THÈSE

présentée devant

L' UNIVERSITE DE VERSAILLES SAINT-QUENTIN

pour obtenir le titre de :

DOCTEUR DE L'UNIVERSITÉ DE VERSAILLES-SAINT-QUENTIN

Mention : Mathématiques et Applications

PAR

Yousri Slaoui

TITRE DE LA THÈSE :

Application des méthodes d'approximations stochastiques à l'estimation de la densité et de la régression

Soutenue publiquement le 18 Décembre 2006 devant le jury composé de :

M.	Bernard	BERCU
Mme.	Marie	DUFLO
M.	Abdelkader	MOKKADEM
Mme.	Mariane	PELLETIER
M.	Jean-Michel	POGGI

RÉSUMÉ

L'objectif de cette thèse est d'appliquer les méthodes d'approximations stochastiques à l'estimation de la densité et de la régression. Dans le premier chapitre, nous construisons un algorithme stochastique à pas simple qui définit toute une famille d'estimateurs récursifs à noyau d'une densité de probabilité. Nous étudions les différentes propriétés de cet algorithme. En particulier, nous identifions deux classes d'estimateurs ; la première correspond à un choix de pas qui permet d'obtenir un risque minimal, la seconde une variance minimale. Dans le deuxième chapitre, nous nous intéressons à l'estimateur proposé par Révész (1973, 1977) pour estimer une fonction de régression $r : x \mapsto \mathbb{E}[Y|X = x]$. Son estimateur r_n, construit à l'aide d'un algorithme stochastique à pas simple, a un gros inconvénient : les hypothèses sur la densité marginale de X nécessaires pour établir la vitesse de convergence de r_n sont beaucoup plus fortes que celles habituellement requises pour étudier le comportement asymptotique d'un estimateur d'une fonction de régression. Nous montrons comment l'application du principe de moyennisation des algorithmes stochastiques permet, tout d'abord en généralisant la définition de l'estimateur de Révész, puis en moyennisant cet estimateur généralisé, de construire un estimateur récursif $\bar{r}_n$ qui possède de bonnes propriétés asymptotiques. Dans le troisième chapitre, nous appliquons à nouveau les méthodes d'approximation stochastique à l'estimation d'une fonction de régression. Mais cette fois, plutôt que d'utiliser des algorithmes stochastiques à pas simple, nous montrons comment les algorithmes stochastiques à pas doubles permettent de construire toute une classe d'estimateurs récursifs d'une fonction de régression, et nous étudions les propriétés asymptotiques de ces estimateurs. Cette approche est beaucoup plus simple que celle du deuxième chapitre : les estimateurs construits à l'aide des algorithmes à pas doubles n'ont pas besoin d'être moyennisés pour avoir les bonnes propriétés asymptotiques.

ABSTRACT

The objective of this thesis is to apply the stochastic approximations methods to the estimation of a density and of a regression function. In the first chapter, we build up a stochastic algorithm with single stepsize, which defines a whole family of recursive kernel estimators of a probability density. We study the properties of this algorithm. In particular, we identify two classes of estimators ; the first one corresponds to a choice of stepsize which allows to get a minimum mean squared error, the second one a minimum variance. In the second chapter, we consider the estimator proposed by Révész (1973, 1977) to estimate a regression function $r : x \mapsto \mathbb{E}[Y|X = x]$. His estimator r_n, built up by using a single-time-scale stochastic algorithm, has a big disadvantage : the assumptions on the marginal density of X necessary to establish the convergence rate of r_n are much stronger than those usually required to study the asymptotic behavior of an estimator of a regression function. We show how the application of the averaging principle of stochastic algorithms allows, by first generalizing the definition of the estimator of Révész and then by averaging this generalized estimator, to build up a recursive estimator $\bar{r}_n$ which has good asymptotic properties. In the third chapter, we still apply stochastic approximation methods to estimate a regression function. But this time, rather than to use single-time-scale stochastic algorithm, we show how the two-time-scale stochastic algorithms allow to build up a whole class of recursive estimators of a regression function, and we study the asymptotic properties of these estimators. This approach is much easier than the one of the second chapter : the estimators built up using the two-time-scale algorithms do not need to be averaged to have good asymptotic properties.

Remerciements

Tout d'abord, mes remerciements s'adressent aux personnes qui m'ont proposé le sujet de thèse et qui m'ont encadré tout au long de ces années d'études : Mariane Pelletier et Abdelkader Mokkadem. Au travers de nos discussions, ils m'ont apporté une compréhension plus approfondie des divers aspects du sujet.

Je tiens à exprimer ma profonde gratitude à Marie Duflo, qui m'a fait le très grand honneur de faire partie du jury de thèse.

Je remercie tout particulièrement Bernard Bercu et Jean-Michel Poggi qui ont accepté d'examiner ce travail et d'en être les rapporteurs.

J'exprime toute mon amitié à Eric pour son aide et aussi à tous les autres doctorants, Clémence, Jean Maxime, Slim, Reda, Baba, Hamdi, Michael, Pascal, Sylvain et Abdoul.

Je clos enfin ces remerciements en dédiant cette thèse de doctorat à mes parents, à mon frère et à mes deux soeurs.

Table des matières

Chapitre 1

Introduction

Introduction

L'objectif de cette thèse est d'appliquer les méthodes d'approximation stochastique à l'estimation de la densité et de la régression. L'utilisation la plus célèbre des algorithmes stochastiques dans le cadre des statistiques non paramétriques est le travail de Kiefer et Wolfowitz (1952). Ces deux auteurs ont construit un algorithme qui permet l'approximation du mode d'une fonction de régression. Leur algorithme a été beaucoup discuté et leur travail prolongé dans plusieurs directions (citons, parmi beaucoup d'autres, Blum (1954), Fabian (1967), Kushner et Clark (1978), Hall et Heyde (1980), Ruppert (1982), Chen (1988), Spall (1988), Polyak et Tsybakov (1990), Dippon et Renz (1997), Spall (1997), Chen, Duncan et Pasik-Duncan (1999), Dippon (2003), et Mokkadem et Pelletier (2004)). Des algorithmes stochastiques d'approximation ont été également présentés par Révész (1973, 1977) pour estimer une fonction de régression en un point donné, et par Tsybakov (1990) pour approximer le mode d'une densité de probabilité. Ce travail est composé de trois chapitres ; dans le premier chapitre, nous nous intéressons au problème de l'estimation d'une densité de probabilité, dans les deux autres chapitres à celui de l'estimation d'une fonction de régression.

Avant de présenter nos résultats de façon détaillée, nous en donnons tout d'abord les grandes lignes.

Dans le premier chapitre, nous construisons un algorithme stochastique à pas simple qui définit toute une famille d'estimateurs récursifs à noyau d'une densité de probabilité. Nous étudions les différentes propriétés de cet algorithme. En particulier, nous identifions deux classes d'estimateurs ; la première correspond à un choix de pas qui permet d'obtenir un risque minimal, la seconde une variance minimale.

Dans le deuxième chapitre, nous nous intéressons à l'estimateur proposé par Révész (1973, 1977) pour estimer une fonction de régression $r : x \mapsto \mathbb{E}[Y|X = x]$. Son estimateur r_n, construit à l'aide d'un algorithme stochastique à pas simple, a un gros inconvénient : les hypothèses sur la densité marginale de X nécessaires pour établir la vitesse de convergence de r_n sont beaucoup plus fortes que celles habituellement requises pour étudier le comportement asymptotique d'un estimateur d'une fonction de régression. Nous montrons comment l'application du principe de moyennisation des algorithmes stochastiques permet, tout d'abord en généralisant la définition de l'estimateur de Révész, puis en moyennisant cet estimateur généralisé, de construire un estimateur récursif $\bar{r}_n$ qui possède de bonnes propriétés asymptotiques.

Dans le troisième chapitre, nous appliquons à nouveau les méthodes d'approximation stochastique à l'estimation d'une fonction de régression. Mais cette fois, plutôt que d'utiliser des algorithmes stochastiques à pas simple, nous montrons comment les algorithmes stochastiques à pas doubles permettent de construire toute une classe d'estimateurs récursifs d'une fonction de régression, et nous étudions les propriétés asymptotiques de ces estimateurs. Cette approche est beaucoup plus simple que celle du deuxième chapitre : les estimateurs construits à l'aide des algorithmes à pas doubles n'ont pas besoin d'être moyennisés pour avoir les bonnes propriétés asymptotiques.

Avant de détailler nos résultats, nous introduisons une classe de suites à variations régulières que nous utiliserons tout au long de la thèse pour définir, entre autres, les pas des algorithmes.

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

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{1.1}$$

La condition (1.1) a été introduite par Galambos et Seneta (1973) pour définir les suites à variations régulières (voir aussi Bojanic et Seneta (1973)). Des exemples typiques de suites dans $\mathcal{GS}(\gamma)$ sont, pour $h^* > 0$ et $b \in \mathbb{R}$, $(h_n) = (h^* n^\gamma)$, $(h_n) = (h^* n^\gamma [\log n]^b)$, $(h_n) = (h^* n^\gamma [\log\log n]^b)$, *etc.*

Nous donnons ici une présentation détaillée des trois chapitres qui composent cette thèse.

1.1 Algorithmes stochastiques et estimateurs récursifs de la densité

Les algorithmes stochastiques de recherche du zéro z^* d'une fonction inconnue $h : \mathbb{R} \to \mathbb{R}$ sont construits de la façon suivante : (i) on choisit $Z_0 \in \mathbb{R}$; (ii) on définit récursivement la suite (Z_n) en posant

$$Z_n = Z_{n-1} + \gamma_n W_n$$

où W_n est une observation de la fonction h au point Z_{n-1} et où le pas (γ_n) est une suite de réels positifs qui tend vers zéro.

