using the Kakizawa's estimators  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17) and  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19), the *LSCV* function is written as in (3.4.1).

To highlight our proposed estimator (3.1.2), we look, first, at the data coming from a study of prostate cancer (Stamey et al. [112]). The predictors are eight clinical measures : `lcavol`, which is the logarithm of the cancer volume, `lweight`, which is the logarithm of the prostate weight, `age`, `lbph`, which is the logarithm of the amount of benign prostatic hyperplasia, `svi`, which is the seminal vesicle invasion, `lcp`, which is the logarithm of the capsular penetration, `gleason`, which is the Gleason score and `pgg45`, which is the percentage Gleason score 4 or 5. The response `lpsa`, which is the logarithm of prostate-specific antigen. The prostate cancer data were divided into two parts : a training set with 67 observations and a test set with 30 observations. We apply our estimator  $f_n$  on the training data for the variable prostate-specific antigen. It is clear in this setup that the density to be estimated is defined only for  $x > 0$ . For convenience, we also assume that the maximum prostate-specific antigen is 19 (the data are such that  $\min_i(x_i) = 0.65$  and  $\max_i(x_i) = 18.55$ ).

The *LSCV* procedure was performed and resulted in  $m = 9$  for Vitale's estimator  $\tilde{f}_n$  defined in (3.1.1),  $(m_n) = (n^{0.687})$  for our proposed estimator  $f_n$  defined in (3.1.2),  $m = 12$  for Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12),  $m = 12$  for the estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14),  $m = 9$  for the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,2,0.00001}$  (3.3.17) and  $m = 9$  for the normalized estimator  $\tilde{f}_{n,m,2,0.00001}^N$  (3.3.19). These estimators are shown in Figure 3.2 along with an histogram of the data and a Gaussian kernel density estimate using the *LSCV*-based bandwidth  $h = 0.1111$ .

Second, we consider the Old Faithful data given in Tab.2.2 of Silverman [101]. These data consist of the eruption lengths (in minutes) of 107 eruptions of the Old Faithful geyser in Yellowstone National Park, U.S.A. The data are such that  $\min_i(x_i) = 1.67$  and  $\max_i(x_i) = 4.93$ , so 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 unit interval.

The *LSCV* procedure was performed and resulted in  $m = 104$  for Vitale's estimator  $\tilde{f}_n$  defined in (3.1.1),  $(m_n) = (n^{0.987})$  for our proposed estimator  $f_n$  defined in (3.1.2),  $m = 66$  for Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12),  $m = 52$  for the estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14),  $m = 66$  for the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,2,0.00001}$  defined in (3.3.17) and  $m = 66$  for the normalized estimator  $\tilde{f}_{n,m,2,0.00001}^N$  defined in (3.3.19). These estimators are shown in Figure 3.3 along with an histogram of the data and a Gaussian kernel density estimate using the *LSCV*-based bandwidth  $h = 0.3677$ .

The last example is the tuna data given in Chen [12]. The data come from an aerial line transect survey of Southern Bluefin Tuna in the Great Australian Bight. An aircraft with two spotters on board flies randomly allocated line transects. The data are the perpendicular sighting distances (in miles) of 64 detected tuna schools to the transect lines. The survey was conducted in summer when tuna tend to stay on the surface. We analyzed the transformed data divided by  $w = 18$  (the data are such that  $\min_i(x_i) = 0.19$  and  $\max_i(x_i) = 16.26$ ).

The *LSCV* procedure was performed and resulted in  $m = 14$  for Vitale's estimator  $\tilde{f}_n$  (3.1.1),  $(m_n) = (n^{0.633})$  for our proposed estimator  $f_n$  defined in (3.1.2),  $m = 4$  for Leblanc's estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12),  $m = 4$  for the estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14),  $m = 8$  for the multiplicative bias corrected Bernstein estimator  $\tilde{f}_{n,m,2,0.00001}$  defined in (3.3.17) and  $m = 4$  for the normalized estimator  $\tilde{f}_{n,m,2,0.00001}^N$  defined in (3.3.19). These estimators are shown in Figure 3.4 along with an histogram of the data and a Gaussian kernel density estimate using the *LSCV*-based bandwidth  $h = 1.291$ .

From Figures 3.2, 3.3 and 3.4, we conclude that all these estimators are smooth and seem to capture the pattern highlighted by the histogram. Besides, we can observe that our recursive estimator is better near the boundaries.

### 3.5 Conclusion

In this chapter, we propose a recursive estimator of a density function based on a stochastic algorithm derived from Robbins-Monro's scheme, using Bernstein polynomials. We first study its asymptotic properties. We show that our estimator of density function have very good boundary properties, including the absence of boundary bias and a reduced rate of bias (a rate of  $m^{-2}$ ) than that of Vitale's estimator with a bias rate of  $m^{-1}$ . For almost all the cases, the average *ISE* of the proposed estimator (3.1.2) with a stepsize  $(\gamma_n) = (n^{-1})$  and the corresponding order  $(m_n)$  is smaller than that of Vitale's estimator and the multiplicative bias-corrected estimator defined by Kakizawa. Also, our recursive density estimator has a slightly larger average *ISE* than that of Leblanc's estimator. In addition, a major advantage of our proposal is that its update, when new sample points are available, require less computational cost than the non recursive estimators. Our proposed estimator always integrates to unity, but is not necessarily non negative. However, we found that truncation and renormalisation may solve this issue.



Finally, through simple real-life examples (Prostate specific antigen data, Old Faithful data and tuna data) and a simulation study, we demonstrated how the Bernstein polynomial density estimators can lead to very satisfactory estimates of the underlying density. In conclusion, using the proposed recursive estimator  $f_n$ , we can obtain better results than those given by Vitale's estimator, Leblanc's estimator and the multiplicative bias-corrected estimator defined by Kakizawa especially near the boundaries.

## 3.6 Proofs

In this section, we present proofs for the results presented in Section 3.3. To do so, we recall the technical lemma introduced in the previous chapter which will be used throughout the proofs.

**Lemma 3.6.1.** *Let  $(v_n) \in \mathcal{GS}(v^*)$ ,  $(\gamma_n) \in \mathcal{GS}(-\alpha)$ , and  $l > 0$  such that  $l - v^*\xi > 0$ .*

*We have*

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

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

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

For the same reasons mentioned in the previous chapter, we can consider  $\alpha = 1$  in all the results given in this chapter

### 3.6.1 Proof of Proposition 3.3.1 and Proposition 3.3.2

In view of (3.1.2), we have

$$\mathbb{E}[f_n(x)] - f(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k (\mathbb{E}[Z_k(x)] - f(x)). \quad (3.6.1)$$

Leblanc showed that, for  $x \in [0, 1]$ , we have (see Theorem 6 in Leblanc [55])

$$\mathbb{E}[Z_n(x)] - f(x) = -2 \frac{\Delta_2(x)}{m_n^2} + o(m_n^2).$$

Substituting this result into (3.6.1) leads to

$$\mathbb{E}[f_n(x)] - f(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k \left[ \frac{-2\Delta_2(x)}{m_k^2} (1 + o(1)) \right], \quad x \in [0, 1].$$

For  $x \in (0, 1)$ , we obtain

- In the case  $0 < a \leq \frac{2}{9}\alpha$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > 2a$ . Lemma 3.6.1 gives (3.3.1).
- In the case  $\frac{2}{9}\alpha < a < 1$ , we have  $m_n^{-2} = o\left(\sqrt{\gamma_n m_n^{1/2}}\right)$ , since  $2a > (2\alpha - a)/4$ . Lemma 3.6.1 gives (3.3.2).

For  $x = 0, 1$ , we obtain

- In the case  $0 < a \leq \frac{\alpha}{5}$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > 2a$ . Lemma 3.6.1 gives (3.3.5).
- In the case  $\frac{\alpha}{5} < a < 1$ , we have  $m_n^{-2} = o\left(\sqrt{\gamma_n m_n}\right)$  and  $2a > (\alpha - a)/2$ . Lemma 3.6.1 gives (3.3.6).

On the other hand, we have

$$\text{Var}[f_n(x)] = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{Var}[Z_k(x)],$$

Leblanc derived that (see Theoreme 6 in Leblanc [55])

$$\text{Var}[Z_n(x)] = \begin{cases} \left( \frac{1}{\sqrt{2}} + 4 \left(1 - \sqrt{\frac{2}{3}}\right) \right) m_n^{1/2} f(x) \psi(x) + o_x(m_n^{1/2}), & \text{for } x \in (0, 1), \\ \frac{5}{2} m_n f(x) + o(m_n), & \text{for } x = 0, 1. \end{cases}$$

It follows that

$$\text{Var}[f_n(x)] = \begin{cases} C_3 f(x) \psi(x) \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k^{1/2} (1 + o_x(1)), & \text{for } x \in (0, 1), \\ \frac{5}{2} f(x) \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k (1 + o(1)), & \text{for } x = 0, 1. \end{cases}$$

Using this result, for  $x \in (0, 1)$ , we obtain

- In the case  $\frac{2}{9}\alpha \leq a < 1$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > (2\alpha - a)/4$ , the application of lemma 3.6.1 then gives

$$\text{Var}[f_n(x)] = C_3 \gamma_n m_n^{1/2} \frac{2}{4 - (2\alpha - a)\xi} f(x) \psi(x) + o\left(\gamma_n m_n^{1/2}\right),$$

and (3.3.3) follows.

— In the case  $0 < a < \frac{2}{9}\alpha$ , we have  $2a < (2\alpha - a)/4$  and  $\gamma_n m_n^{1/2} = o(m_n^{-4})$ , then lemma 3.6.1 ensures that

$$\begin{aligned} \text{Var}[f_n(x)] &= \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k o(m_k^{-4}), \\ &= o(m_n^{-4}), \end{aligned}$$

which gives (3.3.4).

Similarly, for  $x = 0, 1$ , we have

— In the case  $\frac{\alpha}{5} \leq a < 1$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a)/2$ . Lemma 3.6.1 gives (3.3.7).

— In the case  $0 < a < \frac{\alpha}{5}$ , we have  $2a < (\alpha - a)/2$  and  $\gamma_n m_n = o(m_n^{-4})$ . Lemma 3.6.1 gives (3.3.8).

### 3.6.2 Proof of Proposition 3.3.3

Using Proposition 3.3.1, we have

• In the case  $0 < a \leq \frac{2}{9}\alpha$ , we have

$$\begin{aligned} \int_0^1 \{\mathbb{E}[f_n(x)] - f(x)\}^2 dx &= \int_0^1 \left[ -2m_n^{-2} \frac{1}{1 - 2a\xi} \Delta_2(x) + o(m_n^{-2}) \right]^2 dx \\ &= 4C_2 m_n^{-4} \frac{1}{(1 - 2a\xi)^2} + o(m_n^{-4}). \end{aligned} \quad (3.6.2)$$

• In the case  $\frac{2}{9}\alpha < a < 1$ , we have

$$\int_0^1 \{\mathbb{E}[f_n(x)] - f(x)\}^2 dx = o(\gamma_n m_n^{1/2}). \quad (3.6.3)$$

On the other hand, we note that

$$\int_0^1 \text{Var}[f_n(x)] dx = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \int_0^1 \text{Var}[Z_k(x)] dx,$$

We know (see the proof of Theorem 7 in [55]) that

$$\int_0^1 \text{Var}[Z_k(x)] dx = C_1 C_3 m_k^{1/2} + o(m_k^{1/2}),$$

and this gives

$$\int_0^1 \text{Var}[f_n(x)] dx = C_1 C_3 \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k^{1/2} (1 + o(1)).$$

- In the case  $\frac{2}{9}\alpha \leq a < 1$ , we have  $\lim_{n \rightarrow +\infty} (n\gamma_n) > (2\alpha - a)/4$ . Lemma 3.6.1 gives

$$\int_0^1 \text{Var} [f_n(x)] dx = C_1 C_3 \gamma_n m_n^{1/2} \frac{2}{4 - (2\alpha - a)\xi} + o(\gamma_n m_n^{1/2}). \quad (3.6.4)$$

- In the case  $0 < a < \frac{2}{9}$ , we have  $\gamma_n m_n^{1/2} = o(m_n^{-4})$  and  $\lim_{n \rightarrow +\infty} (n\gamma_n) > 2a$ .

Lemma 3.6.1 gives

$$\int_0^1 \text{Var} [f_n(x)] dx = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} o(m_k^{-4}) = o(m_n^{-4}). \quad (3.6.5)$$

Part 1 of Proposition 3.3.3 follows from the combination of (3.6.2) and (3.6.5), Part 2 from that of (3.6.2) and (3.6.4), Part 3 from (3.6.3) and (3.6.4).

### 3.6.3 Proof Corollary 3.3.1

Set

$$K_1(\xi) = \frac{C_1 C_3}{2 - (\alpha - \frac{a}{2})\xi}, \quad K_2(\xi) = \frac{4C_2}{(1 - 2\alpha\xi)^2}.$$

It follows from Proposition 3.3.3 that

$$MISE(f_n) = \begin{cases} K_2(\xi) m_n^{-4} [1 + o(1)] & \text{if } 0 < a < \frac{2}{9}\alpha \\ m_n^{-4} [K_2(\xi) + K_1(\xi) \gamma_n m_n^{9/2} + o(1)] & \text{if } a = \frac{2}{9}\alpha \\ K_1(\xi) \gamma_n m_n^{1/2} [1 + o(1)] & \text{if } a < \frac{2}{9}\alpha. \end{cases} \quad (3.6.6)$$

If  $a = \frac{2}{9}\alpha$ ,  $(K_2(\xi) m_n^{-4} + K_1(\xi) \gamma_n m_n^{1/2}) \in \mathcal{GS}(-\frac{8}{9}\alpha)$  with  $\frac{8}{9}\alpha = 4a$ . If  $0 < a < \frac{2}{9}\alpha$ ,  $(K_2(\xi) m_n^{-4}) \in \mathcal{GS}(-4a)$  with  $-4a > -\frac{8}{9}\alpha$ , and, if  $\frac{2}{9}\alpha < a < 1$ ,  $(K_1(\xi) \gamma_n m_n^{1/2}) \in \mathcal{GS}(-(2\alpha - a)/2)$  with  $-(2\alpha - a)/2 > -4a$ . It follows that, for a given  $\alpha$ , to minimize the  $MISE(f_n)$ , the parameter  $a$  must be chosen equal to  $\frac{2}{9}\alpha$ . Moreover, in view of (3.6.6) the parameter  $\alpha$  must be equal to 1. We conclude that to minimize the  $MISE(f_n)$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$  and the order  $(m_n)$  in  $\mathcal{GS}(2/9)$ .

Since the function  $x \mapsto K_2(\xi)x^{-4} + K_1(\xi)\gamma_n x^{1/2}$  attains its minimum at the point  $\left[\frac{8K_2(\xi)}{\gamma_n K_1(\xi)}\right]^{2/9}$ , to minimize  $MISE(f_n)$ , the order  $(m_n)$  must be equal to  $\left(\frac{8K_2(\xi)}{K_1(\xi)}\right)^{2/9} \gamma_n^{-2/9}$ .

For such a choice, the  $MISE$  of  $f_n$  becomes

$$MISE(f_n) = \frac{9}{8} 8^{1/9} K_1(\xi)^{8/9} K_2(\xi)^{1/9} \gamma_n^{8/9} [1 + o(1)].$$

Now, we assume that  $(\gamma_n) = (\gamma_0 n^{-1})$  for some  $\gamma_0 \in (0, \infty)$ . In this case, note that  $\xi = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1} = \gamma_0^{-1}$ . Then, (3.3.9) (hence (3.3.1)) is a consequence of the standard trade-off argument.

### 3.6.4 Proof Theorem 3.3.1 and Theorem 3.3.2

To prove Theorem 3.3.1, we will use the fact that if  $x \in (0, 1)$  and  $\frac{2}{9}\alpha \leq a \leq 1$ , then

$$\gamma_n^{-1/2} m_n^{-1/4} (f_n(x) - \mathbb{E}[f_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( 0, C_3 \frac{2}{4 - (2\alpha - a)\xi} f(x)\psi(x) \right), \quad (3.6.7)$$

which will be proved later. In the case  $0 < a < \frac{2}{9}\alpha$ , we have

$$\begin{aligned} \gamma_n^{-1/2} m_n^{-1/4} (f_n(x) - f(x)) &= \gamma_n^{-1/2} m_n^{-1/4} (f_n(x) - \mathbb{E}[f_n(x)]) + \gamma_n^{-1/2} m_n^{-1/4} (\mathbb{E}[f_n(x)] - f(x)) \\ &= \gamma_n^{-1/2} m_n^{-1/4} (f_n(x) - \mathbb{E}[f_n(x)]) - \gamma_n^{-1/2} m_n^{-9/4} \frac{2}{1 - 2a\xi} [\Delta_2(x) + o(1)], \end{aligned}$$

so if  $\gamma_n^{-1/2} m_n^{-9/4} \xrightarrow[n \rightarrow +\infty]{} c$ , for some  $c \geq 0$ , Part 1 of Theorem 3.3.1 follows from (3.6.7).

In the case when  $a = \frac{2}{9}\alpha$ , Parts 1 and 2 of Theorem 3.3.1 follow from the combination of (3.3.1) and (3.6.7). In the case  $\frac{2}{9}\alpha < a \leq 1$ , Part 1 of Theorem 3.3.1 follows from the combination of (3.3.2) and (3.6.7).

Now in the case  $\gamma_n^{-1/2} m_n^{-1/4} \xrightarrow[n \rightarrow +\infty]{} \infty$  and  $0 < a < \frac{2}{9}\alpha$ , we have

$$\begin{aligned} m_n^2 (f_n(x) - f(x)) &= m_n^2 (f_n(x) - \mathbb{E}[f_n(x)]) + m_n^2 (\mathbb{E}[f_n(x)] - f(x)) \\ &= m_n^2 (f_n(x) - \mathbb{E}[f_n(x)]) - \frac{2}{1 - a\xi} \Delta_2(x) [1 + o(1)]. \end{aligned}$$

Noting that the equation (3.3.4) implies

$$m_n^2 (f_n(x) - \mathbb{E}[f_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0,$$

then, we obtain Part 2 of Theorem 3.3.1. We now prove (3.6.7). We have

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

We set

$$Y_k(x) = \Pi_k^{-1} \gamma_k (Z_k(x) - \mathbb{E}[Z_k(x)]).$$

The application of Lemma 3.6.1 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n \text{Var}[Y_k(x)] \\ &= \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{Var}[Z_k(x)] \\ &= \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k^{1/2} [C_3 f(x) \psi(x) + o(1)] \\ &= \frac{\gamma_n}{\Pi_n^2} m_n^{1/2} \left[ \frac{2}{4 - (2\alpha - a)\xi} C_3 f(x) \psi(x) + o(1) \right]. \end{aligned}$$

On the other hand, for all  $p > 0$ , we have

$$\mathbb{E}[|Z_k(x)|^{2+p}] = O\left(m_n^{\frac{3(2+p)}{4}}\right),$$

and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - \frac{a}{2})/2 = (2\alpha - a)/4$ , there exists  $p > 0$  such that

$\lim_{n \rightarrow \infty} (n\gamma_n) > \frac{1+p}{2+p}(\alpha - \frac{a}{2}) = \frac{(1+p)\alpha - \frac{1+p}{2}a}{p+2} > \frac{(1+p)\alpha - \frac{3(2+p)}{4}a}{p+2}$ , so Lemma 3.6.1 gives

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}[|Y_k(x)|^{2+p}] &= O\left(\sum_{k=1}^n \Pi_k^{-2-p} \gamma_k^{2+p} \mathbb{E}[|Z_k(x)|^{2+p}]\right) \\ &= O\left(\sum_{k=1}^n \frac{\gamma_k^{2+p}}{\Pi_k^{2+p}} m_k^{\frac{3(2+p)}{4}}\right) \\ &= O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}} m_n^{\frac{3(2+p)}{4}}\right), \end{aligned} \tag{3.6.8}$$

hence

$$\frac{1}{v_n^{2+p}} \sum_{k=1}^n \mathbb{E}[|Y_k(x)|^{2+p}] = O\left(m_n (\gamma_n m_n)^{p/2}\right).$$

Then the convergence in (3.6.7) follows from the application of Lyapounov's Theorem.

To prove Theorem 3.3.2, we will use the fact that if  $x = 0, 1$  and  $\frac{\alpha}{5} \leq a \leq 1$ , then

$$\gamma_n^{-1/2} m_n^{-1/2} (f_n(x) - \mathbb{E}[f_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{5}{2(2 - (\alpha - a)\xi)} f(x)\right), \quad (3.6.9)$$

which will be proved later. In the case  $0 < a < \frac{\alpha}{5}$ , we have

$$\begin{aligned} \gamma_n^{-1/2} m_n^{-1/2} (f_n(x) - f(x)) &= \gamma_n^{-\frac{1}{2}} m_n^{-1/2} (f_n(x) - \mathbb{E}[f_n(x)]) + \gamma_n^{-1/2} m_n^{-1/2} (\mathbb{E}[f_n(x)] - f(x)) \\ &= \gamma_n^{-1/2} m_n^{-1/4} (f_n(x) - \mathbb{E}[f_n(x)]) - \gamma_n^{-1/2} m_n^{-5/2} \frac{2}{1 - 2a\xi} [\Delta_2(x) + o(1)], \end{aligned}$$

so if  $\gamma_n^{-1/2} m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} c$ , for some  $c \geq 0$ , Part 1 of Theorem 3.3.2 follows from (3.6.9).

In the case when  $a = \frac{\alpha}{5}$ , Parts 1 and 2 of Theorem 3.3.2 follow from the combination of (3.3.5) and (3.6.9). In the case  $\frac{\alpha}{5} < a \leq 1$ , Part 1 of Theorem 3.3.2 follows from the combination of (3.3.6) and (3.6.9).

Now in the case  $\gamma_n^{-1/2} m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} \infty$  and  $0 < a < \frac{\alpha}{5}$ , we have

$$\begin{aligned} m_n^2 (f_n(x) - f(x)) &= m_n^2 (f_n(x) - \mathbb{E}[f_n(x)]) + m_n^2 (\mathbb{E}[f_n(x)] - f(x)) \\ &= m_n^2 (f_n(x) - \mathbb{E}[f_n(x)]) - \frac{2}{1 - a\xi} \Delta_2(x) [1 + o(1)]. \end{aligned}$$

Noting that the equation (3.3.8) implies

$$m_n^2 (f_n(x) - \mathbb{E}[f_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0,$$

then, we obtain Part 2 of Theorem 3.3.2. We now prove (3.6.9). The application of Lemma 3.6.1 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n \text{Var}[Y_k(x)] \\ &= \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{Var}[Z_k(x)] \\ &= \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 m_k \left[ \frac{5}{2} f(x) + o(1) \right] \\ &= \frac{\gamma_n}{\Pi_n^2} m_n \left[ \frac{5}{2(2 - (\alpha - a)\xi)} f(x) + o(1) \right]. \end{aligned}$$

On the other hand, for all  $p > 0$ , we have

$$\mathbb{E}[|Z_k(x)|^{2+p}] = O\left(m_n^{\frac{3(2+p)}{4}}\right),$$

and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a)/2$ , there exists  $p > 0$  such that

$\lim_{n \rightarrow \infty} (n\gamma_n) > \frac{1+p}{2+p}(\alpha - a) > \frac{(1+p)\alpha - \frac{3(2+p)}{4}a}{p+2}$ , so Lemma 3.6.1 gives

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}[|Y_k(x)|^{2+p}] &= O\left(\sum_{k=1}^n \Pi_k^{-2-p} \gamma_k^{2+p} \mathbb{E}[|Z_k(x)|^{2+p}]\right) \\ &= O\left(\sum_{k=1}^n \frac{\gamma_k^{2+p}}{\Pi_k^{2+p}} m_k^{\frac{3(2+p)}{4}}\right) \\ &= O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}} m_n^{\frac{3(2+p)}{4}}\right), \end{aligned} \tag{3.6.10}$$

hence

$$\frac{1}{v_n^{2+p}} \sum_{k=1}^n \mathbb{E}[|Y_k(x)|^{2+p}] = O\left(m_n^{1/2} (\gamma_n m_n^{1/2})^{p/2}\right).$$

Then the convergence in (3.6.9) follows from the application of Lyapounov's Theorem.