Soit $(X_1, \ldots, X_n)$ un échantillon de la loi d'une variable aléatoire X de densité de probabilité f. Pour construire un estimateur de f en un point x par la méthode des algorithmes stochastiques, on définit un algorithme de recherche du zéro de la fonction $h : y \mapsto f(x) - y$. On procède donc de la façon suivante : (i) on se donne $f_0(x) \in \mathbb{R}$; (ii) pour tout $n \geq 1$, on pose

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

où $W_n(x)$ doit être une “observation” de la fonction h au point $f_{n-1}(x)$. Soient K un noyau (*i.e.* une fonction telle que $\int_{\mathbb{R}} K(x)dx = 1$) et (h_n) une fenêtre (*i.e.* une suite déterministe positive qui tend vers zéro) ; $f(x)$ peut être estimée par $Z_n(x) = h_n^{-1} K(h_n^{-1}[x - X_n])$, ce qui mène à poser $W_n(x) = Z_n(x) - f_{n-1}(x)$. L'algorithme que nous introduisons pour estimer récursivement la densité f au point x s'écrit alors sous la forme

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right). \tag{1.2}$$

La relation (1.2) définit toute une classe d'estimateurs récursifs à noyau d'une densité de probabilité. Notons que si l'on pose $(\gamma_n) = (n^{-1})$, alors l'estimateur f_n défini par l'algorithme (1.2) se réécrit sous la forme

$$f_n(x) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{h_k} K\left(\frac{x - X_k}{h_k}\right); \tag{1.3}$$

dans ce cas, f_n est l'estimateur récursif introduit par Wolverton et Wagner (1969). D'autre part, dans le cas où on pose $(\gamma_n) = (h_n [\sum_{k=1}^n h_k]^{-1})$, l'estimateur f_n défini par l'algorithme (1.2) se réécrit sous la forme

$$f_n(x) = \frac{1}{\sum_{k=1}^n h_k} \sum_{k=1}^{n} K\left(\frac{x - X_k}{h_k}\right); \tag{1.4}$$

f_n est alors l'estimateur récursif introduit par Deheuvels (1973) et étudié par Duflo (1997).

La question qui se pose naturellement est de savoir quel est le choix optimal du pas. Dans la partie 1.1.1 nous explicitons le biais et la variance de l'estimateur f_n. Dans la partie 1.1.2, nous déterminons le choix optimal du pas selon le point de vue de l'estimation ponctuelle, tandis que dans la partie 1.1.3 nous considérons le point de vue de l'estimation par intervalles de confiance. Dans la partie 1.1.4, nous donnons la vitesse de convergence presque sûre de l'estimateur f_n.

1.1.1 Biais et variance de l'estimateur f_n

Nous nous plaçons sous les hypothèses suivantes :

(H1) $K : \mathbb{R} \to \mathbb{R}$ est continue, bornée et vérifie $\int_{\mathbb{R}} K(z)\, dz = 1$, $\int_{\mathbb{R}} zK(z)\, dz = 0$ et $\int_{\mathbb{R}} z^2 K(z)\, dz < \infty$.

(H2) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ avec $\alpha \in \left]\frac{1}{2}, 1\right]$.
ii) $(h_n) \in \mathcal{GS}(-a)$ avec $a \in \left]0, \frac{\alpha}{2}\right]$.
iii) $\lim_{n\to\infty}(n\gamma_n) \in \left]\min\{2a, (1-a)/2\}, \infty\right]$.

(H3) f est bornée, deux fois différentiable, et $f^{(2)}$ est bornée.

L'hypothèse $(H2)\, iii)$ sur la limite de $(n\gamma_n)$ quand n tend vers l'infini est habituelle dans le cadre des algorithmes stochastiques. Elle implique en particulier que la limite de $\left([n\gamma_n]^{-1}\right)$ est finie. Nous introduisons les notations suivantes :

$$\begin{aligned} \xi &= \lim_{n\to+\infty} (n\gamma_n)^{-1}, \\ \mu_2 &= \int_{\mathbb{R}} z^2 K(z)\, dz. \end{aligned}$$

Proposition 1 ***(Biais et Variance de f_n)***
Supposons que les hypothèses $(H1)-(H3)$ sont vérifiées, et que $f^{(2)}$ est continue au point x.

1. *Si $a \leq \alpha/5$, alors*

$$\mathbb{E}(f_n(x)) - f(x) = \frac{1}{2(1-2a\xi)} h_n^2 \mu_2 f^{(2)}(x) + o\left(h_n^2\right), \tag{1.5}$$

si $a > \alpha/5$, alors

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

2. *Si $a \geq \alpha/5$, alors*

$$Var(f_n(x)) = \frac{1}{2-(1-a)\xi} \frac{\gamma_n}{h_n} f(x) \int_{\mathbb{R}} K^2(z)\, dz + o\left(\frac{\gamma_n}{h_n}\right), \tag{1.6}$$

si $a < \alpha/5$, alors

$$Var(f_n(x)) = o\left(h_n^4\right).$$

Remarque 1 *La seconde assertion dans la partie 1 de la proposition 1 est vérifiée dans le cas $a > \alpha/5$, c'est-à-dire dans le cas $h_n^4 = o\left(\gamma_n h_n^{-1}\right)$. De même, la deuxième assertion dans la partie 2 de la proposition 1 est satisfaite quand $a < \alpha/5$, c'est-à-dire quand $\gamma_n h_n^{-1} = o\left(h_n^4\right)$. Notons que dans le cas $a = \alpha/5$, (1.5) et (1.6) sont toutes les deux vérifiées.*

1.1.2 Choix optimal du pas selon le point de vue de l'estimation ponctuelle

Pour déterminer le choix optimal du pas selon le point de vue de l'estimation ponctuelle, nous cherchons le pas qui permet de minimiser le risque. Dans la partie 1.2.1 du chapitre 1, nous montrons comment le corollaire suivant se déduit de la proposition 1.

Corollaire 1
Supposons que les hypothèses $(H1)-(H3)$ *sont vérifiées, que* $f(x)>0$*, que* $f^{(2)}(x)\neq 0$ *et que* $f^{(2)}$ *est continue au point* x*. Pour minimiser le risque de* f_n*, le pas* (γ_n) *doit être choisi dans* $\mathcal{GS}(-1)$ *et tel que* $\lim_{n\to\infty} n\gamma_n = 1$ *et la fenêtre* (h_n) *doit être égale à*

$$\left(\left[\frac{3}{10}\frac{f(x)\int_{\mathbb{R}}K^2(z)\,dz}{\mu_2^2\left(f^{(2)}(x)\right)^2}\right]^{\frac{1}{5}}\gamma_n^{\frac{1}{5}}\right).$$

Dans ce cas, le risque de f_n *est égal à*

$$n^{-\frac{4}{5}}\frac{5^{\frac{11}{5}}}{4^{\frac{7}{5}}3^{\frac{6}{5}}}\left[\mu_2 f^{(2)}(x)\right]^{\frac{2}{5}}\left[f(x)\int_{\mathbb{R}}K^2(z)\,dz\right]^{\frac{4}{5}}[1+o(1)].$$

Un exemple typique de pas (γ_n) appartenant à $\mathcal{GS}(-1)$ et vérifiant $\lim_{n\to\infty} n\gamma_n = 1$ est $(\gamma_n) = \left(n^{-1}\right)$. Pour ce choix de pas, l'estimateur f_n défini par (1.2) se réécrit sous la forme compacte (1.3) ; il est alors égal à l'estimateur récursif introduit par Wolverton et Wagner (1969), puis étudié entre autres par Yamato (1971), Davies (1973), Devroye (1979), Wegman et Davies (1979), Menon, Prasad et Singh (1984), Wertz (1985), Roussas (1992) et Duflo (1997). L'estimateur de Wolverton et Wagner appartient donc à une classe bien particulière des estimateurs à noyau récursifs : celle des estimateurs dont le risque peut être rendu minimal grâce à un choix adéquat de fenêtre. Notons que des pas (γ_n) appartenant à $\mathcal{GS}(-1)$ et satisfaisant la condition $\lim_{n\to\infty} n\gamma_n = 1$ peuvent être construits en choisissant une suite $(u_n)\in\mathcal{GS}(u^*)$ avec $u^*>-1$ et en posant

$$\gamma_n = \frac{u_n}{(1+u^*)\sum_{k=1}^n u_k}.$$

Soulignons également que le risque optimal obtenu pour l'estimateur non-récursif de Rosenblatt défini par

$$\tilde{f}_n(x) = \frac{1}{nh_n}\sum_{k=1}^n K\left(\frac{x-X_k}{h_n}\right), \tag{1.7}$$

est égal à

$$n^{-\frac{4}{5}}\frac{5}{4}\left[\mu_2 f^{(2)}(x)\right]^{\frac{2}{5}}\left[f(x)\int_{\mathbb{R}}K^2(z)\,dz\right]^{\frac{4}{5}}[1+o(1)].$$

Ainsi, le risque optimal des estimateurs à noyau récursif est plus grand que le risque optimal de l'estimateur à noyau non-récursif de Rosenblatt : dans le cadre de l'estimation ponctuelle, mieux vaut utiliser l'estimateur non récursif $\tilde{f}_n$.

1.1.3 Choix optimal du pas selon le point de vue de l'estimation par intervalle de confiance

Avant de déterminer le choix optimal du pas (γ_n) pour la construction d'intervalles de confiance de f, nous donnons la vitesse de convergence faible de l'estimateur f_n défini par l'algorithme (1.2).

Théorème 1 ***(Convergence en loi de*** f_n***)***
Supposons que les hypothèses $(H1)-(H3)$ *sont vérifiées, que* $f(x)>0$ *et que* $f^{(2)}$ *est continue au point* x*.*

1. *S'il existe $c \geq 0$ tel que $\gamma_n^{-1} h_n^5 \to c$, alors*

$$\sqrt{\gamma_n^{-1} h_n} \left(f_n(x) - f(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{c^{\frac{1}{2}}}{2(1-2a\xi)} f^{(2)}(x) \mu_2, \frac{1}{(2-(1-a)\xi)} f(x) \int_{\mathbb{R}} K^2(z)\, dz \right).$$

2. *Si $\gamma_n^{-1} h_n^5 \to \infty$, alors*

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

Pour construire un intervalle de confiance de $f(x)$, il faut utiliser la première partie du théorème 1. De plus, Hall (1992) montre que pour construire des intervalles de confiance pour une densité de probabilité, mieux vaut avoir recours à un léger sous-lissage plutôt qu'à une estimation du biais. Sous-lisser signifie rendre le biais négligeable devant le terme de variance. Donc, pour choisir le pas optimal selon le point de vue de l'estimation par intervalles de confiance, le critère que nous utilisons est : minimiser la variance de f_n.

Notons que, dans notre contexte, le sous-lissage s'obtient en choisissant (γ_n) et (h_n) tels que $\gamma_n^{-1} h_n^5 \to 0$ et donc $a \geq \alpha/5$. Dans la partie 1.2.2 du chapitre 1, nous montrons comment le corollaire suivant se déduit de la proposition 1.

Corollaire 2
Supposons que les hypothèses $(H1) - (H3)$ sont vérifiées avec $a \geq \alpha/5$, que $f(x) > 0$ et que $f^{(2)}$ est continue en x. Pour minimiser la variance de f_n, α doit être choisi égale à 1, (γ_n) doit être tel que $\lim_{n\to\infty} n\gamma_n = 1 - a$, et on a alors

$$Var\left[f_n(x)\right] = \frac{1-a}{nh_n} f(x) \int_{\mathbb{R}} K^2(z)\, dz + o\left(\frac{1}{nh_n}\right).$$

De plus, dans ce cas, si la fenêtre (h_n) est choisie telle que $\lim_{n\to\infty} nh_n^5 = 0$, alors

$$\sqrt{nh_n} \left(f_n(x) - f(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, (1-a) f(x) \int_{\mathbb{R}} K^2(z)\, dz \right). \tag{1.8}$$

Un exemple typique de pas (γ_n) appartenant à $\mathcal{GS}(-1)$ et vérifiant $\lim_{n\to\infty} n\gamma_n = 1 - a$ est $(\gamma_n) = \left((1-a) n^{-1}\right)$. Par ailleurs, pour une fenêtre donnée $(h_n) \in \mathcal{GS}(-a)$, le pas défini par $(\gamma_n) = \left(h_n \left[\sum_{k=1}^n h_k\right]^{-1}\right)$ appartient à $\mathcal{GS}(-1)$ et vérifie la condition $\lim_{n\to\infty} n\gamma_n = 1 - a$. Pour ce choix de pas, l'estimateur f_n défini par (1.2) se réécrit sous la forme compacte (1.4) et est alors égal à l'estimateur récursif introduit par Duflo (1997). L'estimateur de Duflo appartient donc à une classe bien particulière des estimateurs à noyau récursifs : celle des estimateurs de variance minimale.

Rappelons que la variance de l'estimateur non-récursif de Rosenblatt $\tilde{f}_n$ (défini en (1.7)) est égale à

$$Var\left[\tilde{f}_n(x)\right] = \frac{1}{nh_n} f(x) \int_{\mathbb{R}} K^2(z)\, dz + o\left(\frac{1}{nh_n}\right)$$

et que, si la fenêtre (h_n) est choisie telle que $\lim_{n\to\infty} nh_n^5 = 0$, alors

$$\sqrt{nh_n} \left(\tilde{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, f(x) \int_{\mathbb{R}} K^2(z)\, dz \right). \tag{1.9}$$

La variance de l'estimateur de Rosenblatt est plus grande que celle de l'estimateur de Duflo (ou que celle de l'estimateur f_n défini par (1.2) avec le pas $(\gamma_n) = \left((1-a)\,n^{-1}\right)$). Pour construire des intervalles de confiance pour la densité, il est donc préférable d'utiliser un de ces deux estimateurs récursifs (et la propriété (1.8)) plutôt que l'estimateur de Rosenblatt (et le théorème de la limite centrale (1.9)). Les résultats de simulations que nous présentons dans le chapitre 1 corroborent ces résultats théoriques.

1.1.4 Vitesse de convergence presque sûre de f_n

Notons que le pas de l'algorithme (1.2) peut également être choisi tel que $\lim_{n\to\infty} n\gamma_n = \infty$. Ce choix conduit à la fois à un risque plus grand et à une variance plus grande et peut donc sembler avoir peu d'intérêt. Cependant, nous verrons dans le troisième chapitre que pour construire un algorithme à pas double pour estimer une fonction de régression, l'approximation de la densité doit se faire à l'aide de l'algorithme (1.2) avec un pas tel que $\lim_{n\to\infty} n\gamma_n = \infty$. Nous donnons maintenant la vitesse de convergence presque sûre ponctuelle de f_n, ainsi qu'une majoration de sa vitesse de convergence presque sûre uniforme ; ces deux résultats seront beaucoup utilisés dans les démonstrations du chapitre 3.

Théorème 2 ***(Vitesse de convergence presque sûre ponctuelle de f_n)***
Supposons que les hypothèses $(H1)-(H3)$ sont vérifiées, et que $f^{(2)}$ est continue au point x.

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

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

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

$$\left[\sqrt{\frac{c_1}{2}}\frac{1}{2(1-2a\xi)}f^{(2)}(x)\,\mu_2 - \sqrt{\frac{f(x)}{(2-(1-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz},\right.$$
$$\left.\sqrt{\frac{c_1}{2}}\frac{1}{2(1-2a\xi)}f^{(2)}(x)\,\mu_2 + \sqrt{\frac{f(x)}{(2-(1-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz}\right].$$

2. *Si $\gamma_n^{-1}h_n^5/(\ln s_n) \to \infty$, alors, avec probabilité un,*

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(f_n(x)-f(x)\right) = \frac{1}{2(1-2a\xi)}f^{(2)}(x)\,\mu_2.$$

Pour établir une majoration de la vitesse de convergence uniforme, nous avons besoin de rajouter l'hypothèse suivante.

$(H4)$ K est une fonction lipschitzienne.

Théorème 3 ***(Majoration de la vitesse de convergence uniforme presque sûre de f_n)***
Soit I un intervalle borné de $\mathbb{R}$. Supposons que les hypothèses $(H1)-(H4)$ sont vérifiées et que $f^{(2)}$ est uniformément continue sur I.

1. *S'il existe $c \geq 0$ tel que $\gamma_n^{-1}h_n^5/(\ln n)^2 \to c$, alors*

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

2. *Si $\gamma_n^{-1}h_n^5/(\ln n)^2 \to \infty$, alors*

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

1.2 Application du principe de moyennisation des algorithmes stochastiques à l'estimateur de Révész

Soient $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ des variables aléatoires indépendantes et de même loi, 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 (1973, 1977) a introduit les méthodes d'approximation stochastique pour construire un estimateur à noyau récursif de r. L'objectif de ce chapitre est d'appliquer le principe de moyennisation des algorithmes stochastiques pour construire un estimateur plus performant que celui de Révész. Dans la partie 1.2.1, nous rappelons la définition de l'estimateur introduit dans Révész (1973) ainsi que les principaux résultats de Révész (1977). Pour appliquer le principe de moyennisation, nous avons tout d'abord besoin de définir une version généralisée de l'estimateur de Révész et de connaitre la vitesse de convergence de cet estimateur généralisé : c'est l'objet de la partie 1.2.2. Dans la partie 1.2.3, nous définissons l'estimateur moyennisé et donnons ses propriétés asymptotiques.

1.2.1 L'estimateur de Révész

Pour construire un algorithme d'approximation de la fonction de régression r en un point x tel que $f(x)\neq 0$, Révész (1973) définit un algorithme de recherche du zéro de la fonction $h\ :\ y\mapsto a(x)-f(x)y$ où $a(x)=r(x)f(x)$. Il procède donc de la façon suivante : (i) il fixe $r_0(x)\in\mathbb{R}$; (ii) pour tout $n\geq 1$, il pose

$$r_n(x) \quad = \quad r_{n-1}(x)+\frac{1}{n}\mathcal{W}_n(x) \tag{1.10}$$

où $\mathcal{W}_n(x)$ est une "observation" de la fonction h au point $r_{n-1}(x)$. Estimant $a(x)$ par $h_n^{-1}Y_nK\left(h_n^{-1}[x-X_n]\right)$ et $f(x)$ par $h_n^{-1}K\left(h_n^{-1}[x-X_n]\right)$ où K est un noyau et (h_n) une fenêtre, il pose

$$\mathcal{W}_n(x)=h_n^{-1}Y_nK\left(\frac{x-X_n}{h_n}\right)-h_n^{-1}K\left(\frac{x-X_n}{h_n}\right)r_{n-1}(x).$$

Son algorithme (1.10) se réécrit ainsi sous la forme

$$r_n(x) \quad = \quad \left(1-\frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right)\right)r_{n-1}(x)+\frac{1}{nh_n}Y_nK\left(\frac{x-X_n}{h_n}\right). \tag{1.11}$$

Révész (1977) étudie les propriétés asymptotiques de cet estimateur lorsque la fenêtre (h_n) est choisie égale à (n^{-a}) avec $a\in\,]1/2,1[$. Sous la condition $f(x)>(1-a)/2$, il montre que

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

De plus, soit I un intervalle borné de $\mathbb{R}$. Révész (1977) établit que, sous la condition $\inf_{x\in I}f(x)>(1-a)/2$,

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

L'estimateur proposé par Révész a donc deux gros inconvénients : (i) la fenêtre étant $(h_n)=(n^{-a})$ avec $a>1/2$, la vitesse de convergence de r_n est plus petite que $n^{1/4}$, alors que la vitesse optimale des estimateurs à noyau est atteinte pour une fenêtre égale à $n^{-1/5}$ et vaut $n^{2/5}$; (ii) pour montrer la convergence ponctuelle (respectivement uniforme) de r_n, Révész (1977) a besoin de l'hypothèse $f(x) > (1-a)/2$ (respectivement $\inf_{x \in I} f(x) > (1-a)/2$), alors que les hypothèses usuelles dans le cadre de l'estimation d'une fonction de régression sont $f(x) > 0$ et $\inf_{x\in I} f(x) > 0$. Ce deuxième inconvénient est inhérent à la définition de l'estimateur r_n ; nous verrons dans la partie suivante que l'hypothèse de Révész sur la densité marginale f est en fait une hypothèse classique sur le pas de l'algorithme qui, dans la définition (1.11), est $(\gamma_n) = \left(n^{-1}\right)$. C'est pour supprimer cette hypothèse que nous introduisons le principe de moyennisation des algorithmes stochastiques.

Le principe de moyennisation des algorithmes stochastiques a été introduit indépendemment par Ruppert (1988) et Polyak (1990), puis repris entre autres par Yin (1991), Delyon et Juditsky (1992), Polyak et Juditsky (1992), Kushner et Yang (1993), Le Breton (1993), Le Breton et Novikov (1995), Dippon et Renz (1996, 1997), et Pelletier (2000). La moyennisation d'un algorithme stochastique à pas simple (γ_n) se fait en deux étapes ; (i) dans un premier temps, on “ralentit” l'algorithme, c'est-à-dire on choisit un pas (γ_n) tel que $\lim_{n\to\infty} n\gamma_n = \infty$; (ii) dans un deuxième temps, on calcule une moyenne de la suite calculée à la première étape.

Pour appliquer le principe de moyennisation à l'estimateur de Révész, nous avons donc besoin, dans un premier temps, de généraliser la définition (1.11) de r_n en autorisant d'autres pas que le pas $\left(n^{-1}\right)$.