Case	n	Vitale's estimator	Recursive 1	Recursive 2	Recursive 3
(a)	50	<b>0.085389</b>	0.088252	0.089350	0.092738
	200	0.028167	<b>0.025737</b>	0.026057	0.027045
	500	0.013533	<b>0.011398</b>	0.011540	0.011977
(b)	50	0.145441	<b>0.129385</b>	0.130996	0.135963
	200	0.047977	<b>0.037733</b>	0.038202	0.039651
	500	0.023050	<b>0.016710</b>	0.016918	0.017560
(c)	50	0.074634	<b>0.061163</b>	0.061925	0.064273
	200	0.024620	<b>0.017837</b>	0.018059	0.0187441
	500	0.011828	<b>0.007899</b>	0.007997	0.0083012
(d)	50	<b>0.108664</b>	0.119492	0.120980	0.125567
	200	0.035845	<b>0.034847</b>	0.035281	0.036619
	500	0.017222	<b>0.015433</b>	0.015625	0.016217
(e)	50	<b>0.124816</b>	0.157438	0.159398	0.165442
	200	<b>0.041174</b>	0.045914	0.046485	0.048248
	500	<b>0.019782</b>	0.020333	0.020587	0.021367
(f)	50	<b>0.084119</b>	0.109237	0.110596	0.114790
	200	<b>0.027749</b>	0.031857	0.032253	0.033476
	500	<b>0.013332</b>	0.014108	0.014284	0.014825
(g)	50	<b>0.063962</b>	0.065611	0.066428	0.068947
	200	0.021099	<b>0.019134</b>	0.019372	0.020107
	500	0.010137	<b>0.008474</b>	0.008579	0.008904
(h)	50	<b>0.046147</b>	0.042890	0.043424	0.04507
	200	0.015222	<b>0.012508</b>	0.012664	0.013144
	500	0.007313	<b>0.005539</b>	0.005608	0.005821
(i)	50	<b>0.031391</b>	0.036611	0.037066	0.038472
	200	<b>0.010355</b>	0.010676	0.010809	0.011219
	500	0.004975	<b>0.004728</b>	0.004787	0.004968
(j)	50	0.071621	<b>0.065671</b>	0.066489	0.069010
	200	0.023626	<b>0.019152</b>	0.019390	0.020125
	500	0.011351	<b>0.008481</b>	0.008587	0.008913

TABLE 3.1 – The average integrated squared error (*ISE*) of Vitale's estimator  $\tilde{f}_n$  and the three recursive estimators; **recursive 1** correspond to the estimator  $f_n$  with the choice  $(\gamma_n) = (n^{-1})$ , **recursive 2** correspond to the estimator  $f_n$  with the choice  $(\gamma_n) = ([1 - \frac{a}{2}] n^{-1})$  ( $a = 2/9$ ) and **recursive 3** correspond to the estimator  $f_n$  with the choice  $(\gamma_n) = ([1 - a] n^{-1})$  ( $a = 1/5$ ).

Case	n	$\tilde{f}_{n,m,m/2}$	$\tilde{f}_{n,m,b}$		$\tilde{f}_{n,m,b,\varepsilon}, \varepsilon = 0.00001$			$\tilde{f}_{n,m,b,\varepsilon}^N, \varepsilon = 0.00001$		
			b = 3	b = 4	b = 2	b = 3	b = 4	b = 2	b = 3	b = 4
(a)	50	<b>0.08504</b>	0.08687	0.08874	0.10597	0.10824	0.11057	0.09852	0.10064	0.10280
	200	<b>0.02480</b>	0.02533	0.02587	0.03090	0.03156	0.03224	0.02873	0.02935	0.02998
	500	<b>0.01098</b>	0.01122	0.01146	0.01368	0.01398	0.01428	0.01272	0.01299	0.01327
(b)	50	<b>0.12469</b>	0.12736	0.13010	0.13284	0.13569	0.13861	0.14319	0.14626	0.14940
	200	<b>0.03636</b>	0.03714	0.03794	0.03874	0.03957	0.04042	0.04176	0.04265	0.04357
	500	<b>0.01610</b>	0.01644	0.01680	0.01715	0.01752	0.01790	0.01849	0.01889	0.01929
(c)	50	<b>0.05894</b>	0.06020	0.06150	0.08278	0.08455	0.08637	0.08721	0.08908	0.09100
	200	<b>0.01719</b>	0.01755	0.01793	0.02414	0.02465	0.02518	0.02543	0.02598	0.02653
	500	<b>0.00761</b>	0.00777	0.00794	0.01069	0.01092	0.01115	0.01126	0.01150	0.01175
(d)	50	<b>0.11515</b>	0.11762	0.12015	0.14391	0.14699	0.15015	0.14656	0.14971	0.15292
	200	<b>0.03358</b>	0.03430	0.03504	0.04196	0.04286	0.04379	0.04274	0.04366	0.04459
	500	<b>0.01487</b>	0.01519	0.01551	0.01858	0.01898	0.01939	0.01892	0.01933	0.01975
(e)	50	<b>0.15172</b>	0.15498	0.15830	0.15311	0.15640	0.15976	0.15463	0.15795	0.16134
	200	<b>0.04424</b>	0.04519	0.04616	0.04465	0.04561	0.04659	0.04509	0.04606	0.04705
	500	<b>0.01959</b>	0.02001	0.02044	0.01977	0.02020	0.02063	0.01997	0.02040	0.02083
(f)	50	0.10527	0.10753	0.10984	<b>0.10527</b>	0.11383	0.11627	0.11420	0.11665	0.11916
	200	<b>0.03070</b>	0.03135	0.03203	0.03250	0.03319	0.03391	0.03330	0.03402	0.03475
	500	<b>0.01359</b>	0.01388	0.01418	0.01439	0.01470	0.01501	0.01475	0.01506	0.01539
(g)	50	0.06323	0.06458	0.06597	0.06298	0.06433	0.06571	<b>0.06252</b>	0.06386	0.06524
	200	0.01844	0.01883	0.01924	0.01836	0.01876	0.01916	<b>0.01823</b>	0.01862	0.01902
	500	0.00816	0.00834	0.00852	0.00813	0.00830	0.00848	<b>0.00807</b>	0.00824	0.00842
(h)	50	0.04133	0.04222	0.04312	0.03872	0.03955	0.04040	<b>0.03566</b>	0.03643	0.03721
	200	0.01205	0.01231	0.01257	0.01129	0.01153	0.01178	<b>0.01040</b>	0.01062	0.01085
	500	0.00533	0.00545	0.00557	0.00500	0.00510	0.00521	<b>0.00460</b>	0.00470	0.00480
(i)	50	<b>0.03528</b>	0.03603	0.03681	0.03594	0.03671	0.03750	0.03589	0.03666	0.03745
	200	<b>0.01028</b>	0.01051	0.01073	0.01048	0.01070	0.01093	0.01046	0.01069	0.01092
	500	<b>0.00455</b>	0.00465	0.00475	0.00464	0.00474	0.00484	0.00463	0.00473	0.00483
(j)	50	0.06328	0.06464	0.06603	0.06439	0.06577	0.06718	<b>0.06236</b>	0.06370	0.06507
	200	0.01845	0.01885	0.01925	0.01877	0.01918	0.01959	<b>0.01818</b>	0.01857	0.01897
	500	0.00817	0.00817	0.00852	0.00831	0.00849	0.00867	<b>0.00805</b>	0.00822	0.00840

TABLE 3.2 – The average integrated squared error (*ISE*) of Leblanc estimator's  $\tilde{f}_{n,m,m/2}$  and the three estimators introduced by Kakizawa :  $\tilde{f}_{n,m,m/b}$  with  $b = 3, 4$ ,  $\tilde{f}_{n,m,b,\varepsilon}$  and  $\tilde{f}_{n,m,b,\varepsilon}^N$  with  $b = 2, 3, 4$  and  $\varepsilon = 0.00001$ .

FIGURE 3.1 – Qualitative comparison between the proposed density estimator  $f_n$  given in (3.1.2) with stepsize  $(\gamma_n) = (n^{-1})$  (solid red line), the Vitale’s estimator  $\tilde{f}_n$  defined in (3.1.1) (dashed magenta line), Leblanc’s estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12) (dotted blue line), the generalized estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14) (dotdashed cyan line), Kakizwa’s estimator  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17) (longdashed green line) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) (twodashed yellow line) with  $b = 2$  and  $\varepsilon = 0.00001$  for 500 samples respectively of size 50 (left panel) and of size 250 (right panel) of the beta density  $\mathcal{B}(3, 5)$ .

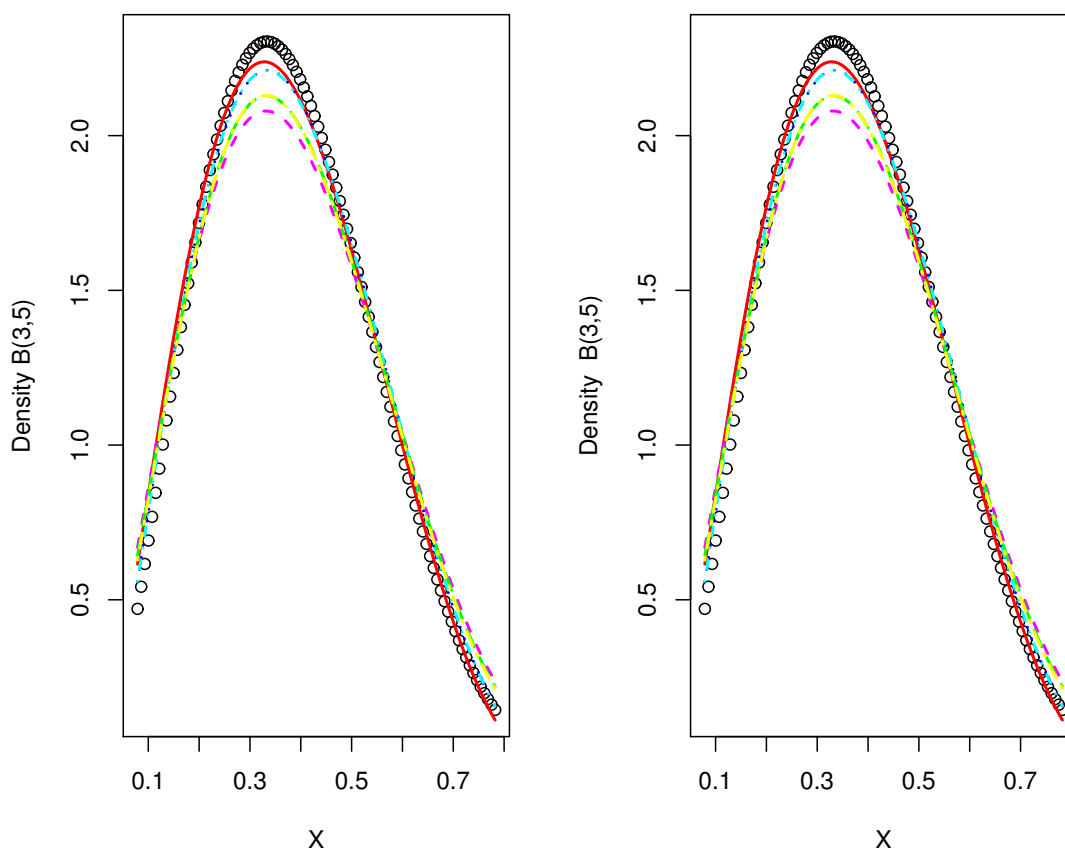


FIGURE 3.2 – Density estimates for the prostate-specific antigen data : recursive estimator  $f_n$  defined in (3.1.2) (solid red line), Vitale’s estimator  $\tilde{f}_n$  defined in (3.1.1) (dashed magenta line), Leblanc’s estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12) (dotted blue line), the generalized estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14) (dotdashed cyan line), Kakizwa’s estimator  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17) (longdashed green line) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) (twodashed yellow line) with  $b = 2$  and  $\varepsilon = 0.00001$  (left panel) and Gaussian kernel density estimate using the *LSCV*-based bandwidth  $h = 0.1111$  (right panel).

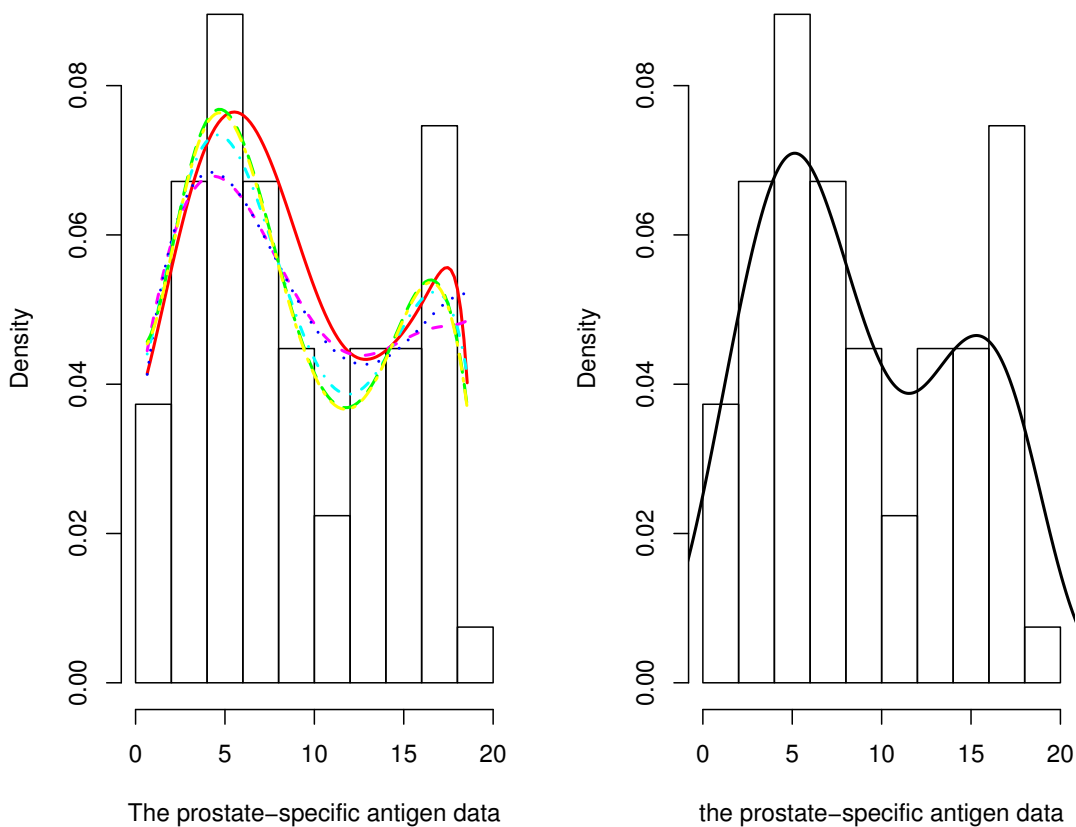


FIGURE 3.3 – Density estimates for the Old Faithful data : recursive estimator  $f_n$  defined in (3.1.2) (solid red line), Vitale’s estimator  $\tilde{f}_n$  defined in (3.1.1) (dashed magenta line), Leblanc’s estimator defined in  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12) (dotted blue line), the generalized estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14) (dotdashed cyan line), Kakizwa’s estimator  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17) (longdashed green line) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) (twodashed yellow line) with  $b = 2$  and  $\varepsilon = 0.00001$  (left panel) and Gaussian kernel density estimate using the *LSCV*-based bandwidth  $h = 0.3677$  (right panel).

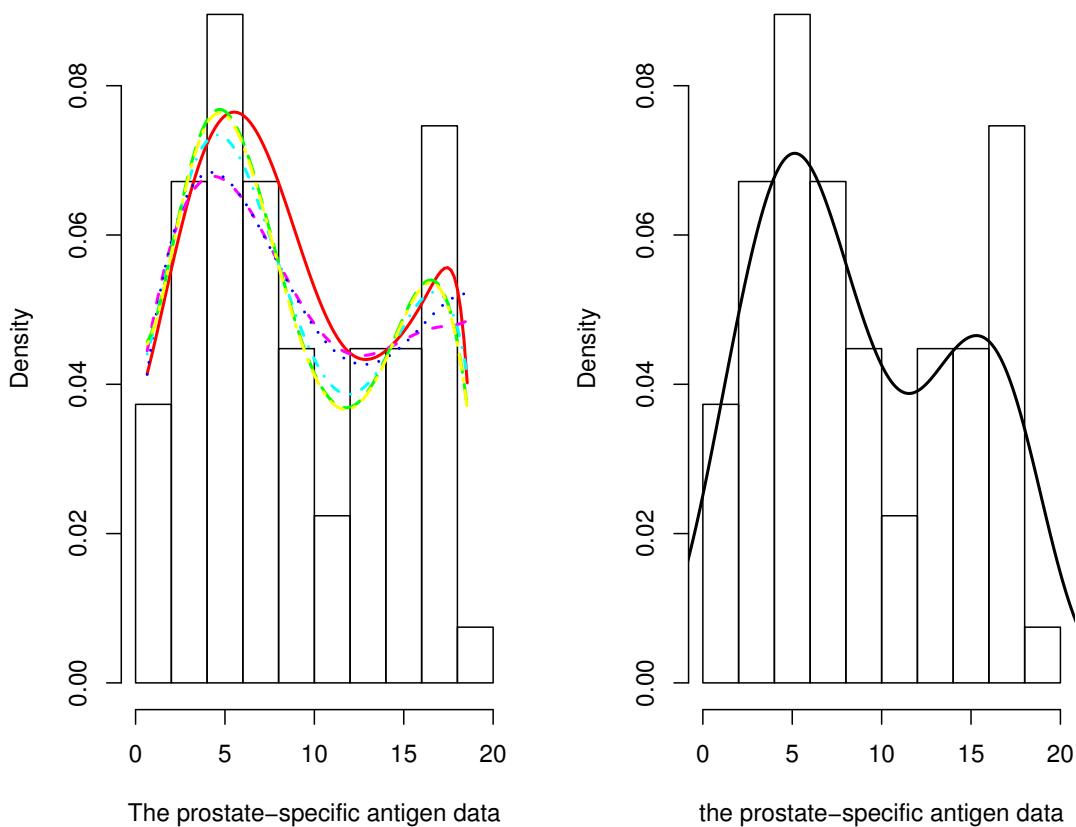
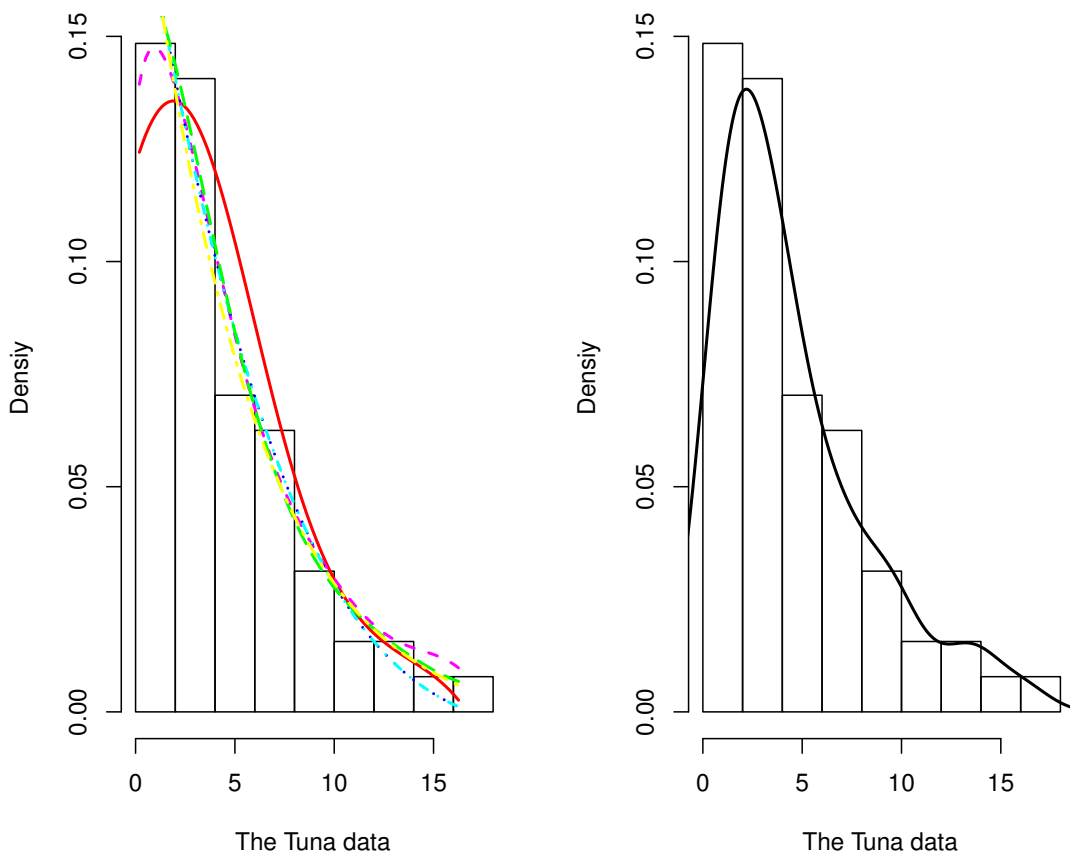


FIGURE 3.4 – Density estimates for the tuna data : recursive estimator  $f_n$  defined in (3.1.2) (solid red line), Vitale’s estimator  $\tilde{f}_n$  defined in (3.1.1) (dashed magenta line), Leblanc’s estimator  $\tilde{f}_{n,m,m/2}$  defined in (3.3.12) (dotted blue line), the generalized estimator  $\tilde{f}_{n,m,m/4}$  defined in (3.3.14) (dotdashed cyan line), Kakizwa’s estimator  $\tilde{f}_{n,m,b,\varepsilon}$  defined in (3.3.17) (longdashed green line) and the normalized estimator  $\tilde{f}_{n,m,b,\varepsilon}^N$  defined in (3.3.19) (twodashed yellow line) with  $b = 2$  and  $\varepsilon = 0.00001$  (left panel) and Gaussian kernel density estimate using the *LSCV*-based bandwidth  $h = 1.291$  (right panel)



# Chapitre 4

## Regression estimators using Bernstein polynomials

### 4.1 Introduction

The goal in any data analysis is to extract from raw information the accurate estimation. One of the most important and common questions concerning if there is a statistical relationship between a response variable ( $Y$ ) and explanatory variables ( $X_i$ ). An option to answer this question is to employ regression analysis in order to model this relationship.

There were many ways to estimate the regression function  $r : x \mapsto \mathbb{E}(Y|X = x)$ . The most known are the kernel regression estimators. On the non-recursive approach, we refer, among many others, to the estimator proposed by Nadaraya [72] and Watson [120], the alternative kernel estimators given by Priestley et Chao [83] and Gasser et Müller [31]. On the other hand, the recursive estimation were widely discussed, we refer to the approach of Révész [85, 86] and Tsybakov [117] which was studied by Mokkadem et al. [67], Slaoui [106, 107, 108], also we find the semi-recursive approach introduced by Slaoui [109]. Each of these estimators has its own particular strengths and weaknesses.

However, the common problem is the edge effect. In fact, when the regression function has bounded support, kernel estimates often overspill the boundaries and are consequently biased at and near these edges. To overcome this problem, we propose a non-recursive and recursive approach of regression estimation using Bernstein polynomials.

The estimation using Bernstein polynomial for density and distribution functions have been widely discussed in several frameworks. See, for instance, the original work of Vitale [119] and extensions given by Tenbusch [115], Ghosa [33], Kakizawa [47, 48], Igarashi and Kakizawa [42], Rao [84], Leblanc [55], Leblanc [56, 57], Babu et al. [2], Babu et Chaubey [3] and Jmaei et al. [43].

The remainder of the chapter is organized as follows. In Section 4.2, we list our assumptions and notations. In Section 4.3, we introduce our non-recursive estimator and we compute its bias, variance, mean squared error ( $MSE$ ), the mean integrated squared error ( $MISE$ ) and we establish a weak convergence rate. In Section 4.4 we introduce our non-recursive estimator and we state the main theoretical results. Section 4.5 is devoted to some numerical studies : first, a simulation study is presented in Subsection 4.5.1 and, then, an application to a real dataset is described in Subsection 4.5.2. Finally, we discuss our conclusion in Section 4.6. The proofs are given in 4.7.

## 4.2 Assumptions and Notations

Let us first define the class of positive sequences that will be used in the statement of our assumptions.

### **Definition 4.2.1.**

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



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

This condition was introduced by Galambos et Seneta [30] to define regularly varying sequences (see also Bojanic and Seneta [7]). 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 obtain the behavior of our estimators, we make to the following assumptions :

(A1) The functions  $r$  and  $f$  are continuous and admit two continuous and bounded derivatives.

(A2)  $(\gamma_n) \in \mathcal{GS}(-\alpha)$ ,  $\alpha \in (\frac{3}{4}, 1]$ .

(A3)  $(m_n) \in \mathcal{GS}(a)$ ,  $a \in (\frac{1-\alpha}{4}, \frac{2}{3}\alpha)$ .

Assumptions (A1) and (A2) are standard on regression estimation. Assumption (A3) on  $(m_n)$  was introduced similarly to the assumption on the bandwidth used for the recursive kernel regression estimator (see Mokkadem et al. [66, 67]).

Throughout this chapter we will use the following notations :

$$\begin{aligned} \Delta_1(x) &= \frac{1}{2} [(1-2x)f'(x) + x(1-x)f''(x)], \quad \psi(x) = (4\pi x(1-x))^{-1/2}, \quad \xi = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}, \\ \Delta_2(x) &= \frac{1}{2} \left\{ (1-2x)(r'(x)f(x) + r(x)f'(x)) + x(1-x) \left( r''(x)f(x) + f''(x)r(x) \right. \right. \\ &\quad \left. \left. + 2r'(x)f'(x) \right) \right\}, \\ \Delta(x) &= \frac{1}{2} \left\{ x(1-x)r''(x) + \left[ (1-2x) + 2x(1-x)\frac{f'(x)}{f(x)} \right] r'(x) \right\}, \\ C_1 &= \int_0^1 \Delta^2(x)dx, \quad C_2 = \int_0^1 \frac{Var[Y|X=x]}{f(x)} \psi(x)dx, \\ K_1 &= \int_0^1 \left\{ \frac{\Delta(x)f(x)}{f(x) - a\xi} \right\}^2 dx, \quad K_2 = \int_0^1 \frac{2f(x)}{4f(x) - (2\alpha - a)\xi} Var(Y|X=x)\psi(x)dx. \end{aligned}$$

### 4.3 Non-Recursive estimator

Let  $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$  be independent, identically distributed pairs of random variables with joint density function  $g(x, y)$ , and let  $f$  denote the probability density of  $X$  which is supported on  $[0, 1]$ . We follow the approach of Vitale [119] and Leblanc [55, 56], used for distribution and density estimation, to define a Bernstein estimator of the regression  $r : x \mapsto \mathbb{E}(Y|X = x)$  at a given point  $x \in [0, 1]$  such that  $f(x) \neq 0$

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x)}{\sum_{i=1}^n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x)}, \quad (4.3.1)$$

where  $b_k(m, x) = \binom{m}{k} x^k (1-x)^{m-k}$  is the Bernstein polynomial of order  $m$ .

The following proposition gives the bias, the variance and the  $MSE$  of  $\hat{r}_n(x)$ , for  $x \in [0, 1]$  such that  $f(x) > 0$ .

**Proposition 4.3.1.**

*Let Assumption (A1) hold. For  $x \in [0, 1]$ , such that  $f(x) > 0$ , we have*

$$\mathbb{E}[\hat{r}_n(x)] - r(x) = \Delta(x)m^{-1} + o(m^{-1}). \quad (4.3.2)$$

$$\text{Var}[\hat{r}_n(x)] = \begin{cases} \frac{m^{1/2}}{n} \frac{\text{Var}(Y|X=x)}{f(x)} \psi(x) + o_x\left(\frac{m^{1/2}}{n}\right) & \text{if } x \in (0, 1), \\ \frac{m}{n} \frac{\text{Var}(Y|X=x)}{f(x)} + o_x\left(\frac{m}{n}\right) & \text{if } x = 0, 1. \end{cases} \quad (4.3.3)$$

$$MSE[\hat{r}_n(x)] = \begin{cases} \Delta^2(x)m^{-2} + \frac{m^{1/2}}{n} \frac{\text{Var}(Y|X=x)}{f(x)} \psi(x) + o(m^{-2}) + o_x\left(\frac{m^{1/2}}{n}\right) & \text{if } x \in (0, 1), \\ \Delta^2(x)m^{-2} + \frac{m}{n} \frac{\text{Var}(Y|X=x)}{f(x)} + o(m^{-2}) + o_x\left(\frac{m}{n}\right) & \text{if } x = 0, 1. \end{cases}$$

To minimize the  $MSE$  of  $\hat{r}_n$ , for  $x \in [0, 1]$  such that  $f(x) > 0$ , the order  $m$  must equal to

$$m_{opt} = \begin{cases} \left[ \frac{4\Delta^2(x)f(x)}{\text{Var}(Y|X=x)\psi(x)} \right]^{2/5} n^{2/5} & \text{if } x \in (0, 1), \\ \left[ \frac{2\Delta^2(x)f(x)}{\text{Var}(Y|X=x)} \right]^{1/3} n^{1/3} & \text{if } x = 0, 1, \end{cases}$$

then

$$MSE [\hat{r}_{n,m_{opt}}(x)] = \begin{cases} \frac{5(\Delta(x))^{2/5}(\text{Var}(Y|X=x)\psi(x))^{4/5}}{(4f(x))^{4/5}} n^{-4/5} + o(n^{-4/5}) & \text{if } x \in (0, 1), \\ \frac{3(\Delta(x)\text{Var}(Y|X=x))^{2/3}}{(2f(x))^{2/3}} n^{-2/3} + o(n^{-2/3}) & \text{if } x = 0, 1. \end{cases}$$

The following proposition gives the  $MISE$  of  $\hat{r}_n$

**Proposition 4.3.2.**

Let Assumption (A1) hold, we have

$$MISE(\hat{r}_n) = C_2 \frac{m^{1/2}}{n} + C_1 m^{-2} + o\left(\frac{m^{1/2}}{n}\right) + o(m^{-2}). \quad (4.3.4)$$

Hence, the asymptotically optimal choice of  $m$  is

$$m_{opt} = \left[ \frac{4C_1}{C_2} \right]^{2/5} n^{2/5},$$

for which we get

$$MISE(\hat{r}_{n,m_{opt}}) = \frac{5C_1^{1/5}C_2^{4/5}}{4^{4/5}} n^{-4/5} + o(n^{-4/5}).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator  $\hat{r}_n(x)$  defined in (4.3.1), for  $x \in [0, 1]$  such that  $f(x) > 0$ .

**Theorem 4.3.1.** (Weak pointwise convergence rate).

Let Assumption (A1) hold.

For  $x \in (0, 1)$ , we have

1. If  $nm^{-5/2} \xrightarrow{n \rightarrow +\infty} c$  for some constant  $c \geq 0$ , then

$$n^{1/2}m^{-1/4}(\hat{r}_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\Delta(x), \frac{\text{Var}(Y|X=x)\psi(x)}{f(x)}\right).$$

2. If  $nm^{-5/2} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m(\widehat{r}_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0\Delta(x),$$

where  $\xrightarrow[n \rightarrow +\infty]{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the Gaussian-distribution and  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$  the convergence in probability.

For  $x = 0, 1$ , we have

1. If  $nm^{-3} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\sqrt{\frac{n}{m}}(\widehat{r}_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\Delta(x), \frac{\text{Var}(Y|X=x)}{f(x)}\right).$$

2. If  $nm^{-3} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m(\widehat{r}_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \Delta(x).$$

## 4.4 Recursive estimator

In order to construct a stochastic algorithm for the estimation of the regression function  $r : x \mapsto \mathbb{E}(Y|X=x)$  at a point  $x$  such as  $f(x) \neq 0$ , Révész [85] defines an algorithm, which approximates the zero of the function  $h : y \mapsto f(x)r(x) - f(x)y$ . Using the procedure of Robbins et Monro [87], the proposed algorithm is defined by setting  $r_0(x) \in \mathbb{R}$  and for  $n \geq 1$

$$r_n(x) = r_{n-1}(x) + \gamma_n \mathcal{W}_n(x),$$

where  $\gamma_n$  is the stepsize and  $\mathcal{W}_n$  is an observation of the function  $h$  at the point  $r_{n-1}(x)$ . We define  $\mathcal{W}_n$ , using Bernstein polynomials

$$\mathcal{W}_n(x) = m_n Y_n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x) - m_n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x) r_{n-1}(x),$$

then, the estimator  $r_n$  can be rewritten as

$$\begin{aligned}
r_n(x) &= \left( 1 - \gamma_n m_n \sum_{k=0}^{m-1} \mathbb{I}_{\left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\}} b_k(m-1, x) \right) r_{n-1}(x) \\
&\quad + \gamma_n m_n Y_n \sum_{k=0}^{m-1} \mathbb{I}_{\left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\}} b_k(m-1, x), \\
&= (1 - \gamma_n f(x)) r_{n-1}(x) + \gamma_n \left( f(x) - m_n \sum_{k=0}^{m-1} \mathbb{I}_{\left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\}} b_k(m-1, x) \right) r_{n-1}(x) \\
&\quad + \gamma_n m_n Y_n \sum_{k=0}^{m-1} \mathbb{I}_{\left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\}} b_k(m-1, x).
\end{aligned}$$

We set

$$\begin{aligned}
Z_n(x) &= m_n \sum_{k=0}^{m-1} \mathbb{I}_{\left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\}} b_k(m-1, x), \\
W_n(x) &= m_n Y_n \sum_{k=0}^{m-1} \mathbb{I}_{\left\{ \frac{k}{m} < X_n \leq \frac{k+1}{m} \right\}} b_k(m-1, x).
\end{aligned}$$

Then, the proposed algorithm can be rewritten as follows :

$$r_n(x) = (1 - \gamma_n f(x)) r_{n-1}(x) + \gamma_n (f(x) - Z_n(x)) r_{n-1}(x) + \gamma_n W_n(x). \quad (4.4.1)$$

In order, to establish the behaviour of  $r_n$ , we introduce the auxiliary stochastic approximation algorithm defined by setting  $\rho_n(x) = r(x)$  for all  $n \leq n_0 - 2$ ,  $\rho_{n_0-1}(x) = r_{n_0-1}(x)$ , and, for  $n \geq n_0$ ,

$$\rho_n(x) = (1 - \gamma_n f(x)) \rho_{n-1}(x) + \gamma_n (f(x) - Z_n(x)) r(x) + \gamma_n W_n(x). \quad (4.4.2)$$

We first give the behaviour of  $\rho_n$ . Then, we show how the behaviour of  $r_n$  can be deduced from that of  $\rho_n$ .

#### 4.4.1 Within the interval $[0, 1]$

To obtain the bias, the variance and the  $MSE$  of  $r_n(x)$ , for  $x \in (0, 1)$  such that  $f(x) > 0$ , we set

$$(A4) \lim_{n \rightarrow \infty} (n\gamma_n) \in \left( \min \left( \frac{a}{f(x)}, \frac{2\alpha - a}{4f(x)} \right), \infty \right].$$

**Proposition 4.4.1.**

Let Assumptions (A1) – (A4) hold. For  $x \in (0, 1)$ , such that  $f(x) > 0$ , we have

1. If  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , then

$$\mathbb{E}[r_n(x)] - r(x) = m_n^{-1} \frac{f(x)}{f(x) - a\xi} \Delta(x) + o(m_n^{-1}). \quad (4.4.3)$$

If  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , then

$$\mathbb{E}[r_n(x)] - r(x) = o\left(\sqrt{\gamma_n m_n^{1/2}}\right).$$

2. If  $\frac{2}{5}\alpha \leq a < \frac{2}{3}\alpha$ , then

$$\begin{aligned} \text{Var}[r_n(x)] &= \frac{2f(x)}{4f(x) - (2\alpha - a)\xi} \text{Var}(Y|X = x)\psi(x)\gamma_n m_n^{1/2} \\ &\quad + o(\gamma_n m_n^{1/2}). \end{aligned} \quad (4.4.4)$$

If  $\frac{1-\alpha}{4} < a < \frac{2}{5}\alpha$ , then

$$\text{Var}[r_n(x)] = o(m_n^{-2}).$$

3. If  $\lim_{n \rightarrow \infty} (n\gamma_n) > \max\left(\frac{a}{f(x)}, \frac{2\alpha - a}{4f(x)}\right)$ , then (4.4.3) and (4.4.4) hold simultaneously.

4. If  $\frac{1-\alpha}{4} < a < \frac{2}{5}\alpha$ , then

$$\text{MSE}[r_n(x)] = \Delta^2(x)m_n^{-2} \frac{f^2(x)}{(f(x) - a\xi)^2} + o(m_n^{-2}).$$

If  $a = \frac{2}{5}\alpha$ , then

$$\begin{aligned} \text{MSE}[r_n(x)] &= \Delta^2(x)m_n^{-2} \frac{f^2(x)}{(f(x) - a\xi)^2} \\ &\quad + \frac{2f(x)}{4f(x) - (2\alpha - a)\xi} \text{Var}(Y|X = x)\psi(x)\gamma_n m_n^{1/2} + o(m_n^{-2} + \gamma_n m_n^{1/2}). \end{aligned}$$

If  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , then

$$\text{MSE}[r_n(x)] = \frac{2f(x)}{4f(x) - (2\alpha - a)\xi} \text{Var}(Y|X = x)\psi(x)\gamma_n m_n^{1/2} + o(\gamma_n m_n^{1/2}).$$

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

$$\left(4^{3/5} \left(f(x) - \frac{2}{5}\xi\right)^{-2/5} \left[\frac{f(x)\Delta^2(x)}{\text{Var}(Y|X = x)\psi(x)}\right]^{2/5} \gamma_n^{-2/5}\right),$$

then

$$MSE[r_n(x)] = \frac{5(f(x))^{6/5}(\Delta(x))^{2/5}(\text{Var}(Y|X=x)\psi(x))^{4/5}}{4^{6/5}\left(f(x) - \frac{2}{5}\xi\right)^{6/5}}\gamma_n^{4/5} + o(\gamma_n^{4/5}).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator  $r_n(x)$  defined in (4.4.1), for  $x \in (0, 1)$  such that  $f(x) > 0$ .

**Theorem 4.4.1.** (*Weak pointwise convergence rate*).

Let Assumption (A1)-(A4) hold, we have

1. If  $\gamma_n^{-1}m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2}m_n^{-1/4}(r_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\frac{f(x)\Delta(x)}{f(x) - a\xi}, \frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi}\right).$$

2. If  $\gamma_n^{-1}m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n(r_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{f(x)\Delta(x)}{f(x) - a\xi},$$

where  $\xrightarrow[n \rightarrow +\infty]{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the Gaussian-distribution and  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}}$  the convergence in probability.

#### 4.4.2 The edges of the interval $[0, 1]$

For the cases  $x = 0, 1$ , such that  $f(x) > 0$ , we need to consider the additional Assumption

$$(A'4) \quad \lim_{n \rightarrow \infty} (n\gamma_n) \in \left(\min\left(\frac{a}{f(x)}, \frac{\alpha - a}{2f(x)}\right), \infty\right].$$

The following proposition gives the bias, the variance and the  $MSE$  of  $r_n(x)$ , for  $x = 0, 1$ .

**Proposition 4.4.2.**

Let Assumptions (A1) – (A3) and (A'4) hold. For  $x = 0, 1$ , such that  $f(x) > 0$ , we have

1. If  $\frac{1-\alpha}{4} < a \leq \frac{\alpha}{3}$ , then

$$\mathbb{E}[r_n(x)] - r(x) = m_n^{-1} \frac{f(x)}{f(x) - a\xi} \Delta(x) + o(m_n^{-1}). \quad (4.4.5)$$

If  $\frac{\alpha}{3} < a < \frac{2}{3}\alpha$ , then

$$\mathbb{E}[r_n(x)] - r(x) = o(\sqrt{\gamma_n m_n}).$$

2. If  $\frac{\alpha}{3} \leq a < \frac{2}{3}\alpha$ , then

$$\text{Var}[r_n(x)] = \frac{f(x)}{2f(x) - (\alpha - a)\xi} \text{Var}(Y|X = x) \gamma_n m_n + o(\gamma_n m_n). \quad (4.4.6)$$

If  $\frac{1-\alpha}{4} < a < \frac{\alpha}{3}$ , then

$$\text{Var}[r_n(x)] = o(m_n^{-2}).$$

3. If  $\lim_{n \rightarrow \infty} (n\gamma_n) > \max\left(\frac{a}{f(x)}, \frac{\alpha - a}{2f(x)}\right)$ , then (4.4.5) and (4.4.6) hold simultaneously.

4. If  $\frac{1-\alpha}{4} < a < \frac{\alpha}{3}$ , then

$$\text{MSE}[f_n(x)] = \Delta^2(x) m_n^{-2} \frac{f^2(x)}{(f(x) - a\xi)^2} + o(m_n^{-2}).$$

If  $a = \frac{\alpha}{3}$ , then

$$\begin{aligned} \text{MSE}[r_n(x)] &= \Delta^2(x) m_n^{-2} \frac{f^2(x)}{(f(x) - a\xi)^2} + \frac{f(x)}{2f(x) - (\alpha - a)\xi} \text{Var}(Y|X = x) \gamma_n m_n \\ &\quad + o(m_n^{-2} + \gamma_n m_n). \end{aligned}$$

If  $\frac{\alpha}{3} < a < \frac{2}{3}\alpha$ , then

$$\text{MSE}[r_n(x)] = \frac{f(x)}{2f(x) - (\alpha - a)\xi} \text{Var}(Y|X = x) \gamma_n m_n + o(\gamma_n m_n).$$

To minimize the  $\text{MSE}$  of  $r_n$ , for  $x = 0, 1$  such that  $f(x) > 0$ , the stepsize ( $\gamma_n$ ) must be chosen in  $\mathcal{GS}(-1)$  and ( $m_n$ ) must be in  $\mathcal{GS}(1/3)$  such that

$$\left( 2^{2/3} \left( f(x) - \frac{1}{3}\xi \right)^{-1/3} \left[ \frac{f(x)\Delta^2(x)}{\text{Var}(Y|X = x)} \right]^{1/3} \gamma_n^{-1/3} \right),$$



then

$$MSE[r_n(x)] = \frac{3(f(x))^{4/3}(\Delta(x)\text{Var}(Y|X=x))^{2/3}}{2^{4/3}(f(x) - \frac{1}{3}\xi)^{4/3}}\gamma_n^{2/3} + o(\gamma_n^{2/3}).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator  $r_n(x)$  defined in (4.4.1), for  $x = 0, 1$  such that  $f(x) > 0$ .

**Theorem 4.4.2.** (*Weak pointwise convergence rate*).

Let Assumption (A1) – (A3) and (A'4) hold, we have

1. If  $\gamma_n^{-1}m_n^{-3} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2}m_n^{-1/2}(r_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(\sqrt{c} \frac{f(x)\Delta(x)}{f(x) - a\xi}, \frac{f(x)\text{Var}(Y|X=x)\psi(x)}{2f(x) - (\alpha - a)\xi}\right),$$

2. If  $\gamma_n^{-1}m_n^{-3} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n(r_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{f(x)\Delta(x)}{f(x) - a\xi},$$

where  $\xrightarrow[n \rightarrow +\infty]{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the Gaussian-distribution and  $\xrightarrow[n \rightarrow +\infty]{\mathbb{P}}$  the convergence in probability.

#### 4.4.3 The MISE of $r_n$

To obtain the MISE of  $r_n$ , we add the following assumption

(A''4) Set  $\varphi = \inf_{x \in [0,1]} f(x) > 0$ , we demand that  $\lim_{n \rightarrow \infty} (n\gamma_n) \in \left(\min\left(\frac{a}{\varphi}, \frac{2\alpha - a}{4\varphi}\right), \infty\right]$ .

**Proposition 4.4.3.**

Let Assumptions (A1) – (A3) and (A''4) hold, we have

1. If  $\frac{1-\alpha}{4} < a < \frac{2}{5}\alpha$ , then

$$MISE(r_n) = K_1 m_n^{-2} + o(m_n^{-2}).$$

2. If  $a = \frac{2}{5}\alpha$ , then

$$MISE(r_n) = K_1 m_n^{-2} + K_2 \gamma_n m_n^{1/2} + o(m_n^{-2} + \gamma_n m_n^{1/2}).$$

3. If  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , then

$$MISE(r_n) = K_2\gamma_n m_n^{1/2} + o(\gamma_n m_n^{1/2}).$$

The following result is a consequence of the previous proposition which gives the optimal order  $(m_n)$  of the estimator  $r_n$  (4.4.1) and the corresponding  $MISE$ .

**Corollary 4.4.1.**

Let Assumptions (A1) – (A3) and (A'4) hold. To minimize the  $MISE$  of  $r_n$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$  and  $(m_n)$  must be in  $\mathcal{GS}(2/5)$  such that

$$\left(\frac{4K_1}{K_2}\right)^{2/5} \gamma_n^{-2/5},$$

and then

$$MISE(r_n) = \frac{5K_1^{1/5}K_2^{4/5}}{4^{4/5}}\gamma_n^{4/5} + o(\gamma_n^{4/5}).$$

## 4.5 Applications

We recall the regression function's kernel estimator proposed by Nadaraya [72] and Watson [120], for  $x \in \mathbb{R}$  such that  $f(x) \neq 0$

$$\hat{r}_n^{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}, \quad (4.5.1)$$

where  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a nonnegative, continuous, bounded function satisfying  $\int_{\mathbb{R}} K(z)dz = 1$ ,  $\int_{\mathbb{R}} zK(z)dz = 0$  and  $\int_{\mathbb{R}} z^2K(z)dz < \infty$  known as kernel and  $h = (h_n)$  is a bandwidth (that is, a sequence of positive real numbers that goes to zero).

We recall also the recursive estimator of a regression function which is a generalized version of Révész's estimator (see Révész [85, 86]) and was studied by Mokkadem et al. [67]

$$r_n^{GR}(x) = \left(1 - \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right)\right) r_{n-1}^{GR}(x) + \gamma_n h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right). \quad (4.5.2)$$

The major limitation of these two estimators occurs at the edges of the support. In fact, these estimators are inconsistent at the boundary. This effectively restricts their application to values of  $x$  in the interior of the support of the estimated regression function.

The object of this section is to provide a study comparing between Nadaraya-Watson's estimator  $\widehat{r}_n^{NW}$  defined in (4.5.1), the generalized Révész's estimator  $r_n^{GR}$  defined in (4.5.2), our non-recursive estimator  $\widehat{r}_n$  defined in (4.3.1) and our recursive estimator  $r_n$  introduced in (4.4.1).

### 4.5.1 Simulations

We consider the regression model

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

where  $\varepsilon \sim \mathcal{N}(0, 1)$ .

When using the estimators  $\widehat{r}_n^{NW}$  and  $r_n^{GR}$ , we choose the kernel  $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  and the bandwidth equal to  $(h_n) = n^{-1/5}(\ln(n+1))^{-1}$ . When using our Bernstein estimators  $\widehat{r}_n$  and  $r_n$ , we choose the order equal to  $m_n = \lfloor n^{2/5}(\ln(n+1)) \rfloor$  and we choose two stepsize  $(\gamma_n) = (n^{-0.9})$  and  $(\gamma_n) = (n^{-1})$ .

We consider three sample sizes  $n = 50$ ,  $n = 100$  and  $n = 500$ , three regression functions

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

(b)  $r(x) = 0.3 \exp(-x^2/2) + 0.7 \exp(-(x-1)^2/2)$ ,

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

and three densities of  $X$ , the beta density  $\mathcal{B}(3, 5)$ , the beta mixture density  $0.5\mathcal{B}(2, 1) + 0.5\mathcal{B}(1, 4)$  and the truncated standard normal density  $\mathcal{N}_{[0,1]}(0, 1)$ .

We consider six estimators : our non-recursive Bernstein estimator  $\hat{r}_n$  (4.3.1), Nadaraya-Watson's estimator  $\hat{r}_n^{NW}$  (4.5.1), two proposed recursive Bernstein estimators  $r_{n,1}$  and  $r_{n,2}$  (4.4.1) with stepsize  $(\gamma_n) = (n^{-1})$  and  $(\gamma_n) = (n^{-0.9})$  respectively and finally two Generalized Révész's estimators  $r_{n,1}^{GR}$  and  $r_{n,2}^{GR}$  (4.5.2) using the same stepsizes as in  $r_n$ . For each model and sample of size  $n$ , we approximate the average integrated squared error (*ISE*) of the estimator using  $N = 500$  trials of sample size  $n$

$$\overline{ISE} = \frac{1}{N} \sum_{k=1}^N ISE [\bar{r}_k],$$

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

$$ISE [\bar{r}_k] = \int_0^1 \{\bar{r}(x) - r(x)\}^2 dx.$$

From Table 4.1 we conclude that :

- In all the considered models, the average *ISE* of our non-recursive regression estimator  $\hat{r}_n$  defined in (4.3.1) is the smallest, except the cases with  $X \sim \mathcal{B}(3, 5)$  where the the average *ISE* of Nadaraya-Watson's estimator  $\hat{r}_n^{NW}$  given in (4.5.1) is the smallest when the size is  $n = 50$  and the cases with  $X \sim \mathcal{N}_{[0,1]}(0, 1)$  where the average *ISE* of the generalized Révész's estimator is the smallest.
- In all the models with  $X \sim \mathcal{N}_{[0,1]}(0, 1)$ , the average *ISE* of our recursive regression estimator  $r_n$  defined in (4.4.1) with the choice  $(\gamma_n) = (n^{-1})$  is smaller than that of our non-recursive regression estimator  $\hat{r}_n$  introduced in (4.3.1).
- In all the models, the average *ISE* of our recursive regression estimator  $r_{n,1}$  with the choice  $(\gamma_n) = (n^{-1})$  is smaller than that of our recursive regression estimator  $r_{n,2}$  with the choice  $(\gamma_n) = (n^{-0.9})$ .
- The average *ISE* decreases as the sample size increases.

For qualitative comparison between our non-recursive estimator  $\hat{r}_n$  defined in (4.3.1) and our recursive estimator  $r_n$  given in (4.4.1) with  $(\gamma_n) = (n^{-1})$ . Figure 4.1, shows regression estimates plotted for 500 simulated samples from the model  $r(x) = \cos(x)$  with  $X \sim \mathcal{N}_{[0,1]}(0, 1)$  of sizes  $n = 100$  (left panel) and  $n = 500$  (right panel).

From figure 4.1, we conclude that :

- Both our estimators are close to the true regression function.
- Our recursive regression estimator  $r_n$  defined in (4.4.1) using the stepsize  $(\gamma_n) = (n^{-1})$  is closer to the true regression function than that of the proposed non-recursive estimator  $\hat{r}_n$  given in (4.3.1) especially with the size  $n = 500$ .
- When the sample size increases, we get closer estimation of the true regression function.