1.2.2 L'estimateur de Révész généralisé

L'estimateur de Révész généralisé est défini par l'algorithme stochastique

$$r_n(x) = \left(1 - \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right)\right) r_{n-1}(x) + \gamma_n h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right). \quad (1.12)$$

Nous étudions son comportement asymptotique sous les hypothèses suivantes.

(H1) $K : \mathbb{R} \to \mathbb{R}$ est une fonction positive, continue, bornée telle que $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ et $\int_{\mathbb{R}} z^2 K(z)\,dz < \infty$.

(H2) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ avec $\alpha \in \left]\frac{3}{4}, 1\right]$; de plus la limite de $(n\gamma_n)^{-1}$ quand n tend vers l'infini existe.
ii) $(h_n) \in \mathcal{GS}(-a)$ avec $a \in \left]\frac{1-\alpha}{4}, \frac{\alpha}{3}\right[$.

(H3) *i)* $g(s,t)$ est deux fois continuement 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 $s = x$.
Pour $q \in [2,3]$, $s \mapsto \int_{\mathbb{R}} |t|^q g(s,t)\,dt$ est une fonction bornée.
iii) Pour $q \in \{0,1\}$, $\int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt < \infty$ et la fonction $s \mapsto \int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial s^2}(s,t)\,dt$ est bornée et continue au point $s = x$.

Posons

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

et, pour $f(x) \neq 0$,

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

Notre premier résultat donne la vitesse de convergence en loi ponctuelle de l'estimateur de Révész généralisé.

Théorème 4 ***(Vitesse de convergence en loi de r_n)***
Supposons les hypothèses $(H1)-(H3)$ vérifiées pour $x \in \mathbb{R}$ tel que $f(x) \neq 0$.

1. *S'il existe $c \geq 0$ tel que $\gamma_n^{-1} h_n^5 \to c$, et si $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2f(x))$, alors*
$$\sqrt{\gamma_n^{-1} h_n}\,(r_n(x) - r(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c} f(x) m^{(2)}(x)}{f(x) - 2a\xi}, \frac{Var[Y|X=x] f(x)}{2f(x) - (1-a)\xi} \int_{\mathbb{R}} K^2(z)\,dz\right).$$

2. *Si $\gamma_n^{-1} h_n^5 \to \infty$, et si $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x)$, alors*
$$\frac{1}{h_n^2}(r_n(x) - r(x)) \xrightarrow{\mathbb{P}} \frac{f(x) m^{(2)}(x)}{f(x) - 2a\xi}.$$

Les conditions sur le pas $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2f(x))$ dans la première partie du théorème 4 et $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x)$ dans la seconde partie du théorème 4 sont des hypothèses classiques dans le cadre des algorithmes stochastiques. Elles sont bien-sûr automatiquement satisfaites pour les pas (γ_n) tels que $\lim_{n\to\infty} n\gamma_n = \infty$; pour les pas tels que $\lim_{n\to\infty} n\gamma_n = \gamma_0 < \infty$, elles se réécrivent sous la forme $\gamma_0 > (1-a)/2f(x)$ (respectivement $\gamma_0 > 2a/f(x)$) pour la première partie (respectivement la deuxième partie). Notons que lorsque le pas utilisé est, comme dans l'algorithme (1.11) de Révész, $(\gamma_n) = (n^{-1})$, ces conditions s'expriment comme des conditions sur la densité marginale f. Soulignons également que, pour ce choix de pas, les fenêtres que nous considérons ($(h_n) \in \mathcal{GS}(-a)$ avec $a \in]0, 1/3[$) sont différentes de celles utilisés par Révész ($(h_n) = (n^{-a})$ avec $a \in]1/2, 1[$) et conduisent à de meilleures vitesses que celles de Révész ; en particulier, le choix $(h_n) = (h_0 n^{-1/5})$ qui conduit à la vitesse optimale $n^{-2/5}$ est autorisé par nos hypothèses.

Nous énonçons maintenant deux résultats qui sont nécessaires pour l'étude du comportement asymptotique de l'estimateur de Révész moyennisé : 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 5 ***(Vitesse de convergence presque sûre de r_n)***
Supposons les hypothèses $(H1)-(H3)$ vérifiées pour $x \in \mathbb{R}$ tel que $f(x) \neq 0$.

1. *S'il existe $c \geq 0$ tel que $\gamma_n^{-1} h_n^5 / \ln\left(\sum_{k=1}^n \gamma_k\right) \to c$, et si $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2f(x))$, alors, avec probabilité un, la suite*
$$\left(\sqrt{\frac{\gamma_n^{-1} h_n}{2\ln\left(\sum_{k=1}^n \gamma_k\right)}}\,(r_n(x) - r(x))\right)$$
est relativement compacte et l'ensemble de ses valeurs d'adhérence est l'intervalle
$$\left[\sqrt{\frac{c}{2}} \frac{f(x) m^{(2)}(x)}{f(x) - 2a\xi} - \sqrt{\frac{Var[Y|X=x] f(x) \int_{\mathbb{R}} K^2(z)\,dz}{2f(x) - (1-a)\xi}}, \right.$$
$$\left. \sqrt{\frac{c}{2}} \frac{f(x) m^{(2)}(x)}{f(x) - 2a\xi} + \sqrt{\frac{Var[Y|X=x] f(x) \int_{\mathbb{R}} K^2(z)\,dz}{2f(x) - (1-a)\xi}}\right]$$

2. *Si $\gamma_n^{-1} h_n^5 / \ln\left(\sum_{k=1}^n \gamma_k\right) \to \infty$, et si $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x)$, alors, avec probabilité un,*
$$\lim_{n\to\infty} \frac{1}{h_n^2}(r_n(x) - r(x)) = \frac{f(x) m^{(2)}(x)}{f(x) - 2a\xi}.$$

Pour établir une majoration de la vitesse de convergence uniforme de r_n, nous avons besoin des hypothèses suivantes.

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

Théorème 6 ***(Vitesse de convergence uniforme de r_n)***
Soit I un intervalle borné de $\mathbb{R}$ sur lequel $\varphi = \inf_{x\in I} f(x) > 0$. Supposons les hypothèses $(H1)-(H4)$ vérifiées pour tout $x \in I$.

1. *Si la suite $\left(\gamma_n^{-1} h_n^5 / [\ln n]^2\right)$ est bornée et si $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2\varphi)$, alors*

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

2. *Si $\lim_{n\to\infty}\left(\gamma_n^{-1} h_n^5 / [\ln n]^2\right) = \infty$ et si $\lim_{n\to\infty}(n\gamma_n) > 2a/\varphi$, alors*

$$\sup_{x\in I} |r_n(x) - r(x)| = O\left(h_n^2\right) \quad p.s.$$

1.2.3 L'estimateur de Révész moyennisé

Pour appliquer le principe de moyennisation à l'estimateur de Révész généralisé, nous procédons de la façon suivante : (i) nous "ralentissons" l'estimateur de Révész généralisé, c'est-à-dire nous choisissons un pas (γ_n) tel que $\lim_{n\to\infty} n\gamma_n = \infty$ dans l'algorithme (1.12) qui définit r_n ; (ii) nous calculons une moyenne des estimateurs r_k, autrement dit nous posons

$$\overline{r}_n = \frac{1}{\sum_{k=1}^n q_k} \sum_{k=1}^n q_k r_k$$

où la suite de poids (q_n) est une suite positive telle que $\sum q_n = \infty$. L'objectif de cette partie est de donner les propriétés asymptotiques de l'estimateur $\bar{r}_n$. Pour cela, nous avons besoin des hypothèses suivantes :

(H5) $\lim_{n\to\infty} n\gamma_n \left(\ln\left(\sum_{k=1}^n \gamma_k\right)\right)^{-1} = \infty$, et $a \in \left]1-\alpha, (4\alpha-3)/2\right[$.

(H6) $(q_n) \in \mathcal{GS}(-q)$ avec $q < \min\{1-2a, (1+a)/2\}$.

Théorème 7 ***(Vitesse de convergence en loi de $\bar{r}_n$)***
Supposons les hypothèses $(H1)-(H3)$, $(H5)$ et $(H6)$ vérifiées pour $x \in \mathbb{R}$ tel que $f(x) \neq 0$.

1. *S'il existe $c \geq 0$ tel que $nh_n^5 \to c$, alors*

$$\sqrt{nh_n}\left(\bar{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(c^{\frac{1}{2}} \frac{1-q}{1-q-2a} m^{(2)}(x), \frac{(1-q)^2}{1+a-2q} \frac{Var[Y|X=x]}{f(x)} \int_{\mathbb{R}} K^2(z)\,dz\right).$$

2. *Si $nh_n^5 \to \infty$, alors*

$$\frac{1}{h_n^2}\left(\bar{r}_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} \frac{1-q}{1-q-2a} m^{(2)}(x).$$

Notons que, quel que soit le choix du pas (γ_n) et de la suite de poids (q_n), la vitesse de convergence de l'estimateur moyennisé est soit $\sqrt{nh_n}$, soit h_n^{-2} ; c'est la même vitesse que celle obtenue pour l'estimateur de Révész généralisé lorsque le pas dans l'algorithme (1.12) est choisi appartenant à $\mathcal{GS}(-1)$ et tel que $\lim_{n\to\infty} n\gamma_n = \gamma_0 < \infty$, mais, cette fois, nous n'avons pas besoin de rajouter une hypothèse portant sur la densité marginale f.

Pour construire un intervalle de confiance de $r(x)$, il est recommandé (comme nous l'avons vu au chapitre 1) d'effectuer un léger sous-lissage, c'est-à-dire de choisir la fenêtre (h_n) telle que $\lim_{n\to\infty} nh_n^5 = 0$ (auquel cas $a \geq 1/5$). Il est également préférable de choisir la suite de poids (q_n) qui minimise la variance asymptotique de $\bar{r}_n$: c'est l'objet du corollaire suivant.

Corollaire 3
Supposons les hypothèses $(H1)-(H3)$, $(H5)$ *et* $(H6)$ *vérifiées pour* $x \in \mathbb{R}$ *tel que* $f(x) \neq 0$ *et avec* $a \geq 1/5$. *Supposons de plus que* $\lim_{n\to\infty} nh_n^5 = 0$. *Pour minimiser la variance asymptotique de* $\bar{r}_n$, q *doit être choisi égal à* a, *et on a alors*

$$\sqrt{nh_n}\left(\bar{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1-a)\frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz\right).$$

Rappelons que l'estimateur à noyau classique (non récursif) d'une fonction de régression introduit par Nadaraya (1964) et Watson (1964) est défini par

$$\tilde{r}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(h_n^{-1}(x-X_i)\right)}{\sum_{i=1}^n K\left(h_n^{-1}(x-X_i)\right)},$$

et que, lorsque $\lim_{n\to\infty} nh_n^5 = 0$, il vérifie le théorème de la limite centrale suivant :

$$\sqrt{nh_n}\left(\tilde{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz\right).$$

La variance asymptotique de l'estimateur de Révész moyennisé est donc plus petite que celle de l'estimateur de Nadaraya-Watson ; ainsi, pour construire des intervalles de confiance de $r(x)$, mieux vaut utiliser l'estimateur $\bar{r}_n$ plutôt que l'estimateur classique $\tilde{r}_n$.