## 4.5.2 Real dataset

In any practical situation, to estimate an unknown regression function, it is critical to have a reliable data-dependent rule for order selection. One popular and practical approach is cross-validation. First, we compute the leave-one-out residuals :

$$\forall i \in \{1, \dots, n\}, \quad e_{-i} = Y_i - \bar{r}_{-i}(X_i),$$

where  $\bar{r}_{-i}$  is the regression estimate without the data point  $(X_i, Y_i)$ . Then, the smoothing parameter is chosen by minimizing

$$CV(m) = \frac{1}{n} \sum_{i=1}^n e_{-i}^2.$$

We consider the CO2 dataset which is available in the R package `Stat2Data` and contained 237 observations on two variables ; Day and CO2. Scientists at a research station in Brotjacklriegel, Germany recorded CO2 levels, in parts per million, in the atmosphere for each day from the start of April through November in 2001. We apply our proposed estimators  $\hat{r}_n$  defined in (4.3.1) and  $r_n$  given in (4.4.1) on this model. For convenience, we assume that the minimum of days is 90 and the maximum is 335 (the Day data are such that  $\min_i(x_i) = 91$  and  $\max_i(x_i) = 334$ ). Finally, we used the Cross-validation method to obtain  $m = 220$  for our non-recursive estimator  $\hat{r}_n$  and  $m_n = n$  for our recursive estimator. We observe that both our estimators are very close to Nadaraya-Watson's estimator (4.5.1) especially near the boundaries.

## 4.6 Conclusion

In this chapter, we propose a non-recursive and recursive estimator of regression function based on Bernstein polynomials and stochastic algorithm derived from the Robbins-Monro's scheme. We first study their theoretical behavior. Then, we conduct a simulation study and analyse a real data application on CO2 data.

For all the models, the average *ISE* of our non-recursive regression estimator  $\hat{r}_n$  defined in (4.3.1) is the smallest, except the cases with  $X \sim \mathcal{B}(3, 5)$  where the average *ISE* of Nadaraya-Watson's estimator  $\hat{r}_n^{NW}$  given in (4.5.1) is smaller when the sample size is  $n = 50$  and the cases with  $X \sim \mathcal{N}_{[0,1]}(0, 1)$  where the average *ISE* of the generalized Révész's estimator is smaller. In addition, a major advantage of our recursive estimator is that its update, when new sample points are available, requires less computational cost than Nadaraya-Watson estimator. Finally, the two estimators have nice features and satisfactory improvement in comparison to Kernel estimators especially near the boundaries.

In conclusion, the estimation using Bernstein polynomials allowed us to overcome the edge problem and obtain quite similar results as Nadaraya-Watson's estimator. Moreover, we plan to make extensions of our method in the future and follow Slaoui [109] to create a semi-recursive Bernstein estimator for regression function.

## 4.7 Outlines of the proofs

In this section, we present proofs for the results given in the chapter. First, we recall a series of results, which are proven in Leblanc [55], linked to different sums of Bernstein polynomial, defined by

$$S_m(x) = \sum_{k=0}^m b_k^2(m, x),$$

These results are given in the following lemma.

**Lemma 4.7.1.**

We have

(i)  $0 \leq S_m(x) \leq 1, \forall x \in [0, 1],$

(ii)  $S_m(x) = m^{-1/2} [\psi(x) + o_x(1)], \forall x \in (0, 1)$

(iii)  $S_m(0) = S_m(1) = 1.$

Let  $g$  be any continuous function on  $[0, 1]$ . Then

(iv)  $m^{1/2} \int_0^1 g(x)S_m(x)dx = \int_0^1 g(x)\psi(x)dx + o(1),$

We start by proving the characteristics of our non-recursive estimator  $\hat{r}_n$  defined by (4.3.1). To do so, we note

$$N_n(x) = \frac{m}{n} \sum_{i=1}^n Y_i \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x).$$

Then, we may rewrite  $\hat{r}_n$  as

$$\hat{r}_n(x) = \frac{N_n(x)}{f_n(x)},$$

where  $f_n$  is the Vitale's estimator of the density  $f$  defined, for all  $x \in [0, 1]$ , by

$$\begin{aligned} f_n(x) &= \frac{m}{n} \sum_{i=1}^n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x) \\ &= m \sum_{k=0}^{m-1} \left\{ F_n \left( \frac{k+1}{m} \right) - F_n \left( \frac{k}{m} \right) \right\} b_k(m-1, x), \end{aligned}$$

with  $F_n$  is the empirical distribution function of the variable  $X$ .

### 4.7.1 Proof of Proposition 4.3.1

We start by giving the bias and the variance of  $N_n(x)$

$$\begin{aligned}
 \mathbb{E}[N_n(x)] &= m\mathbb{E}\left[Y \sum_{k=0}^{m-1} \mathbb{I}_{\left\{\frac{k}{m} < X \leq \frac{k+1}{m}\right\}} b_k(m-1, x)\right], \\
 &= m \sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+1}{m}} \left(\int_{\mathbb{R}} yg(z, y)dy\right) dz b_k(m-1, x), \\
 &= m \sum_{k=0}^{m-1} \left(\int_{\frac{k}{m}}^{\frac{k+1}{m}} r(z)f(z)dz\right) b_k(m-1, x).
 \end{aligned}$$

Using Taylor expansion, we have

$$\begin{aligned}
 r(z)f(z) &= \left[r(x) + (z-x)r'(x) + \frac{(z-x)^2}{2}r''(x) + o((z-x)^2)\right] \\
 &\times \left[f(x) + (z-x)f'(x) + \frac{(z-x)^2}{2}f''(x) + o((z-x)^2)\right], \\
 &= r(x)f(x) + (z-x)(r'(x)f(x) + r(x)f'(x)) \\
 &+ \frac{(z-x)^2}{2}(r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)) + o((z-x)^2).
 \end{aligned}$$

Set  $N(x) = r(x)f(x)$ , then we obtain

$$\begin{aligned}
 \mathbb{E}[N_n(x)] &= r(x)f(x)m \sum_{k=0}^{m-1} \left(\frac{k+1}{m} - \frac{k}{m}\right) b_k(m-1, x) \\
 &+ (r'(x)f(x) + r(x)f'(x)) \frac{m}{2} \sum_{k=0}^{m-1} \left\{ \left(\frac{k+1}{m} - x\right)^2 - \left(\frac{k}{m} - x\right)^2 \right\} \\
 &\times b_k(m-1, x) \\
 &+ (r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)) \\
 &\frac{m}{6} \sum_{k=0}^{m-1} \left\{ \left(\frac{k+1}{m} - x\right)^3 - \left(\frac{k}{m} - x\right)^3 \right\} b_k(m-1, x) \\
 &+ o\left(m \sum_{k=0}^{m-1} \left\{ \left(\frac{k+1}{m} - x\right)^3 - \left(\frac{k}{m} - x\right)^3 \right\} b_k(m-1, x)\right),
 \end{aligned}$$



$$\begin{aligned}
&= N(x) + (r'(x)f(x) + r(x)f'(x)) \frac{m}{2} \sum_{k=0}^{m-1} m^{-2}(2k+1-2mx)b_k(m-1, x) \\
&\quad + (r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)) \\
&\quad \frac{m}{6} \sum_{k=0}^{m-1} m^{-3} \left\{ (k+1-mx)^2 + (k-mx)^2 \right. \\
&\quad \left. + (k+1-mx)(k-mx) \right\} b_k(m-1, x) [1 + o(1)] \\
&= N(x) + (r'(x)f(x) + r(x)f'(x)) \frac{m^{-1}}{2} \{2T_{1,m-1}(x) + (1-2x)T_{0,m-1}(x)\} \\
&\quad + (r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)) \frac{m^{-2}}{6} \sum_{k=0}^{m-1} \left\{ 3(k-mx)^2 \right. \\
&\quad \left. + 3(k-mx) + 1 \right\} b_k(m-1, x) [1 + o(1)] \\
&= N(x) + (r'(x)f(x) + r(x)f'(x)) \frac{m^{-1}}{2} \{2T_{1,m-1}(x) + (1-2x)T_{0,m-1}(x)\} \\
&\quad + (r''(x)f(x) + f''(x)r(x) + 2r'(x)f'(x)) \frac{m^{-2}}{6} \left\{ 3T_{2,m-1}(x) \right. \\
&\quad \left. + 3(2x+1)T_{1,m-1}(x) + (x^2+3x+1)T_{0,m-1}(x) \right\} [1 + o(1)],
\end{aligned}$$

where

$$T_{j,m}(x) = \sum_{k=0}^{m-1} (k-mx)^j b_k(m, x), \quad \forall j \in \mathbb{N}.$$

Note that it is easy to obtain

$$T_{0,m}(x) = 1, \quad T_{1,m}(x) = 0 \quad T_{2,m}(x) = mx(1-x),$$

then, we have

$$\mathbb{E}[N_n(x)] = N(x) + \Delta_2(x)m^{-1} + o(m^{-1}) \quad (4.7.1)$$

Moreover, we have

$$\text{Var}[N_n(x)] = \mathbb{E}[N_n^2(x)] - \mathbb{E}^2[N_n(x)],$$

where

$$\begin{aligned}
N_n^2(x) &= \frac{m^2}{n^2} \sum_{i=1}^n Y_i^2 \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x) \right)^2 \\
&\quad + \frac{m^2}{n^2} \sum_{i,j=1, i \neq j}^n Y_i Y_j \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x) \right) \\
&\quad \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_j \leq \frac{k+1}{m}\}} b_k(m-1, x) \right),
\end{aligned}$$

so, we have

$$\begin{aligned}
\mathbb{E}[N_n^2(x)] &= \frac{m^2}{n} \mathbb{E} \left[ Y^2 \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right)^2 \right] \\
&\quad + \frac{m^2 n(n-1)}{n^2} \mathbb{E}^2 \left[ Y \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right], \\
&= \frac{m^2}{n} \mathbb{E} \left[ Y^2 \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right)^2 \right] + \left(1 - \frac{1}{n}\right) \mathbb{E}^2[N_n(x)],
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}[N_n(x)] &= \frac{m^2}{n} \mathbb{E} \left[ Y^2 \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right)^2 \right] - \frac{1}{n} \mathbb{E}^2[N_n(x)], \\
&= \frac{m^2}{n} \mathbb{E} \left[ Y^2 \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k^2(m-1, x) \right] - \frac{1}{n} \mathbb{E}^2[N_n(x)], \\
&= \frac{m^2}{n} \sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+1}{m}} \left( \int_{\mathbb{R}} y^2 g(z, y) dy \right) dz b_k^2(m-1, x) - \frac{1}{n} \mathbb{E}^2[N_n(x)], \\
&= \frac{m^2}{n} \sum_{k=0}^{m-1} \left( \int_{\frac{k}{m}}^{\frac{k+1}{m}} \mathbb{E}[Y^2 | X = z] f(z) dz \right) b_k^2(m-1, x) - \frac{1}{n} \mathbb{E}^2[N_n(x)], \\
&= \frac{m}{n} \mathbb{E}[Y^2 | X = x] f(x) S_m(x) - \frac{1}{n} \mathbb{E}^2[N_n(x)].
\end{aligned}$$

Using Lemma 4.7.1 (ii) and (iii), we obtain

$$\text{Var}[N_n(x)] = \begin{cases} \frac{m^{1/2}}{n} \mathbb{E}[Y^2 | X = x] f(x) \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right) & \text{for } x \in (0, 1), \\ \frac{m}{n} \mathbb{E}[Y^2 | X = x] f(x) + o_x \left( \frac{m}{n} \right) & \text{for } x = 0, 1. \end{cases} \quad (4.7.2)$$

Furthermore, we have

$$\begin{aligned}
Cov(f_n(x), N_n(x)) &= \mathbb{E}[f_n(x)N_n(x)] - \mathbb{E}[f_n(x)]\mathbb{E}[N_n(x)], \\
&= \frac{m^2}{n} \mathbb{E} \left[ Y \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right)^2 \right] \\
&\quad + \frac{n(n-1)m^2}{n^2} \mathbb{E}^2 \left[ Y \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right] \\
&\quad - \mathbb{E}[f_n(x)]\mathbb{E}[N_n(x)], \\
&= \frac{m^2}{n} \mathbb{E} \left[ Y \left( \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right)^2 \right] - \frac{1}{n} \mathbb{E}[f_n(x)]\mathbb{E}[N_n(x)], \\
&= \frac{m^2}{n} \sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+1}{m}} \left( \int_{\mathbb{R}} yg(z, y) dy \right) dz b_k^2(m-1, x) - \frac{1}{n} \mathbb{E}[f_n(x)]\mathbb{E}[N_n(x)], \\
&= \frac{m}{n} r(x) f(x) S_m(x) - \frac{1}{n} \mathbb{E}[f_n(x)]\mathbb{E}[N_n(x)].
\end{aligned}$$

Using lemma 4.7.1 (ii) and (iii), we obtain

$$Cov(f_n(x), N_n(x)) = \begin{cases} \frac{m^{1/2}}{n} r(x) f(x) \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right) & \text{for } x \in (0, 1), \\ \frac{m}{n} r(x) f(x) + o_x \left( \frac{m}{n} \right) & \text{for } x = 0, 1. \end{cases} \quad (4.7.3)$$

To obtain the bias of  $\widehat{r}_n(x)$ , we let  $h(x, y) = \frac{y}{x}$ . Using Taylor expansion, we have

$$\begin{aligned}
h(x_n, y_n) &= h(x, y) + (x_n - x, y_n - y) \nabla h^T(x, y) \\
&\quad + \frac{1}{2} (x_n - x, y_n - y) \mathbb{H}(x, y) (x_n - x, y_n - y)^T + o(\|(x_n - x, y_n - y)\|^2),
\end{aligned}$$

where  $\nabla h$  is the gradient of  $h$  and  $\mathbb{H}$  is its hessian matrix.

$$\begin{aligned}
\nabla h(x, y) &= \left( -\frac{y}{x^2}, \frac{1}{x} \right) \\
\mathbb{H} &= \begin{pmatrix} \frac{2y}{x^3} & -\frac{1}{x^2} \\ -\frac{1}{x^2} & 0 \end{pmatrix}
\end{aligned}$$

Then, we have

$$\begin{aligned}
\frac{y_n}{x_n} &= \frac{y}{x} - \frac{y}{x^2} (x_n - x) + \frac{1}{x} (y_n - y) + \frac{y}{x^3} (x_n - x)^2 - \frac{1}{x^2} (x_n - x)(y_n - y) \\
&\quad + o((x_n - x)^2 + (x_n - x)(y_n - y)).
\end{aligned}$$

We set  $(x_n, y_n) = (f_n(x), N_n(x))$  and  $(x, y) = (f(x), N(x))$ , so we obtain

$$\begin{aligned}\widehat{r}_n(x) &= r(x) - \frac{r(x)}{f(x)} (f_n(x) - f(x)) + \frac{1}{f(x)} (N_n(x) - N(x)) \\ &+ \frac{r(x)}{\{f(x)\}^2} (f_n(x) - f(x))^2 - \frac{1}{\{f(x)\}^2} (f_n(x) - f(x)) (N_n(x) - N(x)) \\ &+ o\left((f_n(x) - f(x))^2 + (f_n(x) - f(x)) (N_n(x) - N(x))\right),\end{aligned}$$

then

$$\begin{aligned}\mathbb{E}[\widehat{r}_n(x)] &= r(x) - \frac{r(x)}{f(x)} (\mathbb{E}[f_n(x)] - f(x)) + \frac{1}{f(x)} (\mathbb{E}[N_n(x)] - N(x)) \\ &+ \frac{r(x)}{\{f(x)\}^2} (\mathbb{E}[f_n(x)] - f(x))^2 - \frac{1}{\{f(x)\}^2} \mathbb{E}[(f_n(x) - f(x)) (N_n(x) - N(x))] \\ &+ o\left(\mathbb{E}[(f_n(x) - f(x))^2] + \mathbb{E}[(f_n(x) - f(x)) (N_n(x) - N(x))]\right).\end{aligned}$$

Let us recall, that for the Vitale's estimator  $f_n$ , we have

$$\mathbb{E}[f_n(x)] = f(x) + \frac{\Delta_1(x)}{m} + o(m^{-1}), \quad \forall x \in [0, 1], \quad (4.7.4)$$

and

$$\text{Var}[f_n(x)] = \begin{cases} \frac{m^{1/2}}{n} f(x) \psi(x) + o_x\left(\frac{m^{1/2}}{n}\right) & \text{for } x \in (0, 1), \\ \frac{m}{n} f(x) + o_x\left(\frac{m}{n}\right) & \text{for } x = 0, 1, \end{cases} \quad (4.7.5)$$

Using (4.7.4) and (4.7.1), we have

$$\begin{aligned}\mathbb{E}[\widehat{r}_n(x)] &= r(x) + \left( \frac{1}{f(x)} \Delta_2(x) - \frac{r(x)}{f(x)} \Delta_1(x) \right) m_n^{-1} + o(m^{-1}), \\ &= r(x) + \Delta(x) m^{-1} + o(m^{-1}), \quad \forall x \in [0, 1]\end{aligned}$$

and we obtain (4.3.2) of Proposition 4.3.1.

Now for the variance of  $\widehat{r}_n(x)$ , we have

$$\text{Var}[h(x_n, y_n)] = \nabla h(x, y) \text{Var}(x_n, y_n) \nabla h^T(x, y) [1 + o(1)],$$

which gives

$$\text{Var}[\widehat{r}_n(x)] = \nabla h(x, y) \Sigma[f_n(x), m_n(x)] \nabla h^T(x, y) [1 + o(1)].$$

Using (4.7.3), (4.7.2) and (4.7.5)

$$\Sigma [f_n(x), N_n(x)] = \begin{cases} \frac{m^{1/2}}{n} \begin{pmatrix} f(x) & f(x)r(x) \\ f(x)r(x) & f(x)\mathbb{E}(Y^2|X=x) \end{pmatrix} \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right) & \text{for } x \in (0, 1), \\ \frac{m}{n} \begin{pmatrix} f(x) & f(x)r(x) \\ f(x)r(x) & f(x)\mathbb{E}(Y^2|X=x) \end{pmatrix} + o \left( \frac{m}{n} \right) & \text{for } x = 0, 1. \end{cases}$$

So, for  $x \in (0, 1)$ , we have

$$\begin{aligned} \text{Var} [\widehat{r}_n(x)] &= \begin{pmatrix} -\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)} \end{pmatrix} \times \begin{pmatrix} f(x) & f(x)r(x) \\ f(x)r(x) & f(x)\mathbb{E}(Y^2|X=x) \end{pmatrix} \\ &\times \begin{pmatrix} -\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)} \end{pmatrix}^T \times \frac{m^{1/2}}{n} \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right), \\ &= \begin{pmatrix} -\frac{N(x)}{f(x)} + r(x), -\frac{N(x)r(x)}{f(x)} + \mathbb{E}(Y^2|X=x) \end{pmatrix} \times \begin{pmatrix} -\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)} \end{pmatrix}^T \\ &\times \frac{m^{1/2}}{n} \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right), \\ &= \frac{1}{f(x)} (0, -\mathbb{E}^2(Y|X=x) + \mathbb{E}(Y^2|X=x)) \times (-1, 1)^T \frac{m^{1/2}}{n} \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right), \\ &= \frac{1}{f(x)} [\mathbb{E}(Y^2|X=x) - \mathbb{E}^2(Y|X=x)] \times \frac{m^{1/2}}{n} \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right), \\ &= \frac{m^{1/2}}{n} \frac{\text{Var}(Y|X=x)}{f(x)} \psi(x) + o_x \left( \frac{m^{1/2}}{n} \right), \end{aligned}$$

and, for  $x = 0, 1$ , we have

$$\text{Var} [\widehat{r}_n(x)] = \frac{m}{n} \frac{\text{Var}(Y|X=x)}{f(x)} + o_x \left( \frac{m}{n} \right).$$

which gives (4.3.3) of Proposition 4.3.1.

## 4.7.2 Proof of Proposition 4.3.2

We have

$$\begin{aligned} \text{MISE}(\widehat{r}_n) &= \int_0^1 (\text{Var} [\widehat{r}_n(x)] + \text{Bias}^2 [\widehat{r}_n(x)]) dx \\ &= \int_0^1 \text{Var} [\widehat{r}_n(x)] dx + C_1 m^{-2} + o(m^{-2}). \end{aligned}$$

Moreover, we have

$$\begin{aligned}
\text{Var} [\widehat{r}_n(x)] &= \left( -\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)} \right) \times \begin{pmatrix} \text{Var} [f_n(x)] & \text{Cov} [f_n(x), N_n(x)] \\ \text{Cov} [f_n(x), N_n(x)] & \text{Var} [N_n(x)] \end{pmatrix} [1 + o(1)] \\
&\times \left( -\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)} \right)^T, \\
&= \frac{r^2(x)}{\{f(x)\}^2} \text{Var} [f_n(x)] - 2 \frac{r(x)}{\{f(x)\}^2} \text{Cov} [f_n(x), N_n(x)] \\
&\quad + \frac{1}{\{f(x)\}^2} \text{Var} [N_n(x)] [1 + o(1)],
\end{aligned}$$

then

$$\begin{aligned}
\int_0^1 \text{Var} [\widehat{r}_n(x)] dx &= \int_0^1 r^2(x) \frac{\text{Var} [f_n(x)]}{\{f(x)\}^2} dx - 2 \int_0^1 r(x) \frac{\text{Cov} [f_n(x), N_n(x)]}{\{f(x)\}^2} dx \\
&\quad + \int_0^1 \frac{\text{Var} [N_n(x)]}{\{f(x)\}^2} dx [1 + o(1)]. \tag{4.7.6}
\end{aligned}$$

First, we have

$$\text{Var} [f_n(x)] = \frac{1}{n} [A_m(x) - f_m^2(x)],$$

where  $f_m^2(x) = \mathbb{E}^2 [f_n(x)] = f^2(x) + O(m^{-1})$ , and

$$\begin{aligned}
A_m(x) &= m^2 \sum_{k=0}^{m-1} \left[ F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right) \right] b_k^2(m-1, x), \\
&= m [f(x) S_{m-1}(x) + O(H_{m-1}(x)) + O(m^{-1})],
\end{aligned}$$

for  $x \in [0, 1]$ , and with

$$H_m(x) = \sum_{k=0}^{m-1} \left| \frac{k}{m} - x \right| b_k^2(m-1, x).$$

Using Cauchy-Schwarz inequality, we write

$$H_m(x) \leq \left[ \sum_{k=0}^m \left( \frac{k}{m} - x \right)^2 b_k(m, x) \right]^{1/2} \left[ \sum_{k=0}^m b_k^3(m, x) \right]^{1/2} \leq \left[ \frac{S_m(x)}{4m} \right]^{1/2},$$

for all  $m \geq 1$  and  $x \in [0, 1]$ , since  $0 \leq b_k(m, x) \leq 1$  and

$$\sum_{k=0}^m \left( \frac{k}{m} - x \right)^2 b_k(m, x) = \frac{x(1-x)}{m} \leq \frac{1}{4m}.$$

Then, applying Jensen's inequality and lemma 4.7.1 (iv), for any continuous function  $g$ , we have

$$\begin{aligned} \int_0^1 g(x)H_m(x)dx &\leq \int_0^1 g(x) \left[ \frac{S_m(x)}{4m} \right]^{1/2} dx, \\ &\leq \left[ \int_0^1 g(x)dx \right]^{1/2} \left[ \frac{1}{4m^{3/2}} \int_0^1 g(x)\psi(x)dx + o(m^{-3/2}) \right]^{1/2} \\ &= O(m^{-3/4}), \end{aligned}$$

So, we obtain

$$\begin{aligned} \int_0^1 r^2(x) \frac{Var[f_n(x)]}{\{f(x)\}^2} dx &= \frac{1}{n} \int_0^1 r^2(x) \frac{A_m(x) - f_m^2(x)}{\{f(x)\}^2} dx, \\ &= \frac{1}{n} \left[ \int_0^1 r^2(x) \frac{A_m(x)}{\{f(x)\}^2} dx - \int_0^1 r^2(x) \right] + O\left(\frac{1}{mn}\right), \\ &= \frac{m}{n} \left[ \int_0^1 \frac{r^2(x)}{\{f(x)\}^2} (S_{m-1}(x) + O(H_{m-1}(x)) + O(m^{-1})) dx \right] \\ &\quad - \frac{1}{n} \int_0^1 r^2(x) + O\left(\frac{1}{mn}\right), \\ &= \frac{m}{n} \left[ \int_0^1 \frac{r^2(x)}{f(x)} S_{m-1}(x) dx + O(m^{-3/4}) \right] \\ &\quad - \frac{1}{n} \int_0^1 r^2(x) + O\left(\frac{1}{mn}\right), \end{aligned}$$

and, using lemma 4.7.1 (iv), we have

$$\begin{aligned} \int_0^1 r^2(x) \frac{Var[f_n(x)]}{\{f(x)\}^2} dx &= \frac{m^{1/2}}{n} \int_0^1 \frac{r^2(x)}{f(x)} \psi(x) dx - \frac{1}{n} \int_0^1 r^2(x) \\ &\quad + o\left(\frac{m^{1/2}}{n}\right) + O\left(\frac{1}{mn}\right), \end{aligned} \quad (4.7.7)$$

Second, we have

$$\begin{aligned} Cov[f_n(x), N_n(x)] &= \frac{1}{n} \left\{ m^2 \sum_{k=0}^{m-1} \left( \int_{\frac{k}{m}}^{\frac{k+1}{m}} r(z)f(x)dz \right) b_k^2(m-1, x) - \mathbb{E}[f_n(x)] \mathbb{E}[N_n(x)] \right\}, \\ &= \frac{m^2}{n} \sum_{k=0}^{m-1} \left( \int_{\frac{k}{m}}^{\frac{k+1}{m}} [r(x)f(x) + O(z-x)] dz \right) b_k^2(m-1, x) \\ &\quad - \frac{1}{n} f(x)N(x) + O\left(\frac{1}{mn}\right), \\ &= \frac{m}{n} [r(x)f(x)S_{m-1}(x) + O(H_{m-1}(x)) + O(m^{-1})] - \frac{1}{n} f(x)N(x) \\ &\quad + O\left(\frac{1}{mn}\right), \end{aligned}$$

then, using the same argument for  $H_{m-1}(x)$  as previously, we obtain

$$\begin{aligned}
\int_0^1 r(x) \frac{Cov[f_n(x), N_n(x)]}{\{f(x)\}^2} dx &= \frac{m}{n} \left[ \int_0^1 \frac{r^2(x)}{f(x)} S_{m-1}(x) dx + O(m^{-3/4}) \right] - \frac{1}{n} \int_0^1 r^2(x) \\
&\quad + O\left(\frac{1}{mn}\right), \\
&= \frac{m^{1/2}}{n} \int_0^1 \frac{r^2(x)}{f(x)} \psi(x) dx - \frac{1}{n} \int_0^1 r^2(x) \\
&\quad + o\left(\frac{m^{1/2}}{n}\right) + O\left(\frac{1}{mn}\right). \tag{4.7.8}
\end{aligned}$$

Third, we have

$$\begin{aligned}
Var[N_n(x)] &= \frac{m^2}{n} \sum_{k=0}^{m-1} \left( \int_{\frac{k}{m}}^{\frac{k+1}{m}} \mathbb{E}[Y^2|X=z] f(z) dz \right) b_k^2(m-1, x) - \frac{1}{n} \mathbb{E}^2[N_n(x)], \\
&= \frac{m^2}{n} \sum_{k=0}^{m-1} \left( \int_{\frac{k}{m}}^{\frac{k+1}{m}} [\mathbb{E}[Y^2|X=x] f(x) + O(z-x)] dz \right) b_k^2(m-1, x) - \frac{1}{n} N^2(x) \\
&\quad + O\left(\frac{1}{mn}\right), \\
&= \frac{m}{n} [\mathbb{E}[Y^2|X=x] f(x) S_{m-1}(x) + O(H_{m-1}(x)) + O(m^{-1})] - \frac{1}{n} N^2(x) \\
&\quad + O\left(\frac{1}{mn}\right),
\end{aligned}$$

then,

$$\begin{aligned}
\int_0^1 \frac{Var[N_n(x)]}{\{f(x)\}^2} dx &= \frac{m}{n} \left[ \int_0^1 \frac{\mathbb{E}[Y^2|X=x]}{f(x)} S_{m-1}(x) dx + O(m^{-3/4}) \right] - \frac{1}{n} \int_0^1 r^2(x) \\
&\quad + O\left(\frac{1}{mn}\right), \\
&= \frac{m^{1/2}}{n} \int_0^1 \frac{\mathbb{E}[Y^2|X=x]}{f(x)} \psi(x) dx - \frac{1}{n} \int_0^1 r^2(x) \\
&\quad + o\left(\frac{m^{1/2}}{n}\right) + O\left(\frac{1}{mn}\right). \tag{4.7.9}
\end{aligned}$$



Finally, substituting (4.7.7), (4.7.8) and (4.7.9) into (4.7.6), we obtain

$$\begin{aligned}
\int_0^1 \text{Var} [\widehat{r}_n(x)] dx &= \left( \int_0^1 \frac{\mathbb{E}[Y^2|X=x]}{f(x)} \psi(x) dx - \int_0^1 \frac{\mathbb{E}^2[Y|X=x]}{f(x)} \psi(x) dx \right) \frac{m^{1/2}}{n} \\
&\quad + o\left(\frac{m^{1/2}}{n}\right), \\
&= \int_0^1 \frac{\mathbb{E}[Y^2|X=x] - \mathbb{E}^2[Y|X=x]}{f(x)} \psi(x) dx \frac{m^{1/2}}{n} + o\left(\frac{m^{1/2}}{n}\right), \\
&= \int_0^1 \frac{\text{Var}[Y|X=x]}{f(x)} \psi(x) dx \frac{m^{1/2}}{n} \\
&\quad + o\left(\frac{m^{1/2}}{n}\right).
\end{aligned}$$

Then, we obtain the result in (4.3.4) of Proposition 4.3.2.

### 4.7.3 Proof of Theorem 4.3.1

To prove the convergence, for  $x \in (0, 1)$ , we will use the fact that

$$n^{1/2}m^{-1/4} (\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{\text{Var}(Y|X=x)}{f(x)} \psi(x)\right), \quad (4.7.10)$$

which will be proved later. We have

$$\begin{aligned}
n^{1/2}m^{-1/4} (\widehat{r}_n(x) - r(x)) &= n^{1/2}m^{-1/4} (\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)] + n^{1/2}m^{-1/4} (\mathbb{E}[\widehat{r}_n(x)] - r(x))), \\
&= n^{1/2}m^{-1/4} (\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) + n^{1/2}m^{-5/4} \Delta(x)[1 + o(1)],
\end{aligned}$$

so if  $nm^{-5/2} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then Part 1 of Theorem 4.3.1 follows.

Now, if  $nm^{-5/2} \xrightarrow[n \rightarrow +\infty]{} \infty$ , we have

$$\begin{aligned}
m (\widehat{r}_n(x) - r(x)) &= m (\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) + m (\mathbb{E}[\widehat{r}_n(x)] - r(x)), \\
&= [n^{-1/2}m^{5/4}] n^{1/2}m^{-1/4} (\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) + \Delta(x)[1 + o(1)],
\end{aligned}$$

and so we obtain Part 2 of Theorem 4.3.1 since in this case we have  $n^{1/2}m^{5/4} \xrightarrow[n \rightarrow +\infty]{} 0$  and (4.7.10).

Now let us prove (4.7.10). We have

$$\begin{aligned}\widehat{r}_n(x) &= \frac{\sum_{i=1}^n Y_i \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x)}{\sum_{i=1}^n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x)}, \\ &= \sum_{i=1}^n w_i Y_i,\end{aligned}$$

with

$$w_i = \frac{\sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} b_k(m-1, x)}{\sum_{i=1}^n \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x)}$$

then

$$\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)] = \sum_{i=1}^n (w_i Y_i - \mathbb{E}[w_i Y_i]).$$

Since  $0 \leq w_i \leq 1$ , for all  $p > 0$ , we obtain  $\mathbb{E}[|w_i Y_i|^{2+p}] = O(1)$  and

$$\sum_{i=1}^n \mathbb{E}[|w_i Y_i|^{2+p}] = O(n).$$

On the other hand, for  $x \in (0, 1)$ , we have

$$v_n^2 = \sum_{i=1}^n \text{Var}[w_i Y_i] = \frac{\text{Var}[Y|X=x]}{f(x)} \psi(x) n m^{1/2} + o(n m^{1/2}),$$

hence

$$\begin{aligned}\frac{1}{v_n^{2+p}} \sum_{i=1}^n \mathbb{E}[|w_i Y_i|^{2+p}] &= O\left(\frac{n}{n^{\frac{2+p}{2}} m^{\frac{2+p}{4}}}\right), \\ &= O\left(n^{-\frac{p}{2}} m^{-\frac{2+p}{4}}\right) = o(1).\end{aligned}$$

Then the convergence in (4.7.10) follows from the application of Lyapounov's theorem.

Now to prove the convergence, for  $x = 0, 1$ , we will use the fact that

$$\sqrt{\frac{n}{m}} (\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{\text{Var}(Y|X=x)}{f(x)}\right), \quad (4.7.11)$$

which will be proved later. We have

$$\begin{aligned}\sqrt{\frac{n}{m}}(\widehat{r}_n(x) - r(x)) &= \sqrt{\frac{n}{m}}\left(\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)] + \sqrt{\frac{n}{m}}(\mathbb{E}[\widehat{r}_n(x)]) - r(x)\right), \\ &= \sqrt{\frac{n}{m}}(\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) + n^{1/2}m^{-3/2}\Delta(x)[1 + o(1)],\end{aligned}$$

so if  $nm^{-3} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then Part 3 of Theorem 4.3.1 follows.

Now, if  $nm^{-3} \xrightarrow[n \rightarrow +\infty]{} \infty$ , we have

$$\begin{aligned}m(\widehat{r}_n(x) - r(x)) &= m(\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) + m(\mathbb{E}[\widehat{r}_n(x)] - r(x)), \\ &= [n^{1/2}m^{3/2}] \sqrt{\frac{n}{m}}(\widehat{r}_n(x) - \mathbb{E}[\widehat{r}_n(x)]) + \Delta(x)[1 + o(1)],\end{aligned}$$

and so we obtain Part 4 of Theorem 4.3.1 since in this case we have  $n^{1/2}m^{5/4} \xrightarrow[n \rightarrow +\infty]{} 0$  and (4.7.11).

To prove (4.7.11), for  $x = 0, 1$ , we have

$$v_n^2 = \sum_{i=1}^n \text{Var} [|w_i Y_i - \mathbb{E}[w_i Y_i]|^{2+p}] = \frac{\text{Var}[Y|X=x]}{f(x)} nm + o(nm),$$

hence

$$\begin{aligned}\frac{1}{v_n^{2+p}} \sum_{i=1}^n \mathbb{E} [|w_i Y_i - \mathbb{E}[w_i Y_i]|^{2+p}] &= O\left(\frac{n}{n^{\frac{2+p}{2}} m^{\frac{2+p}{2}}}\right), \\ &= O\left(n^{-\frac{p}{2}} m^{-\frac{2+p}{2}}\right) = o(1).\end{aligned}$$

Then the convergence in (4.7.11) follows from the application of Lyapounov's theorem.

#### 4.7.4 Proof of the results obtained for $r_n$

From now on, we set  $n_0 \geq 3$  such that  $\forall k \geq n_0$ ,  $\gamma_k \leq (2\|f\|_\infty)^{-1}$  and  $\gamma_k m_k \leq 1$ . Moreover, we introduce the following notations :

$$s_n = \sum_{k=n_0}^n \gamma_k$$

and, for  $s > 0$

$$\begin{aligned} \Pi_n(s) &= \prod_{j=n_0}^n (1 - s\gamma_j), \\ U_{k,n}(s) &= \Pi_n(s)\Pi_k^{-1}(s), \end{aligned}$$

Finally, we define the sequences  $(\lambda_n)$ ,  $(\tilde{\lambda}_n)$ ,  $(\beta_n)$  and  $(\tilde{\beta}_n)$  by setting

$$\begin{aligned} (\lambda_n) &= \begin{cases} \left( \sqrt{\gamma_n m_n^{1/2}} \right) & \text{if } \lim_{n \rightarrow \infty} (\gamma_n m_n^{5/2}) = \infty, \\ (m_n^{-1}) & \text{otherwise.} \end{cases} \\ (\tilde{\lambda}_n) &= \begin{cases} \left( \sqrt{\gamma_n m_n^{1/2}} \ln n \right) & \text{if } \lim_{n \rightarrow \infty} (\gamma_n m_n^{5/2} \ln n) = \infty, \\ (m_n^{-1}) & \text{otherwise.} \end{cases} \\ (\beta_n) &= \begin{cases} \left( \sqrt{\gamma_n m_n} \right) & \text{if } \lim_{n \rightarrow \infty} (\gamma_n m_n^3) = \infty, \\ (m_n^{-1}) & \text{otherwise.} \end{cases} \\ (\tilde{\beta}_n) &= \begin{cases} \left( \sqrt{\gamma_n m_n} \ln n \right) & \text{if } \lim_{n \rightarrow \infty} (\gamma_n m_n^3 \ln n) = \infty, \\ (m_n^{-1}) & \text{otherwise.} \end{cases} \end{aligned} \tag{4.7.12}$$

Note that  $(\lambda_n)$  and  $(\tilde{\lambda}_n)$  belong to  $\mathcal{GS}(-\lambda^*)$  with

$$\lambda^* = \min \left\{ \frac{2\alpha - a}{4}, a \right\},$$

and  $(\beta_n)$  and  $(\tilde{\beta}_n)$  belong to  $\mathcal{GS}(-\beta^*)$  with

$$\beta^* = \min \left\{ \frac{\alpha - a}{2}, a \right\},$$

To establish the characteristics of our recursive estimator  $r_n$  defined by (4.4.1), we state the following technical lemma, which is proved in Mokkadem et al. [67], and which will be used throughout the demonstrations.

**Lemma 4.7.2.**

Let  $(v_n) \in \mathcal{GS}(v^*)$ ,  $(\gamma_n) \in \mathcal{GS}(-\alpha)$  with  $\alpha > 0$ , and set  $l > 0$ . If  $ls - v^*\xi > 0$  (where  $\xi = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}$ ), then

$$\lim_{n \rightarrow \infty} v_n \Pi_n^l(s) \sum_{k=n_0}^n \Pi_k^{-l}(s) \frac{\gamma_k}{v_k} = \frac{1}{ls - v^*\xi}.$$

Moreover, for all positive sequence  $(\alpha_n)$  such that  $\lim_{n \rightarrow \infty} \alpha_n = 0$ , and all  $C$ ,

$$\lim_{n \rightarrow \infty} v_n \Pi_n^l(s) \left[ \sum_{k=n_0}^n \Pi_k^{-l}(s) \frac{\gamma_k}{v_k} \alpha_k + C \right] = 0.$$

We first give the asymptotic behaviour of  $(\rho_n)$  defined in (4.4.2). Then, we show how the asymptotic behaviour of  $(r_n)$  (4.4.1) can be deduced from that of  $(\rho_n)$ .

**4.7.4.1 Asymptotic behaviour of  $\rho_n$**

The following Lemma gives the bias and the variance of the estimator  $\rho_n$  defined in (4.4.2).

**Lemma 4.7.3.** (*Bias and Variance of  $\rho_n$* )

Let Assumptions (A1) – (A4) hold. For  $x \in (0, 1)$ , such that  $f(x) > 0$ , we have

1. If  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , then

$$\mathbb{E}[\rho_n(x)] - r(x) = m_n^{-1} \frac{f(x)}{f(x) - a\xi} \Delta(x) + o(m_n^{-1}). \quad (4.7.13)$$

- If  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , then

$$\mathbb{E}[\rho_n(x)] - r(x) = o\left(\sqrt{\gamma_n m_n^{1/2}}\right). \quad (4.7.14)$$

2. If  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , then

$$\begin{aligned} \text{Var}[\rho_n(x)] &= \frac{2f(x)}{4f(x) - (2\alpha - a)} \text{Var}(Y|X = x) \psi(x) \gamma_n m_n^{1/2} \\ &\quad + o(\gamma_n m_n^{1/2}). \end{aligned} \quad (4.7.15)$$

If  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , then

$$\text{Var} [\rho_n(x)] = o(m_n^{-2}). \quad (4.7.16)$$

Let Assumptions (A1) – (A'4) hold. For  $x = 0, 1$ , such that  $f(x) > 0$ , we have

1. If  $\frac{1-\alpha}{4} < a \leq \frac{\alpha}{3}$ , then

$$\mathbb{E} [\rho_n(x)] - r(x) = m_n^{-1} \frac{f(x)}{f(x) - a\xi} \Delta(x) + o(m_n^{-1}). \quad (4.7.17)$$

If  $\frac{\alpha}{3} < a < \frac{2}{3}\alpha$ , then

$$\mathbb{E} [\rho_n(x)] - r(x) = o(\sqrt{\gamma_n m_n}). \quad (4.7.18)$$

2. If  $\frac{\alpha}{3} < a < \frac{2}{3}\alpha$ , then

$$\begin{aligned} \text{Var} [\rho_n(x)] &= \frac{f(x)}{2f(x) - (\alpha - a)} \text{Var}(Y|X = x) \psi(x) \gamma_n m_n \\ &\quad + o(\gamma_n m_n). \end{aligned} \quad (4.7.19)$$

If  $\frac{1-\alpha}{4} < a \leq \frac{\alpha}{3}$ , then

$$\text{Var} [\rho_n(x)] = o(m_n^{-2}). \quad (4.7.20)$$

We have, for  $n \geq n_0$ ,

$$\begin{aligned} \rho_n(x) - r(x) &= (1 - \gamma_n f(x)) (\rho_{n-1}(x) - r(x)) + \gamma_n (W_n(x) - r(x) Z_n(x)), \\ &= \Pi_n(f(x)) \sum_{k=n_0}^n \Pi_k^{-1}(f(x)) \gamma_k (W_k(x) - r(x) Z_k(x)) \\ &\quad + \Pi_n(f(x)) (\rho_{n_0-1}(x) - r(x)), \\ &= T_n(x) + R_n(x), \end{aligned}$$

where, since  $\rho_{n_0-1}(x) = r_{n_0-1}(x)$ ,

$$\begin{aligned} T_n(x) &= \sum_{k=n_0}^n U_{k,n}(f(x)) \gamma_k (W_k(x) - r(x) Z_k(x)), \\ R_n(x) &= \Pi_n(f(x)) (r_{n_0-1}(x) - r(x)). \end{aligned}$$

Noting that  $|r_{n_0-1}(x) - r(x)| = O(1)$  a.s. and applying Lemma 4.7.2, we get

$$\begin{aligned} R_n(x) &= O(\Pi_n(f(x))) \quad a.s. \\ &= \begin{cases} o(\lambda_n) & \text{for } x \in (0, 1), \\ o(\beta_n) & \text{for } x = 0, 1. \end{cases} \quad a.s. \end{aligned} \quad (4.7.21)$$

This remark implies that Lemma 4.7.3 hold when  $\rho_n(x)$  is replaced by  $T_n(x)$ . Then, for  $x \in [0, 1]$  such that  $f(x) > 0$ , we have

$$\mathbb{E} [T_n(x)] = \sum_{k=n_0}^n U_{k,n}(f(x))\gamma_k (\mathbb{E} [W_k(x)] - r(x)\mathbb{E} [Z_k(x)]),$$

where

$$\begin{aligned} \mathbb{E} [Z_k(x)] &= m \sum_{k=0}^{m-1} \left[ F \left( \frac{k+1}{m} \right) - F \left( \frac{k}{m} \right) \right] b_k(m-1, x), \\ &= f(x) + \Delta_1(x)m_k^{-1} + o(m_k^{-1}), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} [W_k(x)] &= m \mathbb{E} \left[ Y \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X \leq \frac{k+1}{m}\}} b_k(m-1, x) \right], \\ &= r(x)f(x) + \Delta_2(x)m_k^{-1} + o(m_k^{-1}). \end{aligned}$$

Then, we obtain

$$\mathbb{E} [T_n(x)] = f(x)\Delta(x) \sum_{k=n_0}^n U_{k,n}(f(x))\gamma_k [m_k^{-1} + o(m_k^{-1})],$$

On the other hand, we have

$$Var [T_n(x)] = \sum_{k=n_0}^n U_{k,n}^2(f(x))\gamma_k^2 \{Var [W_k(x)] + r^2(x)Var [Z_k(x)] - 2r(x)Cov (W_k(x), Z_k(x))\},$$

where

$$\begin{aligned} Var [W_k(x)] &= \begin{cases} \mathbb{E}[Y^2|X=x]f(x)\psi(x)m_k^{1/2} + o_x(m_k^{1/2}) & \text{for } x \in (0, 1), \\ \mathbb{E}[Y^2|X=x]f(x)m_k + o_x(m_k) & \text{for } x = 0, 1. \end{cases} \\ Var [Z_k(x)] &= \begin{cases} f(x)\psi(x)m_k^{1/2} + o_x(m_k^{1/2}) & \text{for } x \in (0, 1), \\ f(x)m_k + o_x(m_k) & \text{for } x = 0, 1. \end{cases} \\ Cov (W_k(x), Z_k(x)) &= \begin{cases} r(x)f(x)\psi(x)m_k^{1/2} + o_x(m_k^{1/2}) & \text{for } x \in (0, 1), \\ r(x)f(x)m_k + o_x(m_k) & \text{for } x = 0, 1. \end{cases} \end{aligned}$$

which gives

$$\text{Var} [T_n(x)] = \begin{cases} f(x)\text{Var}(Y|X=x)\psi(x) \sum_{k=n_0}^n U_{k,n}^2(f(x))\gamma_k^2 \left[ m_k^{1/2} + o_x \left( m_k^{1/2} \right) \right] & \text{for } x \in (0, 1), \\ f(x)\text{Var}(Y|X=x) \sum_{k=n_0}^n U_{k,n}^2(f(x))\gamma_k^2 [m_k + o_x(m_k)] & \text{for } x = 0, 1. \end{cases}$$

Then, for  $x \in (0, 1)$ , we obtain

— In the case  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > a/f(x)$  and Lemma 4.7.2 gives (4.7.13).

In the case  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , we have  $m_n^{-1} = o\left(\sqrt{\gamma_n m_n^{1/2}}\right)$ , since  $a > (2\alpha - a)/4$  and Lemma 4.7.2 gives (4.7.14).

— In the case  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > (2\alpha - a)/(4f(x))$  and Lemma 4.7.2 gives (4.7.15).

In the case  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , we have  $\gamma_n m_n^{1/2} = o(m_n^{-2})$ , since  $a \leq (2\alpha - a)/4$  and Lemma 4.7.2 gives (4.7.16).

Now, for  $x = 0, 1$ , we obtain

— In the case  $\frac{1-\alpha}{4} < a \leq \frac{\alpha}{3}$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > a/f(x)$  and Lemma 4.7.2 gives (4.7.17).

In the case  $\frac{\alpha}{3} < a < \frac{2}{3}\alpha$ , we have  $m_n^{-1} = o\left(\sqrt{\gamma_n m_n}\right)$ , since  $a > (\alpha - a)/2$  and Lemma 4.7.2 gives (4.7.18).

— In the case  $\frac{\alpha}{3} < a < \frac{2}{3}\alpha$ , we have  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a)/(2f(x))$  and Lemma 4.7.2 gives (4.7.19).

In the case  $\frac{1-\alpha}{4} < a \leq \frac{\alpha}{3}$ , we have  $\gamma_n m_n = o(m_n^{-2})$ , since  $a \leq (\alpha - a)/2$  and Lemma 4.7.2 gives (4.7.20).

The following lemma gives the weak convergence rate of the estimator  $\rho_n$  defined in (4.4.3), for  $x \in [0, 1]$  such that  $f(x) > 0$ .

**Lemma 4.7.4.** *(Weak convergence rate of  $\rho_n$ )*

Let Assumption (A1) – (A4) hold. For  $x \in (0, 1)$ , we have

1. If  $\gamma_n^{-1} m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2} m_n^{-1/4} (\rho_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( \sqrt{c} \frac{f(x)\Delta(x)}{f(x) - a\xi}, \frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi} \right),$$



2. If  $\gamma_n^{-1}m_n^{-5/2} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n (\rho_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{f(x)\Delta(x)}{f(x) - a\xi},$$

Let Assumption (A1) – (A3) and (A'4) hold. For  $x = 0, 1$ , we have

1. If  $\gamma_n^{-1}m_n^{-3} \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then

$$\gamma_n^{-1/2}m_n^{-1/2} (\rho_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( \sqrt{c} \frac{f(x)\Delta(x)}{f(x) - a\xi}, \frac{f(x)\text{Var}(Y|X=x)\psi(x)}{2f(x) - (\alpha - a)\xi} \right),$$

2. If  $\gamma_n^{-1}m_n^{-3} \xrightarrow[n \rightarrow +\infty]{} \infty$ , then

$$m_n (\rho_n(x) - r(x)) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{f(x)\Delta(x)}{f(x) - a\xi},$$

To prove Lemma 4.7.4, for  $x \in (0, 1)$ , we will use the fact that if  $\frac{2}{5}\alpha \leq a < \frac{2}{3}\alpha$ , then

$$\gamma_n^{-1/2}m_n^{-1/4} (\rho_n(x) - \mathbb{E}[\rho_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( \frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi} \right), \quad (4.7.22)$$

which will be proved later. In the case  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , Part 1 of Lemma 4.7.4 follows from the combination (4.7.22) and (4.7.14). In the case  $a = \frac{2}{5}\alpha$ , Part 1 and 2 of Lemma 4.7.4 follows from the combination of (4.7.13) and (4.7.22). In the case  $\frac{1-\alpha}{4} < a < \frac{2}{5}\alpha$ , (4.7.18) implies that

$$m_n (\rho_n(x) - \mathbb{E}[\rho_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$$

and the application of (4.7.13) gives Part 2 of Lemma 4.7.2.