Pour conclure cette partie, nous donnons la vitesse de convergence presque sûre ponctuelle de l'estimateur de Révész moyennisé, ainsi qu'une majoration de sa vitesse de convergence presque sûre uniforme.

Théorème 8 ***(Vitesse de convergence presque sûre de*** $\bar{r}_n$***)***
Supposons les hypothèses $(H1)-(H3)$, $(H5)$ *et* $(H6)$ *vérifiées pour* $x \in \mathbb{R}$ *tel que* $f(x) \neq 0$.

1. *S'il existe* $c_1 \geq 0$ *tel que* $nh_n^5/\ln\ln n \to c_1$, *alors, avec probabilité un, la suite*

$$\left(\sqrt{\frac{nh_n}{2\ln\ln n}}\left(\bar{r}_n(x) - r(x)\right)\right)$$

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

$$\left[c_1^{\frac{1}{2}}\frac{1-q}{1-q-2a}m^{(2)}(x) - \sqrt{\frac{(1-q)^2}{1+a-2q}\frac{Var[Y/X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz},\right.$$
$$\left. c_1^{\frac{1}{2}}\frac{1-q}{1-q-2a}m^{(2)}(x) + \sqrt{\frac{(1-q)^2}{1+a-2q}\frac{Var[Y/X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz}\right].$$

2. Si $nh_n^5/\ln\ln n \to \infty$, alors, avec probabilité un,

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(\bar{r}_n(x)-r(x)\right)=\frac{1-q}{1-q-2a}m^{(2)}(x).$$

Théorème 9 ***(Majoration de la vitesse de convergence uniforme de $\bar{r}_n$)***
Soit I un intervalle borné de $\mathbb{R}$ sur lequel $\varphi=\inf_{x\in I}f(x)>0$. Supposons les hypothèses $(H1)-(H6)$ vérifiées pour tout $x\in I$.

1. Si la suite $\left(nh_n^5/[\ln n]^2\right)$ est bornée et si $\alpha>(3a+3)/4$, alors

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

2. Si $\lim_{n\to\infty}\left(nh_n^5/[\ln n]^2\right)=\infty$ et si, dans le cas où $a\in[\alpha/5,1/5]$, $\alpha>(4a+1)/2$, alors

$$\sup_{x\in I}|\bar{r}_n(x)-r(x)|=O\left(h_n^2\right)\quad p.s.$$

1.3 Algorithmes stochastiques à pas doubles et estimateurs récursifs de la régression

Dans le troisième chapitre, nous nous intéressons à nouveau au problème de l'estimation d'une fonction de régression. Nous reprenons les notations du chapitre précédent : $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ sont des variables aléatoires indépendantes et de même loi, à valeurs dans $\mathbb{R}^2$, $g(x,y)$ est la densité du couple (X,Y), $f(x)$ la densité marginale de X, $r(x)=\mathbb{E}[Y|X=x]$ la régression de Y sur X et $a(x)=r(x)f(x)$. L'objectif de ce chapitre est d'introduire les algorithmes stochastiques à pas doubles pour construire un algorithme d'estimation de la fonction de régression r et d'étudier précisément les propriétés de cet algorithme.

Les algorithmes à pas doubles ont été introduits récemment par Borkar (1997), Konda et Borkar (1999), Baras et Borkar (2000), Bhatnagar, Fu, Marcus et Fard (2001), Bhatnagar, Fu, Marcus et Bathnagar (2001), Konda et Tsitsiklis (2003), leur vitesse de convergence a été établie par Konda et Tsitsiklis (2004) et Mokkadem et Pelletier (2006). Ce sont des algorithmes de recherche du zéro commun (θ^*,μ^*) de deux fonctions inconnues h_1 et h_2. Ils sont utilisés dans des contextes où l'approximation d'un seul des deux paramètres θ^* ou μ^* est intéressante, l'approximation du second paramètre n'étant nécessaire que pour permettre l'approximation du premier. Ils sont construits de la façon suivante : (i) on se donne θ_0 et μ_0 ; (ii) pour $n\geq 1$, on pose

$$\theta_n = \theta_{n-1}+\gamma_n W_n^{(1)} \tag{1.13}$$
$$\mu_n = \mu_{n-1}+\beta_n W_n^{(2)} \tag{1.14}$$

où $W_n^{(1)}$ et $W_n^{(2)}$ sont des observations de $h_1(\theta_{n-1},\mu_{n-1})$ et $h_2(\theta_{n-1},\mu_{n-1})$ respectivement et où les pas de l'algorithme (γ_n) et (β_n) sont deux suites de réels positifs qui tendent vers zéro à des vitesses différentes. Dans le cas où les deux algorithmes (1.13) et (1.14) sont "de type algorithme de Robbins-Monro", on sait que le premier converge en loi à la vitesse $\gamma_n^{-1/2}$ tandis que la vitesse de convergence du second est $\beta_n^{-1/2}$. Aussi, dans le cas où le paramètre d'intérêt est μ^*, on choisit les pas tels que $\lim_{n\to\infty}\beta_n\gamma_n^{-1}=0$.

Pour construire un estimateur de r en un point x par la méthode des algorithmes stochastiques doubles, on définit un algorithme de recherche du zéro commun des fonctions

$$h_1 \ : (y,z) \mapsto f(x) - y \quad \text{et} \quad h_2 \ : (y,z) \mapsto \frac{r(x)f(x)}{y} - z.$$

On procède donc de la façon suivante : (i) on se donne $f_0(x) > 0$ et $r_0(x) \in \mathbb{R}$; (ii) pour tout $n \geq 1$, on définit

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

où $W_n^{(1)}(x)$ et $W_n^{(2)}(x)$ sont des “observations” des fonctions h_1 et h_2 au point $(f_{n-1}(x), r_{n-1}(x))$. Estimant à nouveau $f(x)$ par $h_n^{-1}K\left(h_n^{-1}[x - X_n]\right)$ et $r(x)f(x)$ par $h_n^{-1}Y_nK\left(h_n^{-1}[x - X_n]\right)$ (où (h_n) est une fenêtre et K un noyau), on pose

$$\begin{aligned} W_n^{(1)}(x) &= h_n^{-1}K\left(\frac{x - X_n}{h_n}\right) - f_{n-1}(x), \\ W_n^{(2)}(x) &= h_n^{-1}Y_nK\left(\frac{x - X_n}{h_n}\right)\frac{1}{f_{n-1}(x)} - r_{n-1}(x). \end{aligned}$$

L’algorithme à pas doubles que nous proposons s’écrit donc sous la forme :

$$f_n(x) = (1-\gamma_n)f_{n-1}(x) + \gamma_n h_n^{-1}K\left(\frac{x - X_n}{h_n}\right) \tag{1.15}$$

$$r_n(x) = (1-\beta_n)r_{n-1}(x) + \beta_n h_n^{-1}Y_nK\left(\frac{x - X_n}{h_n}\right)\frac{1}{f_{n-1}(x)}. \tag{1.16}$$

Afin que l’algorithme (1.16) soit toujours bien défini, on choisit le pas de l’algorithme (1.15) tel que $\gamma_n \leq 1$ pour tout n et on utilise un noyau K positif ; puisque l’on a pris $f_0(x) > 0$, ceci assure que $f_n(x) > 0$ pour tout n. De plus, comme le paramètre d’intérêt ici est $r(x)$, on choisit les pas (γ_n) et (β_n) tels que $\lim_{n\to\infty} \beta_n\gamma_n^{-1} = 0$.

Pour étudier le comportement asymptotique de l’estimateur r_n défini par l’algorithme à pas doubles (1.15)-(1.16), nous avons besoin des hypothèses suivantes.

(H1) $K : \mathbb{R} \to \mathbb{R}$ est une fonction positive, continue, bornée telle que $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ et $\int_{\mathbb{R}} z^2K(z)\,dz < \infty$.

(H2) *i)* $(\beta_n) \in \mathcal{GS}(-\beta)$ avec $\beta \in \left]\frac{3}{4}, 1\right]$.
ii) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ avec $\alpha \in \left]\frac{3}{4}, \beta\right]$; de plus, $\gamma_n \leq 1$ pour tout n et $\lim_{n\to\infty} \beta_n^{-1}\gamma_n\left(\ln\left(\sum_{k=1}^n \gamma_k\right)\right)^{-1} = \infty$.
iii) $(h_n) \in \mathcal{GS}(-a)$ avec $a \in \left]0, \frac{\alpha}{3}\right[$.
iv) $\lim_{n\to\infty}(n\beta_n) \in \left]\min\{2a, (1-a)/2\}, \infty\right]$.

(H3) *i)* $g(s,t)$ est deux fois continuement 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 $s = x$. Pour $q \in [2,3]$, $s \mapsto \int_{\mathbb{R}} |t|^q g(s,t)\,dt$ est une fonction bornée.
iii) Pour $q \in \{0,1\}$, $\int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt < \infty$ et la fonction $s \mapsto \int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial s^2}(s,t)\,dt$ est bornée et continue au point $s = x$.

On note

$$\xi = \lim_{n\to\infty} (n\beta_n)^{-1}$$

et, pour $f(x) \neq 0$,

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

Notre premier résultat donne la vitesse de convergence en loi de r_n.

Théorème 10 ***(Convergence en loi de r_n)***
Supposons les hypothèses $(H1)-(H3)$ vérifiées pour $x \in \mathbb{R}$ tel que $f(x) \neq 0$.