In view of (4.7.22), the result in (4.7.22) hold if we replace  $\rho_n(x)$  by  $T_n(x)$ . Set

$$\eta_k(x) = \Pi_k^{-1}(f(x))\gamma_k (W_k(x) - r(x)Z_k(x)), \quad (4.7.23)$$

so that

$$T_n(x) - \mathbb{E}[T_n(x)] = \Pi_n(f(x)) \sum_{k=n_0}^n (\eta_k(x) - \mathbb{E}[\eta_k(x)]).$$

For  $x \in (0, 1)$ , We have

$$\text{Var} [\eta_k(x)] = \Pi_k^{-2}(f(x))\gamma_k^2 m_k^{1/2} [f(x)\psi(x)\text{Var}(Y|X = x) + o(1)],$$

and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > (2\alpha - a)/(4f(x))$ , Lemma 4.7.2 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=n_0}^n \text{Var} [\eta_k(x)] \\ &= \sum_{k=n_0}^n \Pi_k^{-2}(f(x))\gamma_k^2 m_k^{1/2} [f(x)\psi(x)\text{Var}(Y|X = x) + o(1)] \\ &= \frac{2\Pi_n^{-2}(f(x))\gamma_n m_n^{1/2}}{4f(x) - (2\alpha - a)\xi} [f(x)\psi(x)\text{Var}(Y|X = x) + o(1)]. \end{aligned} \quad (4.7.24)$$

On the other hand, for all  $p > 0$  and  $x \in [0, 1]$ , using Lemma 4.7.1 (ii) we have

$$\begin{aligned} &\mathbb{E} \left( |Y_k - r(x)|^{2+p} \left\{ \sum_{i=0}^{m-1} \mathbb{I}_{\{\frac{i}{m} < X_n \leq \frac{i+1}{m}\}} b_k(m-1, x) \right\}^{2+p} \right) \\ &\leq \left\{ \sum_{i=0}^{m-1} b_k^2(m-1, x) \right\}^{(2+p)/2} \int_0^1 \int_{\mathbb{R}} |y - r(x)|^{2+p} g(z, y) dz dy \\ &\leq m_n^{-(2+p)/4} \int_0^1 \left\{ \int_{\mathbb{R}} |y|^{2+p} g(z, y) dy + |r(x)|^{2+p} \int_{\mathbb{R}} g(z, y) dy \right\} dz \\ &= O(m_n^{-(2+p)/4}), \end{aligned} \quad (4.7.25)$$

and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a/2)/(2f(x))$ , there exists a  $p > 0$  such that  $\lim_{n \rightarrow \infty} (n\gamma_n) > (1+p)(\alpha - a/2)/(2+p)(f(x)) > \frac{(1+p)\alpha - (3(2+p)/4)a}{2+p}$ , so Lemma 4.7.2 gives

$$\begin{aligned} &\sum_{k=n_0}^n \mathbb{E} [|\eta_k(x)|^{2+p}] \\ &= O \left( \sum_{k=n_0}^n \Pi_k^{-2-p}(f(x))\gamma_k^{2+p} m_k^{2+p} \mathbb{E} \left( |Y_k - r(x)|^{2+p} \left\{ \sum_{i=0}^{m-1} \mathbb{I}_{\{\frac{i}{m} < X_n \leq \frac{i+1}{m}\}} b_k(m-1, x) \right\}^{2+p} \right) \right) \\ &= O \left( \sum_{k=n_0}^n \Pi_k^{-2-p}(f(x))\gamma_k^{2+p} m_k^{3(2+p)/4} \right) \\ &= O \left( \frac{\gamma_n^{1+p} m_n^{3(2+p)/4}}{\Pi_n^{2+p}(f(x))} \right), \end{aligned}$$

hence

$$\frac{1}{v_n^{2+p}} \sum_{k=n_0}^n \mathbb{E} [|\eta_k(x)|^{2+p}] = O \left( m_n (\gamma_n m_n)^{p/2} \right),$$

and the convergence in (4.7.22) follows from the application of Lyapounov's Theorem.

To prove Lemma 4.7.4, for  $x = 0, 1$ , we will use the fact that if  $\frac{\alpha}{3} \leq a < \frac{2}{3}\alpha$ , then

$$\gamma_n^{-1/2} m_n^{-1/2} (\rho_n(x) - \mathbb{E}[\rho_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( \frac{f(x) \text{Var}(Y|X=x)}{2f(x) - (\alpha - a)\xi} \right), \quad (4.7.26)$$

which will be proved later. In the case  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , Part 3 of Lemma 4.7.4 follows from the combination (4.7.26) and (4.7.16). In the case  $a = \frac{2}{5}\alpha$ , Part 3 and 4 of Lemma 4.7.4 follows from the combination of (4.7.15) and (4.7.26). In the case  $\frac{1-\alpha}{4} < a < \frac{2}{5}\alpha$ , (4.7.20) implies that

$$m_n (\rho_n(x) - \mathbb{E}[\rho_n(x)]) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$$

and the application of (4.7.15) gives Part 4 of Lemma 4.7.2.

In view of (4.7.22), the result in (4.7.26) hold if we replace  $\rho_n(x)$  by  $T_n(x)$ .

For  $x = 0, 1$ , We have

$$\text{Var}[\eta_k(x)] = \Pi_k^{-2}(f(x)) \gamma_k^2 m_k [f(x) \text{Var}(Y|X=x) + o(1)],$$

and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a)/(2f(x))$ , Lemma 4.7.2 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=n_0}^n \text{Var}[\eta_k(x)] \\ &= \sum_{k=n_0}^n \Pi_k^{-2}(f(x)) \gamma_k^2 m_k [f(x) \text{Var}(Y|X=x) + o(1)] \\ &= \frac{2\Pi_n^{-2}(f(x)) \gamma_n m_n}{4f(x) - (2\alpha - a)\xi} [f(x) \text{Var}(Y|X=x) + o(1)]. \end{aligned}$$

On the other hand, there exists a  $p > 0$  such that

$$\sum_{k=n_0}^n \mathbb{E}[|\eta_k(x)|^{2+p}] = O \left( \frac{\gamma_n^{1+p} m_n^{3(2+p)/4}}{\Pi_n^{2+p}(f(x))} \right),$$

hence

$$\frac{1}{v_n^{2+p}} \sum_{k=n_0}^n \mathbb{E}[|\eta_k(x)|^{2+p}] = O \left( m_n^{1/2} (\gamma_n m_n^{1/2})^{p/2} \right),$$

and the convergence in (4.7.26) follows from the application of Lyapounov's Theorem.

The following lemma gives the strong pointwise convergence rate of  $\rho_n$ , for  $x \in [0, 1]$  such that  $f(x) > 0$ .

**Lemma 4.7.5.** *(Strong pointwise convergence rate of  $\rho_n$ )*

Let Assumption (A1) – (A4) hold. For  $x \in (0, 1)$ , we have

1. If  $\gamma_n^{-1}m_n^{-5/2}/\ln(s_n) \xrightarrow[n \rightarrow +\infty]{} c$  for some constant  $c \geq 0$ , then, with probability one, the sequence

$$\left( \sqrt{\frac{\gamma_n^{-1}m_n^{-1/2}}{2\ln(s_n)}} (\rho_n(x) - r(x)) \right)$$

is relatively compact and its limit set is the interval

$$\left[ \sqrt{\frac{c}{2}} \frac{f(x)\Delta(x)}{f(x) - a\xi} - \sqrt{\frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi}}, \right. \\ \left. \sqrt{\frac{c}{2}} \frac{f(x)\Delta(x)}{f(x) - a\xi} + \sqrt{\frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi}} \right].$$

2. If  $\gamma_n^{-1}m_n^{-5/2}/\ln(s_n) \rightarrow \infty$ , then with probability one

$$\lim_{n \rightarrow \infty} m_n (\rho_n(x) - r(x)) = \frac{f(x)\Delta(x)}{f(x) - a\xi}.$$

Let Assumption (A1) – (A3) and (A'4) hold. For  $x = 0, 1$ , we have

1. If  $\gamma_n^{-1}m_n^{-3}/\ln(s_n) \rightarrow c$  for some constant  $c \geq 0$ , then, with probability one, the sequence

$$\left( \sqrt{\frac{\gamma_n^{-1}m_n^{-1}}{2\ln(s_n)}} (\rho_n(x) - r(x)) \right)$$

is relatively compact and its limit set is the interval

$$\left[ \sqrt{\frac{c}{2}} \frac{f(x)\Delta(x)}{f(x) - a\xi} - \sqrt{\frac{f(x)\text{Var}(Y|X=x)\psi(x)}{2f(x) - (\alpha - a)\xi}}, \right. \\ \left. \sqrt{\frac{c}{2}} \frac{f(x)\Delta(x)}{f(x) - a\xi} + \sqrt{\frac{f(x)\text{Var}(Y|X=x)\psi(x)}{2f(x) - (\alpha - a)\xi}} \right].$$

2. If  $\gamma_n^{-1}m_n^{-3}/\ln(s_n) \rightarrow \infty$ , then, with probability one

$$\lim_{n \rightarrow \infty} m_n (\rho_n(x) - r(x)) = \frac{f(x)\Delta(x)}{f(x) - a\xi},$$

In view of (4.7.22), Lemma 4.7.5 hold when  $\rho_n(x) - r(x)$  is replaced by  $T_n(x)$ .

Set

$$B_n(x) = \sum_{k=n_0}^n (\eta_k(x) - \mathbb{E}[\eta_k(x)]),$$

where  $\eta_k$  is defined in (4.7.23). For  $x \in (0, 1)$

— In the case  $a \geq \frac{2}{5}\alpha$  (in which case  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - a/2)/(2f(x))$ ). We set

$H_n^2(f(x)) = \Pi_n^2(f(x))\gamma_n^{-1}m_n^{-1/2}$ , and note that, since

$(\gamma_n^{-1}m_n^{-1/2}) \in \mathcal{GS}(\alpha - a/2)$ , we have

$$\begin{aligned} & \ln(H_n^{-2}(f(x))) \\ &= -2 \ln(\Pi_n(f(x))) + \ln\left(\prod_{k=n_0}^n \frac{\gamma_{k-1}^{-1}m_{k-1}^{-1/2}}{\gamma_k^{-1}m_k^{-1/2}}\right) + \ln(\gamma_{n_0-1}m_{n_0-1}^{1/2}) \\ &= -2 \sum_{k=n_0}^n \ln(1 - f(x)\gamma_k) + \sum_{k=n_0}^n \ln\left(1 - \frac{\alpha - a/2}{k} + o\left(\frac{1}{k}\right)\right) + \ln(\gamma_{n_0-1}m_{n_0-1}^{1/2}) \\ &= \sum_{k=n_0}^n (2f(x)\gamma_k + o(g_k)) - \sum_{k=n_0}^n ((\alpha - a/2)\xi\gamma_k + o(\gamma_k)) + \ln(\gamma_{n_0-1}m_{n_0-1}^{1/2}) \\ &= (2f(x) - (\alpha - a/2)\xi) s_n + o(s_n). \end{aligned} \tag{4.7.27}$$

Since  $2f(x) - (\alpha - a/2)\xi > 0$ , it follows in particular that

$\lim_{n \rightarrow \infty} H_n^{-2}(f(x)) = \infty$ . Moreover, we have  $\lim_{n \rightarrow \infty} H_n^2(f(x))/H_{n-1}^2(f(x)) = 1$ , and

applying Lemma 4.7.2 and (4.7.24), we get

$$\begin{aligned} & \lim_{n \rightarrow \infty} H_n^2(f(x)) \sum_{k=n_0}^n \text{Var}[\eta_k(x)] \\ &= [2f(x) - (\alpha - a/2)\xi]^{-1} f(x)\psi(x)\text{Var}(Y|X = x) \end{aligned}$$

and, in view of (4.7.25)

$$\mathbb{E}[|\eta_k(x)|^3] = O(\Pi_n^{-3}(f(x))\gamma_n^3m_n^{9/4}).$$

Now, since  $(\gamma_n^{-1}m_n^{1/2}) \in \mathcal{GS}(\alpha - a/2)$ , applying Lemma 4.7.2 and (4.7.27), we

get

$$\begin{aligned}
& \frac{1}{n\sqrt{n}} \sum_{k=n_0}^n \mathbb{E} (|H_n(f(x))\eta_k(x)|^3) \\
&= O \left( \frac{H_n^3(f(x))}{n\sqrt{n}} \left( \sum_{k=n_0}^n \Pi_k^{-3}(f(x))g_k^3m_k^{9/4} \right) \right) \\
&= O \left( \frac{\Pi_n^3(f(x))\gamma_n^{-3/2}m_n^{-3/4}}{n\sqrt{n}} \left( \sum_{k=n_0}^n \Pi_k^{-3}(f(x))g_k m_k o \left( \left( g_k m_k^{1/2} \right)^{3/2} \right) \right) \right) \\
&= o \left( \frac{m_n}{n\sqrt{n}} \right)
\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier [65] then ensures that, with probability one, the sequence

$$\left( \frac{H_n(f(x))B_n(x)}{\sqrt{2 \ln \ln (H_n^{-2}(f(x)))}} \right) = \left( \frac{\sqrt{\gamma_n^{-1}m_n^{1/2}} (T_n(x) - \mathbb{E} [T_n(x)])}{\sqrt{2 \ln \ln (H_n^{-2}(f(x)))}} \right)$$

is relatively compact and its limit set is the interval

$$\left[ -\sqrt{\frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi}}, \sqrt{\frac{2f(x)\text{Var}(Y|X=x)\psi(x)}{4f(x) - (2\alpha - a)\xi}} \right]. \quad (4.7.28)$$

In view (4.7.27), we have  $\lim_{n \rightarrow \infty} \ln \ln (H_n^{-2}(f(x))) / \ln(s_n) = 1$ . It, follows that, with probability one, the sequence  $\left( \sqrt{\gamma_n^{-1}m_n^{1/2}} (T_n(x) - \mathbb{E} [T_n(x)]) / \sqrt{2 \ln(s_n)} \right)$  is relatively compact and its limit set is the interval given in (4.7.28).

The application of (4.7.14) (respectively (4.7.13)) concludes the proof of Lemma 4.7.5 in the cases  $a > 2\alpha/5$  (respectively  $a = 2\alpha/5$ )

— In the case  $a > \frac{2}{5}\alpha$  (in which case  $\lim_{n \rightarrow \infty} (n\gamma_n) > a/f(x)$ ).

Set  $H_n^{-2}(f(x)) = \Pi_n^{-2}(f(x))m_n^{-2} (\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2}))^{-1}$ , and note that, since  $(m_n^2) \in \mathcal{GS}(2a)$ , we have

$$\begin{aligned}
& \ln (\Pi_n^{-2}(f(x))m_n^{-2}) \\
&= -2 \ln (\Pi_n(f(x))) + \ln \left( \prod_{k=n_0}^n \frac{m_{k-1}^2}{m_k^2} \right) + \ln (m_{n_0-1}^{-2}) \\
&= -2 \sum_{k=n_0}^n \ln (1 - f(x)\gamma_k) + \sum_{k=n_0}^n \ln \left( 1 - \frac{2a}{k} + o \left( \frac{1}{k} \right) \right) + \ln (\gamma_{n_0-1}m_{n_0-1}^{1/2}) \\
&= \sum_{k=n_0}^n (2f(x)\gamma_k + o(g_k)) - \sum_{k=n_0}^n (2a\xi\gamma_k + o(\gamma_k)) + \ln (\gamma_{n_0-1}m_{n_0-1}^{1/2}) \\
&= (2f(x) - 2a\xi) s_n + o(s_n). \quad (4.7.29)
\end{aligned}$$

Since  $2f(x) - 2a\xi > 0$ , it follows in particular that  $\lim_{n \rightarrow \infty} \Pi_n^{-2}(f(x))m_n^{-2} = \infty$  and thus  $\lim_{n \rightarrow \infty} H_n^{-2}(f(x)) = \infty$ . Moreover, we have  $\lim_{n \rightarrow \infty} H_n^2(f(x))/H_{n-1}^2(f(x)) = 1$ . Now, set  $\varepsilon \in ]0, \alpha - 5a/2[$  such that  $\lim_{n \rightarrow \infty} (n\gamma_n) > a/f(x) + \varepsilon/2$ ; in view of (4.7.24), and applying Lemma 4.7.2, we get

$$\begin{aligned} & H_n^2(f(x)) \sum_{k=n_0}^n \text{Var} [\eta_k(x)] \\ &= O \left( \Pi_n^2(f(x))m_n^2 (\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2})) \sum_{k=n_0}^n \Pi_k^{-2}(f(x))\gamma_k m_k^{1/2} \right) \\ &= O \left( \Pi_n^2(f(x))m_n^2 (\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2})) \sum_{k=n_0}^n \Pi_k^{-2}(f(x))\gamma_k o(m_k^{-2}k^{-\varepsilon}) \right) \\ &= o(1). \end{aligned}$$

Moreover, in view of (4.7.25)

$$\mathbb{E} [|\eta_k(x)|^3] = O(\Pi_n^{-3}(f(x))\gamma_n^3 m_n^{9/4}),$$

and thus in view of (4.7.29), we get

$$\begin{aligned} & \frac{1}{n\sqrt{n}} \sum_{k=n_0}^n \mathbb{E} (|H_n(f(x))\eta_k(x)|^3) \\ &= O \left( \frac{H_n^3(f(x))}{n\sqrt{n}} (\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2}))^{3/2} \left( \sum_{k=n_0}^n \Pi_k^{-3}(f(x))g_k^3 m_k^{9/4} \right) \right) \\ &= O \left( \frac{\Pi_n^3(f(x))m_n^3}{n\sqrt{n}} (\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2}))^{3/2} \left( \sum_{k=n_0}^n \Pi_k^{-3}(f(x))g_k o(m_k^{-3}) \right) \right) \\ &= O \left( \frac{(\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2}))^{3/2}}{n\sqrt{n}} \right) \\ &= o \left( [\ln (H_n^{-2}(f(x)))]^{-1} \right) \end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier [65] then ensures that, with probability one

$$\lim_{n \rightarrow \infty} \frac{H_n(f(x))B_n(x)}{\sqrt{2 \ln \ln (H_n^{-2}(f(x)))}} = \lim_{n \rightarrow \infty} m_n^{-1} \frac{\sqrt{\ln \ln (\Pi_n^{-2}(f(x))m_n^{-2})}}{\sqrt{2 \ln \ln (H_n^{-2}(f(x)))}} (T_n(x) - \mathbb{E} [T_n(x)]) = 0$$

Noting that (4.7.29) ensures that

$$\lim_{n \rightarrow \infty} \ln \ln (H_n^{-2}(f(x))) / \ln \ln (\Pi_n^{-2}(f(x))m_n^{-2}) = 1,$$

we get

$$\lim_{n \rightarrow \infty} m_n^{-1} (T_n(x) - \mathbb{E} [T_n(x)]) = 0 \quad a.s.$$

and Lemma 4.7.5 in the case  $a < 2\alpha/5$  follows from (4.7.13). We use the same method to prove Lemma 4.7.5 in the cases  $x = 0, 1$ .

Now, we give the *MISE* of  $\rho_n$ .

**Lemma 4.7.6.** (*MISE of  $\rho_n$* )

Let Assumptions (A1) – (A3) and (A'4) hold, we have

1. If  $\frac{1-\alpha}{4} < a < \frac{2}{5}\alpha$ , then

$$MISE(\rho_n) = K_1 m_n^{-2} + o(m_n^{-2}).$$

2. If  $a = \frac{2}{5}\alpha$ , then

$$MISE(\rho_n) = K_1 m_n^{-2} + K_2 \gamma_n m_n^{1/2} + o(m_n^{-2} + \gamma_n m_n^{1/2}).$$

3. If  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , then

$$MISE(\rho_n) = K_2 \gamma_n m_n^{1/2} + o(\gamma_n m_n^{1/2}).$$

We have

$$MISE(\rho_n) = \int_0^1 Bias^2[\rho_n(x)] dx + \int_0^1 Var[\rho_n(x)] dx.$$

— In the case  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , since, for all  $x \in (0, 1)$ ,

$\lim_{n \rightarrow \infty} (n\gamma_n) > a/\varphi > a/f(x)$ , we have

$$\begin{aligned} \int_0^1 Bias^2[\rho_n(x)] dx &= \int_0^1 \left[ m_n^{-1} \frac{f(x)}{f(x) - a\xi} \Delta(x) + o(m_n^{-1}) \right]^2 dx \\ &= K_1 m_n^{-2} + o(m_n^{-2}). \end{aligned} \quad (4.7.30)$$

— In the case  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , we have  $m_n^{-1} = o(\gamma_n m_n^{1/2})$ , then Lemma 4.7.2 gives

$$\begin{aligned} \int_0^1 Bias^2[\rho_n(x)] dx &= \int_0^1 \Pi_n(f(x)) \sum_{k=n_0}^n \Pi_k^{-1}(f(x)) \gamma_k o(\gamma_k m_k^{1/2}) dx \\ &= o(\gamma_n m_n^{1/2}) \end{aligned} \quad (4.7.31)$$



On the other hand, we note that

$$\begin{aligned}
\int_0^1 \text{Var} [\rho_n(x)] dx &= \int_0^1 \text{Var} [T_n(x)] dx \\
&= \int_0^1 \sum_{k=n_0}^n U_{k,n}^2(f(x)) \gamma_k^2 \left\{ \text{Var} [W_k(x)] + r^2(x) \text{Var} [Z_k(x)] \right. \\
&\quad \left. - 2r(x) \text{Cov} (W_k(x), Z_k(x)) \right\} dx,
\end{aligned}$$

Using the same argument as in the proof of Proposition 4.3.2, we obtain

— In the case  $\frac{2}{5}\alpha < a < \frac{2}{3}\alpha$ , since, for all  $x \in (0, 1)$ ,

$\lim_{n \rightarrow \infty} (n\gamma_n) > (2\alpha - a)/(4\varphi) > (2\alpha - a)/(4f(x))$ , Lemma 4.7.2 gives

$$\int_0^1 \text{Var} [\rho_n(x)] dx = K_2 \gamma_n^2 m_n^{1/2} + o(\gamma_n m_n^{1/2}). \quad (4.7.32)$$

— In the case  $\frac{1-\alpha}{4} < a \leq \frac{2}{5}\alpha$ , we have  $\gamma_n m_n^{1/2} = o(m_n^{-2})$  and Lemma 4.7.2 gives

$$\begin{aligned}
\int_0^1 \text{Var} [\rho_n(x)] dx &= \int_0^1 \Pi_n^2(f(x)) \sum_{k=n_0}^n \Pi_k^{-2}(f(x)) o(m_k^{-2}) dx \\
&= o(m_n^{-2}).
\end{aligned} \quad (4.7.33)$$

Then, Part 1 of Lemma 4.7.6 follows from the combination of (4.7.30) and (4.7.33), Part 2 from that of (4.7.30) and (4.7.32) and Part 3 from (4.7.31) and (4.7.32).

#### 4.7.4.2 Asymptotic behaviour of $r_n$

We show in this section how to deduce the asymptotic behaviour of  $r_n$  from that of  $\rho_n$ . To do so, we set

$$\delta_n(x) = r_n(x) - \rho_n(x),$$

and we prove that  $\delta_n$  is negligible in front of  $\rho_n$ . Note that, in view of (4.4.1) and (4.4.2), and since  $\rho_{n_0-1} = r_{n_0-1}$ , we have, for  $n \geq n_0$

$$\begin{aligned}
\delta_n(x) &= (1 - \gamma_n f(x)) \delta_{n-1}(x) + \gamma_n (f(x) - Z_n(x)) (r_{n-1}(x) - r(x)) \\
&= \sum_{k=n_0}^n U_{k,n}(f(x)) \gamma_k (f(x) - Z_k(x)) (r_{k-1}(x) - r(x)).
\end{aligned} \quad (4.7.34)$$

To obtain an upper bound of  $\delta_n$ , we must have an upper bound of  $r_n - r$ . To do so, we use the following property given by Mokkadem et al. [67]

( $\mathcal{P}$ ) : if  $(r_n - r)$  is known to be bounded almost surely by a sequence  $(w_n)$ , then it can be shown that  $(\delta_n)$  is bounded almost surely by a sequence  $(w'_n)$  such that  $\lim_{n \rightarrow \infty} w'_n w_n = 0$ , which may allow to upper bound  $r_n - r$  by a sequence smaller than  $(w_n)$ .

We thus proceed as follows. We first establish an upper bound of  $(r_n - r)$ . Then, we apply the Property ( $\mathcal{P}$ ) several times until we obtain an upper bound which allows to prove that  $\delta_n$  is negligible in front of  $\rho_n$ .

The proof of the results given in Section 4.4 relies on the repeated application of the following lemma

**Lemma 4.7.7.**

Let Assumptions (A1) – (A3) hold, and assume that there exists  $(w_n) \in \mathcal{GS}(w^*)$  such that  $|r_n(x) - r(x)| = O(w_n)$  a.s.