1. *S'il existe $c \geq 0$ tel que $\beta_n^{-1} h_n^5 \to c$, alors*

$$\sqrt{\beta_n^{-1} h_n}\,(r_n(x) - r(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{1-2a\xi} m^{(2)}(x), \frac{Var[Y|X=x]}{(2-(1-a)\xi) f(x)} \int_{\mathbb{R}} K^2(z)\, dz \right).$$

2. *Si $\beta_n^{-1} h_n^5 \to \infty$, alors*

$$\frac{1}{h_n^2}(r_n(x) - r(x)) \xrightarrow{\mathbb{P}} \frac{1}{(1-2a\xi)} m^{(2)}(x).$$

Notons que la vitesse de convergence de l'estimateur r_n ne dépend pas du pas (γ_n) utilisé dans l'algorithme (1.15), mais uniquement de celui utilisé pour l'algorithme (1.16) ; ce phénomène est similaire à celui qui apparait dans le cadre des algorithmes stochastiques à pas doubles. Soulignons également que, quel que soit le choix de la fenêtre $(h_n) \in \mathcal{GS}(-a)$, les choix de pas qui conduisent à la meilleur vitesse de convergence de (r_n) sont les pas (β_n) appartenant à $\mathcal{GS}(-1)$ et vérifiant $\lim_{n\to\infty} n\beta_n = \beta^* \in]\min\{2a, (1-a)/2\}, \infty[$; cette vitesse est alors (comme pour l'estimateur de Nadaraya-Watson et comme pour l'estimateur de Révész moyennisé) soit $\sqrt{nh_n}$, soit h_n^{-2}. Considérons enfin le point de vue de l'estimation par intervalles de confiance ; dans ce cas, on effectue un léger sous-lissage et on recherche le choix du pas (β_n) qui permet de minimiser la variance asymptotique de r_n, ce qui fait l'objet du corollaire suivant.

Corollaire 4
Supposons les hypothèses $(H1)-(H3)$ vérifiées pour $x \in \mathbb{R}$ tel que $f(x) \neq 0$, et soit $(h_n) \in \mathcal{GS}(-a)$ telle que $\lim_{n\to\infty} nh_n^5 = 0$. Pour minimiser la variance asymptotique de r_n, le pas (β_n) doit être choisi tel que $\lim_{n\to\infty} n\beta_n = 1-a$, et on a alors

$$\sqrt{nh_n}\,(r_n(x) - r(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1-a)\frac{Var[Y|X=x]}{f(x)} \int_{\mathbb{R}} K^2(z)\, dz \right).$$

Remarquons que l'estimateur r_n défini par l'algorithme double (1.15)-(1.16) avec une fenêtre (h_n) appartenant à $\mathcal{GS}(-a)$ telle que $\lim_{n\to\infty} nh_n^5 = 0$ et un pas (β_n) appartenant à $\mathcal{GS}(-1)$ tel que $\lim_{n\to\infty} n\beta_n = 1-a$, a le même comportement asymptotique que l'estimateur de Révész moyennisé $\bar{r}_n$ défini avec une fenêtre (h_n) appartenant à $\mathcal{GS}(-a)$ telle que $\lim_{n\to\infty} nh_n^5 = 0$ et un poids $(q_n) \in \mathcal{GS}(-a)$ (voir les corollaires 3 et 4). Ces deux estimateurs ont une variance asymptotique plus petite que celle de l'estimateur non récursif de Nadaraya-Watson ; pour construire des intervalles de confiance de $r(x)$, il est donc recommandé d'utiliser l'un ou l'autre de ces estimateurs récursifs plutôt que d'avoir recours à l'estimateur non récursif de Nadaraya-Watson. Il parait difficile de comparer théoriquement plus précisément ces deux estimateurs récursif r_n et $\bar{r}_n$, mais les résultats de simulations semblent indiquer que l'estimateur r_n soit plus performant que $\bar{r}_n$.

Pour conclure l'étude des propriétés asymptotiques de l'estimateur r_n défini par l'algorithme double (1.15)-(1.16), nous donnons sa vitesse de convergence presque sûre ponctuelle, ainsi qu'une majoration de sa vitesse de convergence presque sûre uniforme.

Théorème 11 ***(Convergence presque sûre de r_n)***
Supposons les hypothèses $(H1)-(H3)$ vérifiées pour $x \in \mathbb{R}$ tel que $f(x) \neq 0$.

1. *S'il existe $c \geq 0$ tel que $\beta_n^{-1} h_n^5 / \ln\left(\sum_{k=1}^n \beta_k\right) \to c$ alors, avec probabilité un, la suite*

$$\left(\sqrt{\frac{\beta_n^{-1} h_n}{2\ln\left(\sum_{k=1}^n \beta_k\right)}}\left(r_n(x) - r(x)\right)\right)$$

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

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

2. *Si $\beta_n^{-1} h_n^5 / \ln\left(\sum_{k=1}^n \beta_k\right) \to \infty$, alors, avec probabilité un,*

$$\lim_{n\to\infty} \frac{1}{h_n^2}\left(r_n(x) - r(x)\right) = \frac{1}{1-2a\xi} m^{(2)}(x).$$

Pour établir une majoration de la vitesse de convergence uniforme de r_n, nous avons besoin des hypothèses suivantes.

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

Théorème 12 ***(Convergence uniforme de r_n)***
Soit I un intervalle borné de $\mathbb{R}$ sur lequel $\inf_{x\in I} f(x) > 0$. Supposons les hypothèses $(H1)-(H4)$ vérifiées pour tout $x \in I$.

1. *Si $\left(\beta_n^{-1} h_n^5 / \left[\ln n\right]^2\right)$ est bornée, alors, avec probabilité un,*

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

2. *Si $\lim_{n\to\infty} \beta_n^{-1} h_n^5 / \left[\ln n\right]^2 = \infty$ alors, avec probabilité un,*

$$\sup_{x\in I}\left|r_n(x) - r(x)\right| = O\left(h_n^2\right).$$

Bibliographie

[1] Baras, J.S. and Borkar, V.S. (2000), A learning algorithm for Markov decision processes with adaptive state aggregation. In *Proc. 39th IEEE Conference on Decision and Control.* IEEE, New York

[2] Bhatnagar, S., Fu, M.C., Marcus S.I. and Bathnagar, S. (2001), Two timescale algorithms for simulation optimization of hidden Markov models, *IEEE Transactions* **3**, 245-258

[3] Bhatnagar, S., Fu, M.C., Marcus S.I. and Fard, P.J. (2001), Optimal structured feedback policies for ABR flow control using two timescale SPSA, *IEEE/ACM Transactions on Networkings* **9**, 479-491

[4] Blum, J.R. (1954), Multidimensional stochastic approximation methods, *Ann. Math. Statist.,* **25**, 737-744

[5] Bojanic, R. and Seneta, E. (1973), A unified theory of regularly varying sequences, *Math.* Z., **134**, 91-106.

[6] Borkar, V.S. (1997), Stochastic approximation with two time scales, *Systems Control Lett.* **29**, 291-294

[7] Chen, H. (1988), Lower rate of convergence for locating a maximum of a function, *Ann. Statist.,* **16**, 1330-1334

[8] Chen, H.F., Duncan, T.E., and Pasik-Duncan, B. (1999), A Kiefer-Wolfowitz algorithm with randomized differences, *IEEE Trans. Automat. Control,* **44**, 442-453

[9] Davies, H.L. (1973), Strong consistency of a sequential estimator of a probability density function. *Bull. Math. Statist.* **15**, 49-54.

[10] Deheuvels, P. (1973), *Sur l'estimation séquentielle de la densité, C. R. Acad. Sci. Paris Ser. A-B* **276**, 1119-1121.

[11] Devroye, L. (1979), On the pointwise and integral convergence of recursive kernel estimates of probability densities. *Utilitas Math.* **15**, 113-128.

[12] Dippon, J. and Renz J. (1997), Weighted means in stochastic approximation of minima, *SIAM J. Control Optim.,* **35**, 1811-1827

[13] Dippon, J. (2003), Accelerated randomized stochastic optimization, *Ann. Statist.,* **31**, 1260-1281

[14] Duflo, M. (1997), *Random Iterative Models*, Collection Applications of mathematics, Springer.

[15] Fabian, V. (1967), Stochastic approximation of minima with improved asymptotic speed, *Ann. Math. Statist.,* **38**, 191-200

[16] Galambos, J. and Seneta, E. (1973), Regularly varying sequences, *Proc. Amer. Math. Soc.*, **41**, 110-116.

[17] Hall, P. and Heyde, C.C. (1980), *Martingale limit theory and its application,* Academic Press, Inc., New York-London

[18] Hall, P. (1992), Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Ann. Statist.* **20**, 675-694

[19] Kiefer, J. and Wolfowitz, J. (1952), Stochastic approximation of the maximum of a regression functions, *Ann. Math. Statist.,* **23**, 462-466

[20] Konda, V.R. and Borkar, V.S. (1999), Actor-critic like learning algorithms for Markov decision processes, *SIAM J. Control Optim.* **38**, 94-123

[21] Konda, V.R. and Tsitsiklis, J.N. (2003), On actor-critic algorithms, *SIAM J. Control Optim.* **42**, 1143-1166

[22] Konda, V.R. and Tsitsiklis, J.N. (2004), Convergence rate of linear two-time-scale stochastic approximation, *Ann. of Appl. Probab.* **14**, 796-819

[23] Kushner H.J. and Clark, D.S. (1978) *Stochastic approximation methods for constrained and unconstrained systems,* Springer, New York

[24] 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.

[25] Mokkadem, A. and Pelletier, M. (2004), A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *to appear in Ann. Statist.*

[26] Mokkadem, A. and Pelletier, M. (2006), Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms, *Ann. of Appl. Probab.* **16**, 1671-1702.

[27] Nadaraya, E.A. (1964) On estimating regression. Theor. Probab. Appl. **10**, 186-190.

[28] Polyak, B.T. and Tsybakov, A.B. (1990), Optimal orders of accuracy for search algorithms of stochastic optimization, *Problems Inform. Transmission,* **26**, 126-133.

[29] Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.*, **8**, 391-398.

[30] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119-126.

[31] 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. *J. of Nonparam. Statist.* **3**, 171-195.

[32] Ruppert, D. (1982), Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise, *Ann. of Probab.,* **10**, 178-187.

[33] Spall, J.C. (1988), A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting. In *Proc. Conference on Decision and Control,* 1544-1548. IEEE, New York.

[34] Watson, G. S. (1964), Smooth regression analysis.*Sankhya* Ser. A. **26**, 359-372.

[35] Spall, J.C. (1997), A one-measurement form of simultaneous perturbation stochastic approximation, *Automatica J. IFAC,* **33**, 109-112.

[36] Wegman, E.J. and Davies, H. I. (1979), Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7**, 316-327.

[37] Wertz, W. (1985), Sequential and recursive estimators of the probability density. *Statistics* **16**, 277-295.

[38] Wolverton, C.T. and Wagner, T.J. (1969), Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory* **15**, 258-265.

[39] Yamato, H. (1971), Sequential estimation of a continuous probability density function and mode. *Bull. Math. Satist.* **14**, 1-12.

Chapitre 2

On the application of stochastic approximation in the estimation of a probability density

2.1 Introduction

The most famous use of stochastic approximation algorithms in the framework of nonparametric statistics is the work of Kiefer and Wolfowitz (1952), who build up an algorithm which allows the approximation of the point at which a regression function reaches its maximum. Their well-known algorithm has been widely discussed and extended in many directions (see, among many others, Blum (1954), Fabian (1967), Kushner and Clark (1978), Hall and Heyde (1980), Ruppert (1982), Chen (1988), Spall (1988), Polyak and Tsybakov (1990), Dippon and Renz (1997), Spall (1997), Chen, Duncan and Pasik-Duncan (1999), Dippon (2003), and Mokkadem and Pelletier (2004)). Stochastic approximation algorithms have also been introduced by Révész (1977) to estimate a regression function at a given point, and by Tsybakov (1990) to approximate the mode of a probability density. The aim of this chapter is to use stochastic approximation algorithms to define estimators of a probability density at a given point.

Let us recall that stochastic approximation algorithms used for the search of the zero z^* of an unknown function $h\ :\mathbb{R}\to\mathbb{R}$ are built up in the following way : (i) $Z_0\in\mathbb{R}$ is arbitrarily chosen ; (ii) the sequence (Z_n) is recursively defined by setting

$$Z_n = Z_{n-1} + \gamma_n W_n$$

where W_n is an "observation" of the function h at the point Z_{n-1}, and where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero.

Let $X_1,\dots,X_n$ be independent, identically distributed random variables, and let f denote the probability density of X_1. To construct a stochastic algorithm, which approximates the function f at a given point x, we define an algorithm of search of the zero of the function $h\ : y\mapsto f(x)-y$. We thus proceed in the following way : (i) we set $f_0(x)\in\mathbb{R}$; (ii) for all $n\geq 1$, we set

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

where $W_n(x)$ is an "observation" of the function h at the point $f_{n-1}(x)$. To define $W_n(x)$, we follow the approach of Révész (1977) and of Tsybakov (1990), and introduce a kernel K (that is, a function satisfying $\int_{\mathbb{R}} K(x)dx = 1$) and a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and set $W_n(x) = h_n^{-1}K(h_n^{-1}[x-X_n]) - f_{n-1}(x)$. The stochastic approximation algorithm we introduce to recursively estimate the density f at the point x can thus be written as

$$f_n(x) = (1-\gamma_n)f_{n-1}(x) + \gamma_n h_n^{-1}K\left(\frac{x-X_n}{h_n}\right). \tag{2.1}$$

Relation (2.1) defines a broad class of recursive kernel density estimators. Let us mention that in the case when the stepsize (γ_n) is chosen equal to $\left(n^{-1}\right)$, the estimator f_n defined by (2.1) can be rewritten as

$$f_n(x) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{h_k}K\left(\frac{x-X_k}{h_k}\right); \tag{2.2}$$

f_n then equals the recursive kernel estimator introduced by Wolverton and Wagner (1969). On the other hand, in the case when the the stepsize (γ_n) is chosen equal to $\left(h_n\left[\sum_{k=1}^n h_k\right]^{-1}\right)$, the estimator f_n defined by (2.1) can be rewritten as

$$f_n(x) = \frac{1}{\sum_{k=1}^n h_k}\sum_{k=1}^{n}K\left(\frac{x-X_k}{h_k}\right); \tag{2.3}$$

f_n then equals the recursive kernel estimator introduced by Deheuvels (1973) and studied by Duflo (1997).

The question which naturally arises is to wonder what the optimal choice of stepsize is. To answer this question, we dissociate the two different points of view which are pointwise estimation on the one hand, and estimation by confidence intervals on the other hand. Moreover, we consider stepsizes which belong to the following class of regularly varying sequences.

Definition 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*

$$\lim_{n\to+\infty} n\left[1-\frac{v_{n-1}}{v_n}\right]=\gamma. \tag{2.4}$$

Condition (2.4) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1973)). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\log n)^{b}$, $n^{\gamma}(\log\log n)^{b}$.

Concerning the pointwise estimation point of view, the criteria we consider to find the optimal stepsize is minimizing the mean squared error (MSE) or the integrated mean squared error (MISE). We show that, to minimize the MSE or the MISE of the estimator f_n defined by (2.1), the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1$ (and the bandwidth must be suitably chosen as explicited in Section 2.2.1). A particular example of such a sequence (γ_n) is $(\gamma_n) = \left(n^{-1}\right)$. Thus, the recursive estimator (2.2) introduced by Wolverton and Wagner (1969) (and studied, among others by Yamato (1971), Davies (1973), Devroye (1979), Wegman and Davies (1979), Menon, Prasad and Singh (1984), Wertz (1985), Roussas (1992), and Duflo (1997)) belongs to the subclass of recursive kernel estimators which have a minimum MSE or MISE (thanks to an adequate choice of the bandwidth). Let us mention that the optimal MSE and MISE obtained for recursive kernel estimators are larger than the ones obtained for the well-known nonrecursive Rosenblatt estimator (*see* Rosenblatt (1956) and Parzen (1962)) defined as

$$\tilde{f}_n(x)=\frac{1}{nh_n}\sum_{k=1}^{n}K\left(\frac{x-X_k}{h_n}\right). \tag{2.5}$$

Let us now consider the estimation by confidence intervals point of view. Hall (1992) shows that, to minimize the coverage error of probability density confidence intervals, avoiding bias estimation by a slight undersmoothing is more efficient than explicit bias correction. In the framework of undersmoothing, minimizing the MSE (or the MISE) comes down to minimizing the variance (or the integrated variance), which is thus the criteria we consider here. Let the bandwidth (h_n) belong to $\mathcal{GS}(-a)$, $a \in [1/5, 1[$. We prove that to minimize the variance of the recursive estimator defined by (2.1), the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1-a$. A typical example of such a stepsize is $(\gamma_n) = \left((1-a)\,n^{-1}\right)$. On the other hand, we show that when (h_n) is in $\mathcal{GS}(-a)$ with $a < 1$, then the sequence $\left(h_n\left[\sum_{k=1}^{n} h_k\right]^{-1}\right)$ belongs to $\mathcal{GS}(-1)$ and satisfies the property $\lim_{n\to\infty} n\left(h_n\left[\sum_{k=1}^{n} h_k\right]^{-1}\right) = 1-a$. It follows that the recursive estimator (2.3) introduced by Deheuvels (1973) and studied by Duflo (1997) belongs to the subclass of the recursive kernel estimators whose variance is minimum. Let us underline that the optimal variance obtained for recursive estimators is smaller than the one of the nonrecursive Rosenblatt estimator (2.5).

It thus turns out that, for pointwise estimation, it is preferable to use the nonrecursive Rosenblatt estimator (2.5), or, if the online aspect is important, to choose the stepsize (γ_n) equal to $\left(n^{-1}\right)$, that is, to use the recursive estimator (2.2) introduced by Wolverton and Wagner (1969). Let us mention that, whatever the choice of the estimator is, to minimize the MSE or the MISE, the bandwidth (h_n) must be set equal to $\left(h^*\gamma_n^{1/5}\right)$ with a suitable choice of the constant h^*. On the other hand, for estimation by confidence intervals, it is advised to undersmooth by choosing a bandwidth (h_n) in $\mathcal{GS}(-a)$, $1/5 \leq a < 1/2$, and such that $\lim_{n\to\infty} nh_n^5 = 0$, and to use a recursive kernel estimator. The stepsize (γ_n) in (2.1) can be set equal to $\left(n^{-1}\right)$ (that is, the estimator (2.2) introduced by Wolverton and Wagner (1969) can be used) since this choice gives an estimator with a smaller variance than Rosenblatt estimator. However, it is better to set either $(\gamma_n) = \left((1-a)\,n^{-1}\right)$ or $(\gamma_n) = \left(h_n\left[\sum_{k=1}^n h_k\right]^{-1}\right)$, this lattest choice corresponding to the use of the estimator (2.3) introduced by Deheuvels (1973) and studied by Duflo (1997). Simulations results comparing the empirical levels obtained when confidence intervals are constructed by using Rosenblatt estimator, Wolverton and Wagner estimator, or Deheuvels estimator, are given in Section 2.4.

To conclude this introduction, let us mention that the stepsize of the stochastic approximation algorithm (2.1) may also be chosen such that $\lim_{n\to+\infty} n\gamma_n = \infty$. This choice leads to both a larger integrated mean squared error and a larger variance, and might thus seem to be of poor interest. However, we shall see in the next chapter that it is the choice, which allows to build up two-time-scale stochastic approximation algorithms to estimate a regression function.

This chapter is now organized as follows. In Section 2.2, we precisely state our main results. We first give the bias and the variance of f_n. Then, in Subsection 2.2.1, we give the integrated mean squared error of f_n, and point out the choice of the stepsize (γ_n) and of the bandwidth (h_n), which minimizes it ; in Subsection 2.2.2, we first show how minimizing the variance of f_n, and then give the weak convergence rate of f_n. In Section 2.3, we state two auxiliary results, which will be widely applied in the next chapter : the pointwise strong convergence rate of f_n, and an upper bound of the uniform strong convergence rate of f_n on any bounded interval of $\mathbb{R}$. Section 2.4 is devoted to our simulations results, and Section 2.5 to the proof of our theoretical results.

2.2 Assumptions and main results

The assumptions to which we shall refer are the following.

(A1) $K : \mathbb{R} \to \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\int_{\mathbb{R}} z^2K(z)\,dz < \infty$.

(A2) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in \left]\frac{1}{2}, 1\right]$.
ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in \left]0, \frac{\alpha}{2}\right]$.
iii) $\lim_{n\to\infty}(n\gamma_n) \in]\min\{2a,(1-a)/2\},\infty]$.

(A3) f is bounded, twice differentiable, and $f^{(2)}$ is bounded.

Assumption $(A2)$ *iii)* on the limit of $(n\gamma_n)$ as n goes to infinity is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\gamma_n]^{-1}\right)$ is finite. Throughout this chapter we will use the following notations :

$$\begin{aligned} \xi &= \lim_{n\to+\infty}(n\gamma_n)^{-1}, \qquad (2.6)\\ \mu_2 &= \int_{\mathbb{R}} z^2K(z)\,dz. \end{aligned}$$

Our first result is the following proposition, which gives the bias and the variance of f_n.