For  $x \in (0, 1)$ , we have

1. If the sequence  $(n\gamma_n)$  is bounded, if  $\lim_{n \rightarrow \infty} (n\gamma_n) > \min \{a/f(x), (2\alpha - a)/(4f(x))\}$ , and if  $w^* > 0$ , then, for all  $\delta > 0$ ,

$$|\delta_n(x)| = O\left(\lambda_n w_n (\ln n)^{\frac{1+\delta}{2}}\right) + o(\lambda_n) \quad a.s.$$

2. If  $\lim_{n \rightarrow \infty} (n\gamma_n) = \infty$ , then, for all  $\delta > 0$ ,

$$|\delta_n(x)| = O\left(\lambda_n w_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}}\right) \quad a.s.$$

For  $x = 0, 1$ , we have

1. If the sequence  $(n\gamma_n)$  is bounded, if  $\lim_{n \rightarrow \infty} (n\gamma_n) > \min \{a/f(x), (\alpha - a)/(2f(x))\}$ , and if  $w^* > 0$ , then, for all  $\delta > 0$ ,

$$|\delta_n(x)| = O\left(\beta_n w_n (\ln n)^{\frac{1+\delta}{2}}\right) + o(\beta_n) \quad a.s.$$

2. If  $\lim_{n \rightarrow \infty} (n\gamma_n) = \infty$ , then, for all  $\delta > 0$ ,

$$|\delta_n(x)| = O\left(\beta_n w_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}}\right) \quad a.s.$$

To prove this Lemma, we use (4.7.34) to write

$$\delta_n(x) = \delta_n^{(1)}(x) + \delta_n^{(2)}(x),$$

with

$$\delta_n^{(1)}(x) = \sum_{k=n_0}^n U_{k,n}(f(x)) \gamma_k (\mathbb{E}[Z_k(x)] - Z_k(x)) (r_{k-1}(x) - r(x)),$$

$$\delta_n^{(2)}(x) = \sum_{k=n_0}^n U_{k,n}(f(x)) \gamma_k (f(x) - \mathbb{E}[Z_k(x)]) (r_{k-1}(x) - r(x)).$$

For  $x \in [0, 1]$ , we have

$$\mathbb{E}[Z_k(x)] = f(x) + \Delta_1(x) m_k^{-1} + o(m_k^{-1}),$$

then, by application of Lemma 4.7.2, we have

$$\begin{aligned} |\delta_n^{(2)}(x)| &= O\left(\Pi_n(f(x)) \sum_{k=n_0}^n \Pi_k^{-1}(f(x)) \gamma_k m_k^{-1} w_k\right) \quad a.s. \\ &= O\left(\Pi_n(f(x)) \sum_{k=n_0}^n \Pi_k^{-1}(f(x)) \gamma_k O(\lambda_k) w_k\right) \quad a.s. \\ &= O(\lambda_n w_n) \quad a.s. \end{aligned}$$

Let us now bound  $\delta_n^{(1)}$ . To this end, we set

$$\begin{aligned} \varepsilon_k(x) &= \mathbb{E}[Z_k(x)] - Z_k(x), \\ G_k(x) &= r_k(x) - r(x), \\ Q_n(x) &= \sum_{k=n_0}^n \Pi_k^{-1}(f(x)) \gamma_k \varepsilon_k(x) G_{k-1}(x), \end{aligned}$$

and  $\mathcal{F}_k = \sigma((X_1, Y_1), \dots, (X_k, Y_k))$ . In view of

$$\text{Var}[Z_k(x)] = f(x)\psi(x)m_k^{1/2} + o(m_k^{1/2}), \quad \forall x \in (0, 1),$$

and of Lemma 4.7.2, the increasing process of the martingale  $(Q_n(x))$  satisfies

$$\begin{aligned}
\langle Q \rangle_n(x) &= \sum_{k=n_0}^n \mathbb{E} [\Pi_k^{-2}(f(x)) \gamma_k^2 \varepsilon_k^2(x) G_{k-1}^2(x) | \mathcal{F}_{k-1}] \\
&= \sum_{k=n_0}^n \Pi_k^{-2}(f(x)) \gamma_k^2 G_{k-1}^2(x) \text{Var} [Z_k(x)] \\
&= O \left( \sum_{k=n_0}^n \Pi_k^{-2}(f(x)) \gamma_k^2 w_k^2 m_k^{1/2} \right) \quad a.s. \\
&= O \left( \sum_{k=n_0}^n \Pi_k^{-2}(f(x)) \gamma_k \lambda_k^2 w_k^2 \right) \quad a.s. \\
&= O(\Pi_n^{-2}(f(x)) \lambda_n^2 w_n^2) \quad a.s.
\end{aligned}$$

— Let us first consider the case the sequence  $(n\gamma_n)$  is bounded. We then have  $(\Pi_n^{-1}(f(x))) \in \mathcal{GS}(\xi^{-1}f(x))$ , and thus  $\ln \langle Q \rangle_n(x) = O(\ln n)$  a.s. Theorem 1.3.15 in Duflo [27] then ensures that, for any  $\delta > 0$ ,

$$\begin{aligned}
|Q_n(x)| &= o \left( \langle Q \rangle_n^{\frac{1}{2}}(x) (\ln \langle Q \rangle_n(x))^{\frac{1+\delta}{2}} \right) + O(1) \quad a.s. \\
&= o \left( \Pi_n^{-1}(f(x)) \lambda_n w_n (\ln n)^{\frac{1+\delta}{2}} \right) + O(1) \quad a.s.
\end{aligned}$$

It follows that, for any  $\delta > 0$ ,

$$\begin{aligned}
|\delta_n^{(1)}(x)| &= o \left( \lambda_n w_n (\ln n)^{\frac{1+\delta}{2}} \right) + O(\Pi_n(f(x))) \quad a.s. \\
&= o \left( \lambda_n w_n (\ln n)^{\frac{1+\delta}{2}} \right) + o(\lambda_n) \quad a.s.
\end{aligned}$$

which concludes the proof of Lemma 4.7.7 in this case.

— Let us now consider the case  $\lim_{n \rightarrow \infty} (n\gamma_n) = \infty$ . In this case, for all  $\delta > 0$ , we have

$$\begin{aligned}
\ln(\Pi_n^{-2}(f(x))) &= \sum_{k=n_0}^n \ln(1 - \gamma_k f(x))^{-2} \\
&= \sum_{k=n_0}^n (2\gamma_k f(x) + o(g_k)) \\
&= O \left( \sum_{k=1}^n g_k k^\delta \right).
\end{aligned}$$

Since  $(\gamma_n n^\delta) \in \mathcal{GS}(-(\alpha - \delta))$  with  $(\alpha - \delta) < 1$ , we have

$$\lim_{n \rightarrow \infty} \frac{n (g_n n^\delta)}{\sum_{k=1}^n g_k k^\delta} = 1 - (\alpha - \delta).$$

It follows that  $\ln(\Pi_n^{-2}(f(x))) = O(n^{1+\delta}\gamma_n)$ . The sequence  $(\lambda_n w_n)$  being in  $\mathcal{GS}(-\lambda * +w*)$ , we deduce that, for all  $\delta > 0$ , we have

$$\ln(\langle Q \rangle_n(x)) = O(n^{1+\delta}\gamma_n) \quad a.s.$$

Theorem 1.3.15 in Duflo [27] then ensures that, for any  $\delta > 0$ ,

$$\begin{aligned} |Q_n(x)| &= o\left(\langle Q \rangle_n^{\frac{1}{2}}(x) (\ln \langle Q \rangle_n(x))^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s. \\ &= o\left(\Pi_n^{-1}(f(x))\lambda_n w_n (n^{1+\delta}\gamma_n)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s. \end{aligned}$$

It follows from the application of Lemma 4.7.2 that, for any  $\delta > 0$ ,

$$\begin{aligned} |\delta_n^{(1)}(x)| &= o\left(\lambda_n w_n (n^{1+\delta}\gamma_n)^{\frac{1+\delta}{2}}\right) + O(\Pi_n(f(x))) \quad a.s. \\ &= o\left(\lambda_n w_n (n^{1+\delta}\gamma_n)^{\frac{1+\delta}{2}}\right) \quad a.s. \end{aligned}$$

which concludes the proof of Lemma 4.7.7 for  $x \in (0, 1)$ .

For  $x = 0, 1$ , we use the same method with the sequence  $(\beta_n)$  defined in (4.7.12).

To use Lemma 4.7.7, we must establish an upper bound for  $r_n(x)_r(x)$ . Then, we successively prove Proposition 4.4.1, Theorem 4.4.1 and Proposition 4.4.3 in the case  $(n\gamma_n)$  is bounded and in the case  $\lim_{n \rightarrow \infty} (n\gamma_n) = \infty$ . We use the same idea to prove Proposition 4.4.2 and Theorem 4.4.2 which correspond to the case  $x = 0, 1$ .

**Upper bound of  $r_n(x) - r(x)$ .**

Since  $0 \leq 1 - \gamma_n Z_n(x) \leq 1$  for all  $n \geq n_0$ , it follows from (4.4.1) that, for  $n \geq n_0$ ,

$$\begin{aligned} |r_n(x)| &\leq |r_{n-1}(x)| + \gamma_n m_n |Y_n|, \\ &\leq |r_{n_0-1}(x)| + \left(\sup_{k \leq n} |Y_k|\right) \sum_{k=1}^n \gamma_k m_k. \end{aligned}$$

Since

$$\mathbb{P}\left(\sup_{k \leq n} |Y_k| > n^2\right) \leq n\mathbb{P}(|Y| > n^2) \leq n^{-3}\mathbb{E}(|Y|),$$

then, we have  $\sup_{k \leq n} |Y_k| \leq n^2$  a.s. Moreover, since  $(\gamma_n m_n) \in \mathcal{GS}(-\alpha + a)$  with

$1 - \alpha + a > 0$ , we note that  $\sum_{k=1}^n \gamma_k m_k = O(n\gamma_n m_n)$ . We thus deduce that

$$|r_n(x) - r(x)| = O(n^3 \gamma_n m_n) \quad a.s. \quad (4.7.35)$$

**Proof of Proposition 4.4.1, Theorem 4.4.1 and Proposition 4.4.3 in the case  $(n\gamma_n)$  is bounded**

In this case,  $\alpha = 1$ .

— Lemma 4.7.5 implies that :

$$|\rho_n(x) - r(x)| = O(\lambda_n \ln n) \quad a.s. \quad (4.7.36)$$

— Lemma 4.7.7 implies that : if  $\exists(w_n) \in \mathcal{GS}(w^*)$ ,  $w^* \geq 0$  such that

$$|r_n(x) - r(x)| = O(w_n) \text{ a.s.}, \text{ then}$$

$$|\delta_n(x)| = O(\lambda_n w_n \ln n) + o(\lambda_n) \quad a.s. \quad (4.7.37)$$

Set  $p_0 = \max p$  such that  $-\lambda^* p + 2 + a \geq 0$ , set  $j \in \{0, 1, \dots, p_0 - 1\}$  and assume that

$$|r_n(x) - r(x)| = O(\lambda_n^j (n^3 \gamma_n m_n) (\ln n)^j) \quad a.s. \quad (4.7.38)$$

Since the sequence  $(w_n) = (\lambda_n^j (n^3 \gamma_n m_n) (\ln n)^j)$  belongs to  $\mathcal{GS}(-\lambda^* j + 2 + a)$  with  $-\lambda^* j + 2 + a > 0$ , the application of (4.7.37) implies that

$$|\delta_n(x)| = O(\lambda_n^{j+1} (n^3 \gamma_n m_n) (\ln n)^{j+1}) + o(\lambda_n) \quad a.s.$$

Since  $(\lambda_n^{j+1} w_n (\ln n)^{j+1}) \in \mathcal{GS}(-\lambda^*(j+1) + 2 + a)$  with  $-\lambda^*(j+1) + 2 + a \geq 0$ , whereas  $(\lambda_n) \in \mathcal{GS}(-\lambda^*)$  with  $-\lambda^* < 0$ , it follows that

$$|\delta_n(x)| = O(\lambda_n^{j+1} w_n (\ln n)^{j+1}) \quad a.s.$$

and the application of (4.7.36) leads to

$$\begin{aligned} |r_n(x) - r(x)| &\leq |\rho_n(x) - r(x)| + |\delta_n(x)| \\ &= O(\lambda_n^{j+1} (n^3 \gamma_n m_n) (\ln n)^{j+1}) \quad a.s. \end{aligned}$$

Since (4.7.35) ensures that (4.7.38) is satisfied for  $j = 0$ , we have proved by induction that

$$|r_n(x) - r(x)| = O(\lambda_n^{p_0} (n^3 \gamma_n m_n) (\ln n)^{p_0}) \quad a.s.$$

Applying (4.7.37) with  $(w_n) = O(\lambda_n^{p_0} (n^3 \gamma_n m_n) (\ln n)^{p_0})$  and then (4.7.36), we obtain

$$|r_n(x) - r(x)| = O(\lambda_n^{p_0+1} (n^3 \gamma_n m_n) (\ln n)^{p_0+1}) + O(\lambda_n \ln n) \quad a.s.$$

Since the sequences  $(\lambda_n^{p_0+1} (n^3 \gamma_n m_n) (\ln n)^{p_0+1})$  and  $(\lambda_n \ln n)$  are in  $\mathcal{GS}(-\lambda^*(p_0 + 1) + 2 + a)$  with  $-\lambda^*(p_0 + 1) + 2 + a < 0$  and  $\mathcal{GS}(-\lambda^*)$  with  $-\lambda^* < 0$  respectively, it follows that

$$|r_n(x) - r(x)| = O((\ln n)^{-2}) \quad a.s.$$

Applying once more (4.7.37) with  $(w_n) = ((\ln n)^{-2}) \in \mathcal{GS}(0)$ , we get

$$|\delta_n(x)| = O(\lambda_n (\ln n)^{-1}) + o(\lambda_n) = o(\lambda_n) \quad a.s.$$

Proposition 4.4.1 (respectively Theorem 4.4.1 and Proposition 4.4.3) in the case  $(n\gamma_n)$  is bounded then follows from the application of Lemma 4.7.3 (respectively Lemma 4.7.4 and Lemma 4.7.6).

**Proof of Proposition 4.4.1, Theorem 4.4.1 and Proposition 4.4.3 in the**

**case**  $\lim_{n \rightarrow \infty} (n\gamma_n) = \infty$

In this case, for all  $\delta > 0$ , we have

— Lemma 4.7.5 implies that :

$$|\rho_n(x) - r(x)| = O(\tilde{\lambda}_n) \quad a.s. \quad (4.7.39)$$

— Lemma 4.7.7 implies that : if  $\exists (w_n) \in \mathcal{GS}(w^*)$ ,  $w^* \geq 0$  such that

$|r_n(x) - r(x)| = O(w_n)$  a.s., then

$$|\delta_n(x)| = O\left(\lambda_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}} w_n\right) + o(\lambda_n) \quad a.s. \quad (4.7.40)$$

Now, set  $\delta > 0$  such that  $c(\delta) = -\lambda^* + (1 + \delta)(1 + \delta - \alpha)/2 < 0$  (the existence of such a  $\delta$  being ensured by (A2)). In view of (4.7.35), the application of (4.7.40) with  $(w_n) = (n^3 \gamma_n m_n)$  ensures that

$$|\delta_n(x)| = O\left(\lambda_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}} n^3 \gamma_n m_n\right) + o(\lambda_n) \quad a.s.$$

and, in view of (4.7.39), it follows that

$$|r_n(x) - r(x)| = O(\tilde{\lambda}_n) + O\left(\lambda_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}} n^3 \gamma_n m_n\right) + o(\lambda_n) \quad a.s.$$

Set  $p \geq 1$ , and assume that

$$|r_n(x) - r(x)| = O(\tilde{\lambda}_n) + O\left(\lambda_n^p (n^{1+\delta} \gamma_n)^{p \frac{1+\delta}{2}} n^3 \gamma_n m_n\right) + o(\lambda_n) \quad a.s.$$

The application of (4.7.40) with  $(w_n) = (\tilde{\lambda}_n)$  and with

$$(w_n) = \left( \lambda_n^p (n^{1+\delta} \gamma_n)^{p \frac{1+\delta}{2}} n^3 \gamma_n m_n \right) + o(\lambda_n)$$

ensures that

$$|\delta_n(x)| = O \left( \lambda_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}} \tilde{\lambda}_n \right) + O \left( \lambda_n^{p+1} (n^{1+\delta} \gamma_n)^{(p+1) \frac{1+\delta}{2}} n^3 \gamma_n m_n \right) + o(\lambda_n) \quad a.s.$$

The sequence  $\left( \lambda_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}} \right)$  being in  $\mathcal{GS}(c(\delta))$  with  $c(\delta) < 0$ , it follows that

$$|\delta_n(x)| = o \left( \tilde{\lambda}_n \right) + O \left( \lambda_n^{p+1} (n^{1+\delta} \gamma_n)^{(p+1) \frac{1+\delta}{2}} n^3 \gamma_n m_n \right) + o(\lambda_n) \quad a.s.$$

and, in view of (4.7.39), we obtain

$$|r_n(x)| = O \left( \tilde{\lambda}_n \right) + O \left( \lambda_n^{p+1} (n^{1+\delta} \gamma_n)^{(p+1) \frac{1+\delta}{2}} n^3 \gamma_n m_n \right) + o(\lambda_n) \quad a.s. \quad (4.7.41)$$

We have thus proved by induction that, for all  $p \geq 1$ ,

$$|r_n(x)| = O \left( \tilde{\lambda}_n \right) + O \left( \lambda_n^p (n^{1+\delta} \gamma_n)^{p \frac{1+\delta}{2}} n^3 \gamma_n m_n \right) + o(\lambda_n) \quad a.s.$$

By setting  $p$  large enough, we deduce that

$$|r_n(x)| = O \left( \tilde{\lambda}_n \right)$$

Applying once more (4.7.40) with  $(w_n) = (\tilde{\lambda}_n)$ , we get

$$\begin{aligned} |\delta_n(x)| &= O \left( \lambda_n (n^{1+\delta} \gamma_n)^{\frac{1+\delta}{2}} \tilde{\lambda}_n \right) \quad a.s. \\ &= o(\lambda_n) \end{aligned}$$

Proposition 4.4.1 (respectively Theorem 4.4.1 and Proposition 4.4.3) in the case  $\lim_{n \rightarrow \infty} (n \gamma_n) = \infty$  then follows from the application of Lemma 4.7.3 (respectively Lemma 4.7.4 and Lemma 4.7.6).



Case	Density of $X$	n	$\hat{r}_n$	$\hat{r}_n^{NW}$	$r_{n,1}$	$r_{n,2}$	$r_{n,1}^{GR}$	$r_{n,2}^{GR}$
(a)	$\mathcal{B}(3, 5)$	50	0.055546	0.055112	0.104385	0.115865	<b>0.038028</b>	0.047105
		200	0.032658	0.034140	0.079668	0.081905	<b>0.022588</b>	0.028159
		500	0.026863	0.028092	0.066130	0.069522	<b>0.016560</b>	0.019712
	$0.5\mathcal{B}(2, 1) + 0.5\mathcal{B}(1, 4)$	50	0.068932	<b>0.067473</b>	0.125988	0.134351	0.108101	0.106430
		200	<b>0.038980</b>	0.042949	0.106689	0.104661	0.090068	0.085189
		500	<b>0.030297</b>	0.033478	0.099312	0.101245	0.084980	0.075519
	$\mathcal{N}_{[0,1]}(0, 1)$	50	0.064737	0.063374	0.064036	0.081830	<b>0.041471</b>	0.0486673
		200	0.038685	0.041353	0.025061	0.033815	<b>0.017258</b>	0.024455
		500	0.029203	0.031523	0.012420	0.018851	<b>0.009583</b>	0.014499
(b)	$\mathcal{B}(3, 5)$	50	0.036493	<b>0.036286</b>	0.092803	0.0956917	0.050927	0.0526410
		200	<b>0.014553</b>	0.016068	0.064965	0.063641	0.032833	0.032282
		500	<b>0.009843</b>	0.011627	0.057761	0.057449	0.027077	0.028032
	$0.5\mathcal{B}(2, 1) + 0.5\mathcal{B}(1, 4)$	50	<b>0.030405</b>	0.030367	0.085902	0.092896	0.042517	0.048069
		200	<b>0.012761</b>	0.013743	0.061782	0.063798	0.029937	0.029840
		500	<b>0.007265</b>	0.008007	0.057099	0.057393	0.024920	0.024856
	$\mathcal{N}_{[0,1]}(0, 1)$	50	0.051507	0.050845	0.060607	0.078881	<b>0.038801</b>	0.048378
		200	0.020906	0.023600	0.023876	0.033470	<b>0.017115</b>	0.023916
		500	0.013738	0.015828	0.012567	0.018276	<b>0.010042</b>	0.014245
(c)	$\mathcal{B}(3, 5)$	50	0.067239	<b>0.067031</b>	0.11308	0.107216	0.120559	0.105257
		200	<b>0.041546</b>	0.043186	0.078042	0.079022	0.086307	0.076342
		500	<b>0.036315</b>	0.037728	0.074358	0.070136	0.079413	0.064560
	$0.5\mathcal{B}(2, 1) + 0.5\mathcal{B}(1, 4)$	50	0.055296	<b>0.054980</b>	0.098793	0.105148	0.094618	0.086450
		200	<b>0.039160</b>	0.039796	0.079007	0.072025	0.068825	0.070474
		500	<b>0.033791</b>	0.034482	0.066915	0.063910	0.065890	0.052299
	$\mathcal{N}_{[0,1]}(0, 1)$	50	0.082809	0.082397	0.065922	0.085857	<b>0.055670</b>	0.059021
		200	0.051326	0.053833	0.024884	0.034103	<b>0.021069</b>	0.025526
		500	0.039724	0.042222	0.012545	0.017652	<b>0.010811</b>	0.014033

TABLE 4.1 – The average integrated squared error (*ISE*) of our non-recursive estimator  $\hat{r}_n$ , Nadaraya-Watson's estimator  $\hat{r}_n^{NW}$ ,  $r_{n,1}$  and  $r_{n,2}$  correspond to our recursive estimator with the choice  $(\gamma_n) = (n^{-1})$  and  $(\gamma_n) = (n^{-0.9})$  respectively, and  $r_{n,1}^{GR}$  and  $r_{n,2}^{GR}$  correspond to the generalized Révész's estimator with the choice  $(\gamma_n) = (n^{-1})$  and  $(\gamma_n) = (n^{-0.9})$  respectively.

FIGURE 4.1 – Qualitative comparison between the two proposed regression estimators  $r_n$  given in (4.4.1) with stepsize  $(\gamma_n) = (n^{-1})$  (solid red line) and  $\hat{r}_n$  given in (4.3.1) (dashed blue line) for 500 samples respectively of size 100 (left panel) and of size 500 (right panel) of the model  $r(x) = \cos(x)$  with  $X \sim \mathcal{N}_{[0,1]}(0, 1)$ .

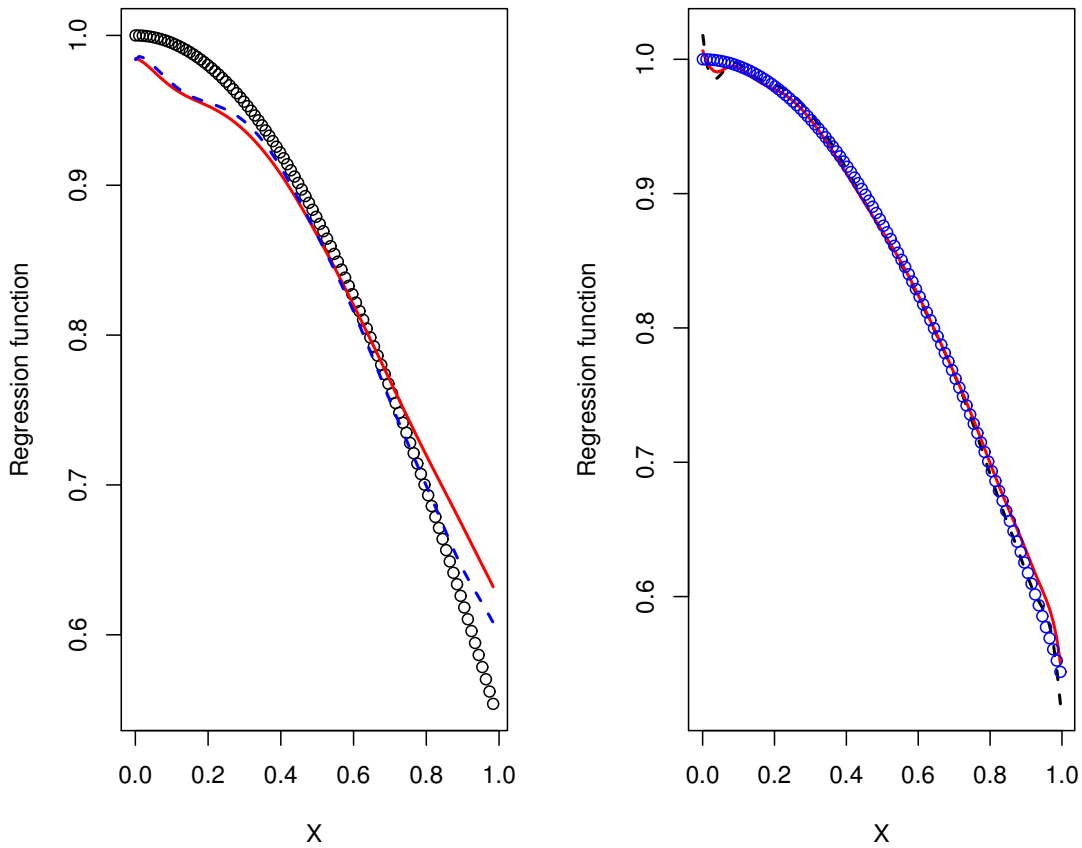
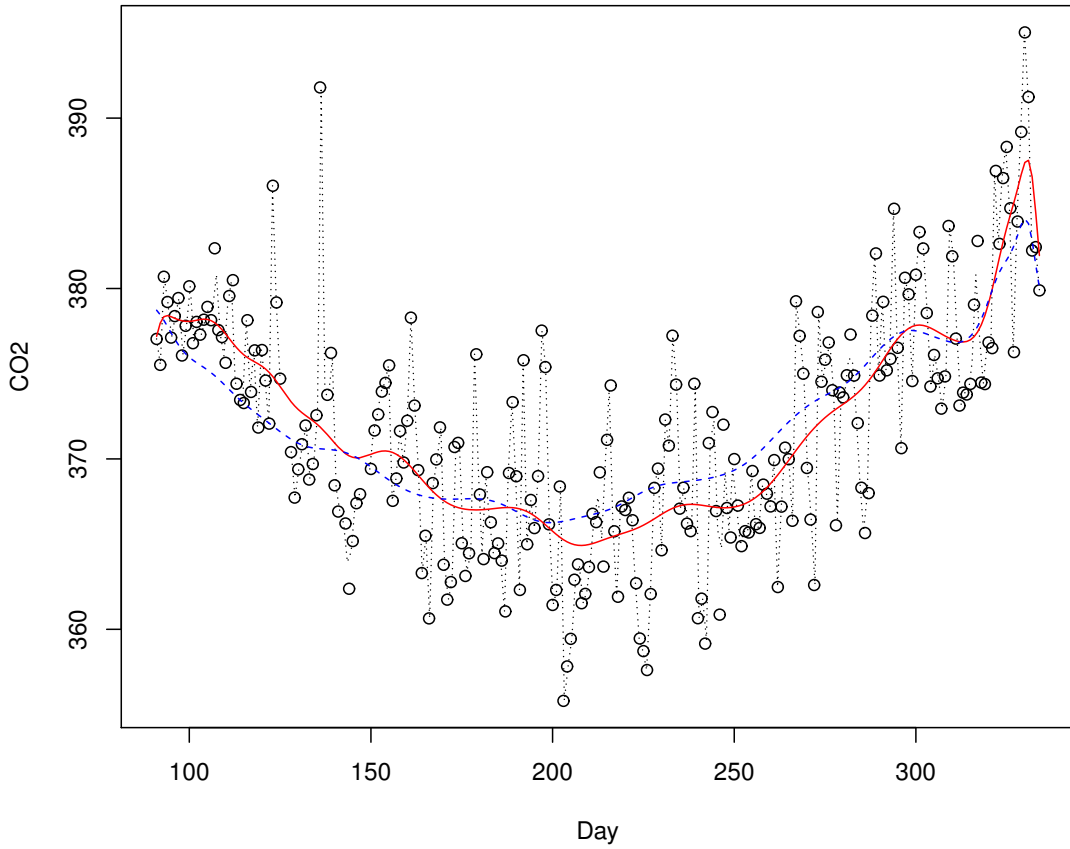


FIGURE 4.2 – Qualitative comparison between the two proposed regression estimators  $r_n$  given in (4.4.1) with stepsize  $(\gamma_n) = (n^{-1})$  (solid red line) and  $\hat{r}_n$  given in (4.3.1) (dashed blue line) for 500 samples respectively of size 100 (left panel) and of size 500 (right panel) of the model  $r(x) = \cos(x)$  with  $X \sim \mathcal{N}_{[0,1]}(0, 1)$ .



# Conclusion et perspectives

Dans ce travail nous avons apporté notre contribution afin de palier aux problèmes du bord créés par les estimateurs à noyau et ce par la construction d'estimateurs fonctionnels. Cette contribution se résume en trois grandes parties :

1. Introduire les estimateurs basés sur les polynômes de Bernstein pour une fonction de distribution, une densité de probabilité et une fonction de régression.
2. Comparer leurs propriétés théoriques avec ceux des estimateurs à noyau et les estimateurs non récursifs utilisant les polynômes de Bernstein.
3. Utiliser la méthode de validation croisée pour la recherche pratique du paramètre de lissage et appliquer nos estimateurs sur des données réelles.

Nous avons montré que nos estimateurs possèdent de bonnes qualités en particulier près du bord. L'avantage de nos estimateurs est qu'ils utilisent un algorithme simple qui facilite la mise à jours de l'estimateur lorsque on passe d'un échantillon de taille  $n$  à un échantillon de taille  $n + 1$ .

Au cours de ce travail de recherche nous avons vu que de nombreux problèmes restent encore à étudier. Parmi les points que nous pouvons ajouter à l'étude des estimateurs basés sur les polynômes de Bernstein, les plus intéressants semblent être les suivants

1. Établir la convergence presque sûre et appliquer le principe de grandes déviations et de déviations modérées pour nos estimateurs.
2. La généralisation de nos estimateurs au cas multidimensionnel.
3. La construction d'un nouvel estimateur basé sur les polynômes de Bernstein d'une fonction de régression en utilisant la méthode d'estimation semi-réursive.

# Bibliographie

1. Altman, N., Léger, C. (1995), 'Bandwidth selection for kernel distribution function estimation', *Journal of Statistical Planning and Inference*, **46**, 195–214.
2. Babu, G. J., Canty, A. J. and Chaubey, Y. P. (2002), 'Application of Bernstein polynomials for smooth estimation of a distribution and density function', *Journal of Statistical Planning and Inference*, **105**, 377–392.
3. 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', *Statistics and Probability Letters*, **76** , 959–969.
4. Berline, A. and Devroye, L. (1989), 'A comparison of kernel density estimates', *Publications de l'Institut de Statistique de l'Université de Paris*, **38**, 03–59.
5. Bernstein, S.N. (1912), 'Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités', *Communications of the Kharkov Mathematical Society*, **13** 1–2.
6. Blum, J.R. (1954), 'Multidimensional stochastic approximation methods', *Annals of Mathematical Statistics*, **25**, 737–744.
7. Bojanic, R., and Seneta, E. (1973), 'A unified theory of regularly varying sequences', *Mathematische Zeitschrift*, **134**, 91–106.
8. Bouezmarni, T. and J. Rolin (2007), 'Bernstein estimator for unbounded density function', *Journal of Nonparametric Statistics*, **19**, 145–161.
9. Bowman, A. (1984), 'An alternative method of cross-validation for the smoothing of density estimate', *Biometrika*, **71** , 352–360.

10. Bowman, A., Hall, P. and Prvan, T. (1998), 'Bandwidth selection for the smoothing of distribution functions', *Biometrika*, **85**, 799–808.
11. Chen, H. (1988), 'Lower rate of convergence for locating a maximum of a function', *Annals of Statistics*, **16**, 1330–1334.
12. Chen, S, X. (1996), 'Empirical likelihood confidence intervals for nonparametric density estimation', *Biometrika*, **83**, 329–341.
13. Chen, H.F., Duncan, T.E., and Pasik-Duncan, B. (1999), 'A Kiefer-Wolfowitz algorithm with randomized differences', *IEEE Transactions on Automatic Control*, **44**, 442–453.
14. Chen, S, X. (1999), 'Beta kernel estimators for density functions', *Computational Statistics and Data Analysis*, **31**, 131–145.
15. Chen, S, X. (2000), 'Probability Density Function Estimation Using Gamma Kernels', *Annals of the Institute of Statistical Mathematics*, **52**, 471–480.
16. Choi, E., Hall, P. and Rousson, V. (2000), 'Data sharpening methods for bias reduction in nonparametric regression', *Annals of Statistics*, **28.5**, 1339–1355.
17. Cline, D.B.H. and Hart, J.D. (1991), 'Kernel estimation of densities with discontinuities or discontinuous derivatives', *Statistics*, **1** 69–84.
18. Collomb, G. (1977), 'Estimation non-paramétrique de la régression : revue bibliographique', *International Statistical Review*, **49**, 75–93.
19. Collomb, G. (1977), 'Quelques propriétés de la méthode du noyau pour l'estimation non-paramétrique de la régression en un point fixé', *Comptes Rendus de l'Académie des Sciences. Paris*, **285 A**, 289–292.
20. Cowling, A. and Hall, P. (1996), 'On pseudodata methods for removing boundary effects in kernel density estimation', *Journal of the Royal Statistical Society. Series B*, **3**, 551–563.
21. Cuevas, A. , Cao, R. and Gonzalez-Manteiga, W. (1994), 'A comparative study of several smoothing methods in density estimation', *Computational Statistics and Data Analysis*, **17**, 235–239.

22. Davies, H.L. (1973), 'Strong consistency of a sequential estimator of a probability density function', *Bulletin of Mathematical Sciences*, **15**, 49–54.
23. Deheuvels, P. (1973), 'Sur l'estimation séquentielle de la densité', *Comptes Rendus de l'Académie des Sciences, Paris. Série A-B*, **276**, 1119–1121.
24. Devroye, L. (1979), 'On the pointwise and integral convergence of recursive kernel estimates of probability densities', *Utilitas Mathematica*, **15**, 113–128.
25. Dippon, J. and Renz J. (1997), 'Weighted means in stochastic approximation of minima', *SIAM Journal on Control and Optimization*, **35**, 1811–1827.
26. Dippon, J. (2003), 'Accelerated randomized stochastic optimization', *Annals of Statistics*, **31**, 1260–1281
27. Duflo, M. (1997), 'Random iterative models. In applications of mathematics stochastic modelling and applied probability', *Springer*.
28. Epanechnikov, V. (1969), 'Nonparametric estimates of a multivariate probability density', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **67**, 175–196.
29. Fabian, V. (1967), 'Stochastic approximation of minima with improved asymptotic speed', *Annals of Mathematical Statistics*, **38**, 191–200.
30. Galambos, J. and Seneta, E. (1973), 'Regularly varying sequences', *Proceedings of the American Mathematical Society*, **41**, 110–116.
31. Gasser, T. and Müller, H.G. (1979), 'Kernel estimation of regression functions. Smoothing techniques for curve estimation', *Proceedings of Workshop, Heidelberg, April*, **2–4**, 1979.
32. Gasser, T., Müller, H.G. and Mammitzsch, V. (1985), 'Kernels for nonparametric curve estimation', *Journal of the Royal Statistical Society*, **2**, 238–252.
33. Ghosal, S. (2000), 'Convergence rates for density estimation with Bernstein polynomials', *The Annals of Statistics*, **29**, 1264–1280.
34. Hall, P. and Heyde, C.C. (1980), 'Martingale limit theory and its application', *Academic Press, Inc., New York-London*.

35. Hall, P. (1984), 'Asymptotic properties of integrated square error and cross validation for kernel estimation of a regression function', *Journal of the American Statistical Association*, **98**, 598-608.
36. Hall, P. and Müller, H.-G. (2003), 'Order-preserving nonparametric regression with applications to conditional distribution and quantile function estimation', *Journal of the American Statistical Association*, **98**, 598-608.
37. Härdle, W. and Marron, J. S. (1985), 'Optimal bandwidth selection in nonparametric regression function estimation', *Annals of Statistics*, **13.4**, 1465–1481.
38. Härdle, W. and Kelly, G. (1987), 'Nonparametric kernel regression estimation optimal choice of bandwidth', *Statistics*, **18.1**, 21–35.
39. Härdle, W. (1990), 'Applied Nonparametric Regression. Cambridge University', *Press, Cambridge*.
40. Härdle, W. (1991), 'Smoothing Techniques with Implementation in S', *Springer, New York*.
41. Hirukawa, M. (2010), 'Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval', *Computational Statistics and Data Analysis*, **54**, 473–495 .
42. Igarashi, I. and Kakizawa, Y. (2014), 'On improving convergence rate of Bernstein polynomial density estimator', *Journal of Nonparametric Statistics*, **26**, 61–84.
43. Jmaei, A., Slaoui, Y. and Dellagi, W. (2017), 'Recursive distribution estimator defined by stochastic approximation method using Bernstein polynomials', *Journal of Nonparametric Statistics*, doi : 10.1080/10485252.2017.1369538.
44. Jones, M. C. (1990), 'The performance of kernel density functions in kernel distribution function estimation', *Statistics and Probability Letters*, **9**, 129–132.
45. Jones, M. C. (1993), 'Simple boundary correction for density estimation kernel', *Statistics and Computing*, **3**, 135–146.
46. Jones, M.C. and Foster, P.J. (1996), 'A simple nonnegative boundary correction method for kernel density estimation', *Statistics Sinica*, **4**, 1005–1013.



47. Kakizawa, Y. (2004), 'Bernstein polynomial probability density estimation', *Journal of Nonparametric Statistics*, **16** , 709–729.
48. Kakizawa, Y. (2011), 'A note on generalized bernstein polynomial density estimators', *Statistical Methodology*, **8**, 136–153.
49. Karunamuni, R.J. and Alberts,T. (2005), 'Ageneralized reflection method of boundary correction in kernel density estimation', *The Canadian Journal of Statistics*, **33**, 497–509.
50. Karunamuni, R.J., Sriram, T.N. and Wu, J. (2006), 'Asymptotic normality of an adaptiv kernel density estimator for finite mixture models', *Statistics and Probability Letters*, **76**, 211–220.
51. Karunamuni, R. J. and Zhang, S. (2008), 'Some improvements on a boundary corrected kernel density estimator', *Statistics and Probability Letters*, **5**, 499–507.
52. Kiefer, J. and Wolfowitz, J. (1952), 'Stochastic approximation of the maximum of a regression functions', *Annals of Mathematical Statistics*, **23**, 462–466.
53. Kruijer, W.T. and van der Vaart,A.W. (2008), 'Posterior convergence rates for Dirichlet mixtures of beta densities', *Journal of Statistical Planning and Inference*, **138**, 1981–1992.
54. Kushner H.J. and Clark, D.S. (1978), 'Stochastic approximation methods for constrained an unconstrained systems', *Springer, New York*.
55. Leblanc, A. (2010), 'A bias-reduced approach to density estimation using Bernstein polynomials', *Journal of Nonparametric Statistics* , **22**, 459–475.
56. Leblanc, A. (2012a), 'On estimating distribution function using Bernstein polynomials', *Annals of the Institute of Statistical Mathematics*, **64**, 919–943.
57. Leblanc, A. (2012b), 'On the boundary properties of bernstein polynomial estimators of density and distribution dunctions', *Journal of Statistical Planning and Inference*, **142**, 2762–2778.
58. Lejeune, M. and Sarda, P. (1992), 'Smooth estimators of distribution and density functions', *Computational Statistics and Data Analysis*, **14**, 457–471.

59. Lorentz, G. G. (1986), 'Bernstein polynomials (2nd ed.)', *New York : Chelsea Publishing*.
60. Mack, Y. P. and Müller, H.-G., (1987), 'Adaptive nonparametric estimation of a multivariate regression function', *Journal of Multivariate Analysis*, **23**, 169–182.
61. Marron, J.S. (1987), 'A comparison of cross-validation techniques in density estimation', *The Annals of Statistics*, **15**, 152–162.
62. Marron, J.S. and Ruppert, D. (1994), 'Transformations to reduce boundary bias in kernel density estimation', *Journal of the Royal Statistical Society*, **4**, 653–671.
63. Menon, V.V., Prasad, B. and Singh, R.S. (1984), 'Non-parametric recursive estimates of a probability density function and its derivatives', *Journal of Statistical Planning and Inference*, **9**, 73–82.
64. Mokkadem, A. and Pelletier, M. (2007), 'A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm', *The Annals of Statistics*, **35**, 1749–1772.
65. Mokkadem, A. and Pelletier, M. (2008), 'Compact law of the iterated logarithm for matrixnormalized sums of random vectors', *Theory of Probability and Its Applications*, **52**, 636–650.
66. Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009a), 'The stochastic approximation method for the estimation of a multivariate probability density', *Journal of Statistical Planning and Inference*, **139**, 2459–2478.
67. Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009b), 'Revisiting Révész's stochastic approximation method for the estimation of a regression function', *ALEA. Latin American Journal of Probability and Mathematical Statistics*, **6**, 63–114.
68. Müller, H.-G. (1991), 'Smooth optimum kernel estimators near endpoints', *Biometrika*, **78**, 521–530.
69. Müller, H.-G. and Prewitt K. A. (1993), 'Multiparameter bandwidth processes and adaptive surface smoothing', *Journal of Multivariate Analysis*, **47**, 1–21.
70. Müller, H.-G. and Wang, J.-L. (1994), 'Hazard rate estimation under random censoring with varying kernels and bandwidths', *Biometrics*, **50**, 61–76.

71. Müller, H.-G. (1997), 'Density adjusted kernel smoothers for random design nonparametric regression', *Statistics and Probability Letters*, **19**, 181–187.
72. Nadaraya, E. A. (1964), 'On estimating regression', *Theory of Probability and its Applications*, **10**, 186–190.
73. Nadaraya, E. A. (1965), 'On nonparametric estimation of density function and regression', *Theoretical Probability for Applications*, **10**, 186–190.
74. Nadaraya, E. A. (1989), 'Nonparametric Estimation of Probability Densities and Regression Curves', *Kluwer, Dordrecht*.
75. Park, B.U. and Marron, J. S. (1992), 'On the use of pilot estimators in bandwidth selection', *Journal of Nonparametric Statistics*, **3**, , 231–240.
76. Parzen, E. (1962), 'On estimation of probability density and mode', *Annals of Mathematical Statistics*, **33**, 1065–1070.
77. Petrone, S. (1999a), 'Bayesian density estimation using Bernstein polynomials', *Canadian Journal of Statistics*, **27**, 105–126.
78. Petrone, S. (1999b), 'Random Bernstein Polynomials', *Scandinavian Journal of Statistics*, **26**, 373–393.
79. Petrone, S. and Wasserman, L. (2002), 'Consistency of Bernstein polynomial posteriors', *Journal of the Royal Statistical Society, Series B*, **64**, 79–100.
82. Politis, D.N. and Romano, J.P. (1995), 'Bias-corrected nonparametric spectral estimation', *Journal of Times Series Analysis*, **16**, 67–103.
81. Polyak, B.T. and Tsybakov, A.B. (1990), 'Optimal orders of accuracy for search algorithms of stochastic optimization', *Problems of Information Transmission*, **26**, 126–133.
82. Politis, D. N. and Romano, J. P. (1995), 'Bias-corrected nonparametric spectral estimation', *Journal of Time Series Analysis*, **16**, 67–103.
83. M. B. Priestley and M. T. Chao (1972), 'Non-parametric function fitting', *Journal of the Royal Statistical Society. Series B*, **34**, 385–392.

84. Rao, B. L. S. P. (2005), 'Estimation of distribution and density functions by generalized Bernstein polynomials', *Indian Journal of Pure and Applied Mathematics*, **36**, 63–88.
85. Révész, P. (1973), 'Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I', *Studia Scientiarum Mathematicarum Hungarica*, **8**, 391–398.
86. Révész, P. (1977), 'How to apply the method of stochastic approximation in the non-parametric estimation of a regression function', *Mathematische Operationsforschung und Statistik, Series Statistics*, **8**, 119–126.
87. Robbins, H. and Monro, S. (1951), 'A Stochastic Approximation Method', *Annals of Mathematical Statistics*, **22**, 400–407.
88. Rosenblatt, M. (1956), 'Remarks on some nonparametric estimates of density functions', *Annals of Mathematical Statistics*, **27**, 832–837.
89. Roussas, G. (1992), 'Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimate', *Journal of Nonparametric Statistics*, **3**, 171–195.
90. Rudemo, M. (1982), 'Empirical choice of histograms and kernel density estimators', *Scandinavian Journal of Statistics*, **9**, 65–78.
91. Ruppert, D. (1982), 'Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise', *Annals of Probability*, **10**, 178–187.
92. Sarda, P. and Vieu, P. (1991), 'Smoothing parameter selection in hazard estimation', *Statistics and Probability Letters*, **5**, 429–434.
93. Sarda, P. (1993), 'Smoothing Parameter Selection for Smooth Distribution Functions', *Journal of Statistical Planning and Inference*, **35**, 65–75.
94. Scaillet, O. (2004), 'Density estimation using inverse and reciprocal inverse Gaussian kernels', *Journal of Nonparametric Statistics*, **16**, 217–226.
95. Schucany, W.R., Gray, H.L. and Owen, B.D. (1971), 'On bias reduction in estimation', *Journal of the American Statistical Association*, **66**, 524–533.

96. Schucany, W.R. and Sommers, j.P. (1977), 'Improvement of kernel type density estimators', *Journal of the American Statistical Association*, **72**, 420–423.
97. Schuster, E. F. (1972), 'Joint asymptotic distribution of the estimated regression function at a finite number of points', *Annals of Mathematical Statistics*, **43**, 84–88.
98. Schuster, E.F. (1985), 'Incorporating support constraints into nonparametric estimators of densities', *Communication in Statistics Theory and Methods*, **5**, 1123–1136.
99. Scott, D.W., Tapia, R.A. and Thompson, J.R. (1977), 'Kernel density estimation revisited', *Nonlinear Analysis*, **1**, 339–372.
100. Sheather, J. and Jones, M.C. (1991), 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of the Royal Statistical Society Series B*, **53**, 683–690.
101. Silverman, B.W. (1986), 'Density estimation for statistics and data analysis', *Chapman and Hall, London*.
102. Singh, R. S. (1977), 'Applications of estimators of a density and its derivatives to certain statistical problems', *Journal of the Royal Statistical Society. Series B* , **39**, 357–363.
103. Slaoui, Y. (2013), 'Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method', *Serdica Matheatical Journal*, **39**, 53–82.
104. Slaoui, Y., (2014a), 'Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method', *Journal of Probability and Statistics*, **2014**, Article ID 739640, 11 pages, doi :10.1155/2014/739640.
105. Slaoui, Y. (2014b), 'The stochastic approximation method for the estimation of a distribution function'. *Mathematical Methods of Statistics*, **23**, 306–325.
106. Slaoui, Y. (2015a), 'Moderate deviation principles for recursive regression estimators defined by stochastic approximation method', *International Journal of Mathematics and Statistics*, **16**, 51–60.

107. Slaoui, Y. (2015b), 'Large and moderate deviation principles for recursive regression estimators defined by stochastic approximation method', *Serdica Mathematical Journal*, **41**, 307–328.
108. Slaoui, Y. (2015c), 'Plug-in bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method', *Statistica Neerlandica*, **69**, 483–509.
109. Slaoui, Y. (2016), 'Optimal bandwidth selection for semi-recursive kernel regression estimators', *Statistics and its interface*, **9**, 375–388.
110. Spall, J.C. (1988), 'A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting', *Proceedings of the IEEE Conference on Decision and Control, New York*, 1544–1548.
111. Spall, J.C. (1997), 'A one-measurement form of simultaneous perturbation stochastic approximation', *Automatica A Journal of The International Federation of Automatic Control*, **33**, 109–112.
112. Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N. (1989) 'Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate ii : radical prostatectomy treated patients'. *Journal of Urology*, **16**, , 1076–1083.
113. Stone, C. J. (1984), 'An asymptotically optimal window selection rule for kernel density estimates', *The Annals of Statistics*, **12**, 1285–1297.
114. Tiago de Oliveira, J. (1963), 'Estatística de densidades : resultados assintóticos', *Revista de la Facultad de Ciencias, Universidad de Lisboa*, **A9**, 111–206.
115. Tenbusch, A., (1994), 'Two-dimensional Bernstein polynomial density estimation', *Metrika*, **41**, 233–253.
116. Terrell, G.R. and Scott, D.W. (1980), 'On Improving Convergence Rates for Non-negative Kernel Density Estimators', *The Annals of Statistics*, **8**, 1160–1163.
117. Tsybakov, A.B. (1990), 'Recurrent estimation of the mode of a multidimensional distribution', *Problems of Information Transmission*, **26**, 31–37.

118. Tsybakov, A.B. (2004), 'Introduction à l'Estimation Non-paramétrique', *Springer, Paris*.
119. Vitale, R. A. (1975), 'A bernstein polynomial approach to density function estimation', *Statistical Inference and Related Topics*, **2**, 87–99.
120. Watson, G.S. (1975), 'Smooth regression analysis', *Sankhya A. Mathematical Statistics and Probability*, **26**, 359–372.
121. Watson, G. S. and Leadbetter, M. R. (1964), 'Hazard analysis II', *Sankhya A*, **26**, 101–116.
122. Wegman, E.J. and Davies, H. I. (1979), 'Remarks on some recursive estimators of a probability density', *Annals of Statistics*, **7**, 316–327.
123. Wertz, W. (1985), 'Sequential and recursive estimators of the probability density', *Statistics*, **16**, 277–295.
124. Wolverton, C.T. and Wagner, T.J. (1969), 'Asymptotically optimal discriminant functions for pattern classification', *IEEE Transactions on Information Theory*, **15**, 258–265.
125. Yamato, H. (1971), 'Sequential estimation of a continuous probability density function and mode', *Bulletin of Mathematical Sciences*, **14**, 1–12.
126. Youndjé, E., Sarda, P. and Vieu, P. (1996a), 'Optimal smooth hazard estimates', *Test*, **2**, 379–394.
127. Youndjé, E., Sarda, P. and Vieu, P. (1996b) 'Choix de paramètres de lissage en estimation de densité conditionnelle', *Publications de l'Institut de Statistique de Paris*, **38**, 57–80.
128. Zhang, S. and Karunamuni, R.J. (2010), 'Boundary performance of the beta-kernal estimator', *Journal of Nonparametric Statistics*, **22**, 81–104.