Proposition 2 (Bias and Variance of f_n)
Let Assumptions $(A1)-(A3)$ hold, and assume that $f^{(2)}$ is continuous at x.

1. If $a \leq \alpha/5$, then

$$\mathbb{E}\left(f_n\left(x\right)\right) - f\left(x\right) = \frac{1}{2\left(1-2a\xi\right)} h_n^2 \mu_2 f^{(2)}\left(x\right) + o\left(h_n^2\right). \tag{2.7}$$

If $a > \alpha/5$, then

$$\mathbb{E}\left(f_n\left(x\right)\right) - f\left(x\right) = o\left(\sqrt{\gamma_n h_n^{-1}}\right). \tag{2.8}$$

2. If $a \geq \alpha/5$, then

$$Var\left(f_n\left(x\right)\right) = \frac{1}{2-\left(1-a\right)\xi} \frac{\gamma_n}{h_n} f\left(x\right) \int_{\mathbb{R}} K^2\left(z\right) dz + o\left(\frac{\gamma_n}{h_n}\right). \tag{2.9}$$

If $a < \alpha/5$, then

$$Var\left(f_n\left(x\right)\right) = o\left(h_n^4\right). \tag{2.10}$$

Remark 1 *The second assertion in Part 1 of Proposition 2 holds in the case $a > \alpha/5$, that is, in the case $h_n^4 = o\left(\gamma_n h_n^{-1}\right)$. Similarly, the second assertion in Part 2 of Proposition 2 holds when $a < \alpha/5$, i.e. when $\gamma_n h_n^{-1} = o\left(h_n^4\right)$. Note that in the case $a = \alpha/5$,* (2.7) *and* (2.9) *hold simultaneously.*

2.2.1 Choice of the optimal stepsize according to the pointwise estimation point of view

Assume $f(x) > 0$, $f^{(2)} \neq 0$, and set

$$\begin{aligned} C_1\left(\xi\right) &= \frac{1}{4\left(1-2a\xi\right)^2} \mu_2^2 \left(f^{(2)}\left(x\right)\right)^2, \\ C_2\left(\xi\right) &= \frac{1}{2-\left(1-a\right)\xi} f\left(x\right) \int_{\mathbb{R}} K^2\left(z\right) dz. \end{aligned}$$

The application of Proposition 2 ensures that

$$MSE = \begin{cases} C_1\left(\xi\right) h_n^4 + o\left(h_n^4\right) & \text{if} \quad a < \alpha/5, \\ C_1\left(\xi\right) h_n^4 + C_2\left(\xi\right)\gamma_n h_n^{-1} + o\left(h_n^4 + \gamma_n h_n^{-1}\right) & \text{if} \quad a = \alpha/5, \\ C_2\left(\xi\right)\gamma_n h_n^{-1} + o\left(\gamma_n h_n^{-1}\right) & \text{if} \quad a > \alpha/5. \end{cases} \tag{2.11}$$

Set $\alpha \in]1/2, 1]$. If $a = \alpha/5$, $\left(C_1\left(\xi\right) h_n^4 + C_2\left(\xi\right)\gamma_n h_n^{-1}\right) \in \mathcal{GS}\left(-4\alpha/5\right)$. If $a < \alpha/5$, $\left(h_n^4\right) \in \mathcal{GS}\left(-4a\right)$ with $-4a > -4\alpha/5$, and, if $a > \alpha/5$, $\left(\gamma_n h_n^{-1}\right) \in \mathcal{GS}\left(-\alpha + a\right)$ with $-\alpha + a > -4\alpha/5$. It follows that, for a given α, to minimize the MSE of f_n, the parameter a must be chosen equal to $\alpha/5$. Moreover, in view of (2.11), the parameter α must be chosen equal to 1. In other words, to minimize the MSE of f_n, the stepsize (γ_n) must be chosen in $\mathcal{GS}\left(-1\right)$, the bandwidth (h_n) in $\mathcal{GS}\left(-1/5\right)$ (and, in view of $(A2)iii$), the condition $\lim_{n\to\infty} n\gamma_n > 2/5$ must be fulfilled). For this choice of stepsize and bandwidth, set $\mathcal{L}_n = n\gamma_n$ and $\tilde{\mathcal{L}}_n = n^{1/5}h_n$. The MSE of f_n can then be rewritten as

$$MSE = n^{-\frac{4}{5}} \left[C_1\left(\xi\right)\tilde{\mathcal{L}}_n^4 + C_2\left(\xi\right)\mathcal{L}_n\tilde{\mathcal{L}}_n^{-1}\right]\left[1 + o\left(1\right)\right].$$

Now, set $\mathcal{L}_n$. Since the function $x \mapsto C_1\left(\xi\right)x^4 + C_2\left(\xi\right)\mathcal{L}_n x^{-1}$ reaches its minimum at the point $\left(C_2\left(\xi\right)\mathcal{L}_n / \left[4C_1\left(\xi\right)\right]\right)^{1/5}$, to minimise the MSE of f_n, $\tilde{\mathcal{L}}_n$ must be chosen equal to $\left(C_2\left(\xi\right)\mathcal{L}_n / \left[4C_1\left(\xi\right)\right]\right)^{1/5}$, that is, (h_n) must equal $\left(\gamma_n C_2(\xi)/\left[4C_1(\xi)\right]\right)^{1/5}$. For such a choice, the MSE of f_n can be rewritten as

$$MSE = n^{-\frac{4}{5}}\mathcal{L}_n^{\frac{4}{5}}\frac{5}{4^{\frac{4}{5}}}\left[C_1(\xi)\right]^{\frac{1}{5}}\left[C_2(\xi)\right]^{\frac{4}{5}}\left[1+o(1)\right].$$

It follows that to minimize the MSE of f_n, the limit of $\mathcal{L}_n$ (that is, of $n\gamma_n$) must be finite (and larger than 2/5). Now, set $\gamma_0 > 2/5$ and $\mathcal{L}_n = \gamma_0\delta_n$ with $\lim_{n\to\infty}\delta_n = 1$ (so that $\lim_{n\to\infty} n\gamma_n = \gamma_0$). In this case, we have $\xi = \gamma_0^{-1}$,

$$\begin{array}{lll} C_1(\xi) = \frac{\gamma_0^2}{4\left(\gamma_0-\frac{2}{5}\right)^2}c_1 & , & c_1 = \mu_2^2\left(f^{(2)}(x)\right)^2, \\ C_2(\xi) = \frac{\gamma_0}{2\left(\gamma_0-\frac{2}{5}\right)}c_2 & , & c_2 = f(x)\int_{\mathbb{R}}K^2(z)\,dz, \end{array}$$

and the MSE of f_n can be rewritten as

$$MSE = n^{-\frac{4}{5}}\delta_n^{\frac{4}{5}}\frac{5}{4^{\frac{7}{5}}}\frac{\gamma_0^2}{\left(\gamma_0-\frac{2}{5}\right)^{\frac{6}{5}}}c_1^{\frac{1}{5}}c_2^{\frac{4}{5}}\left[1+o(1)\right].$$

The function $x\mapsto x^2/(x-2/5)^{6/5}$ reaching its minimum at the point $x=1$, to minimize the MSE of f_n, γ_0 must be chosen equal to 1. We can now state the following corollary of Proposition 2.

Corollary 1
Let Assumptions $(A1)-(A3)$ *hold, and assume that* $f(x)>0$, $f^{(2)}(x)\neq 0$, *and that* $f^{(2)}$ *is continuous at* x. *To minimize the MSE of* f_n *at the point* x, *the stepsize* (γ_n) *must be chosen in* $\mathcal{GS}(-1)$ *and such that* $\lim_{n\to\infty} n\gamma_n = 1$, *the bandwidth* (h_n) *must equal*

$$\left(\left[\frac{3}{10}\frac{f(x)\int_{\mathbb{R}}K^2(z)\,dz}{\mu_2^2\left(f^{(2)}(x)\right)^2}\right]^{\frac{1}{5}}\gamma_n^{\frac{1}{5}}\right),$$

and we then have

$$MSE = n^{-\frac{4}{5}}\frac{5^{\frac{11}{5}}}{4^{\frac{7}{5}}3^{\frac{6}{5}}}\left[\mu_2 f^{(2)}(x)\right]^{\frac{2}{5}}\left[f(x)\int_{\mathbb{R}}K^2(z)\,dz\right]^{\frac{4}{5}}\left[1+o(1)\right].$$

As mentioned in the introduction, a typical example of stepsize belonging to $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1$ is $(\gamma_n) = \left(n^{-1}\right)$. For this choice of stepsize, the estimator f_n defined by (2.1) can be rewritten as (2.2) and equals the recursive kernel estimator introduced by Wolverton and Wagner (1969). This lattest estimator thus belongs to the subclass of recursive kernel estimators whose MSE can be made minimum thanks to an adequate choice of the bandwidth. Let us underline that the optimal MSE obtained for the recursive kernel estimators is larger than the one obtained for the nonrecursive Rosenblatt kernel estimator defined in (2.5), and whose optimal MSE is known to equal

$$n^{-\frac{4}{5}}\frac{5}{4}\left[\mu_2 f^{(2)}(x)\right]^{\frac{2}{5}}\left[f(x)\int_{\mathbb{R}}K^2(z)\,dz\right]^{\frac{4}{5}}\left[1+o(1)\right].$$

Let us also mention that stepsizes belonging to $\mathcal{GS}(-1)$ and satisfying $\lim_{n\to\infty} n\gamma_n = 1$ can be computed by choosing a sequence $(u_n)\in\mathcal{GS}(u^*)$ with $u^*>-1$ and by setting

$$\gamma_n = \frac{u_n}{(1+u^*)\sum_{k=1}^n u_k}.$$

As a matter of fact, since $(u_n) \in \mathcal{GS}(u^*)$ with $u^* > -1$, we have

$$\lim_{n\to\infty} \frac{nu_n}{\sum_{k=1}^n u_k} = 1 + u^*, \tag{2.12}$$

which guarantees that $\lim_{n\to\infty} n\gamma_n = 1$. Moreover, applying (2.12), we note that

$$\begin{aligned}\frac{\sum_{k=1}^{n-1} u_k}{\sum_{k=1}^n u_k} &= 1 - \frac{u_n}{\sum_{k=1}^n u_k} \\ &= 1 - \frac{1+u^*}{n} + o\left(\frac{1}{n}\right),\end{aligned}$$

so that

$$\lim_{n\to\infty} n\left[1 - \frac{\sum_{k=1}^{n-1} u_k}{\sum_{k=1}^n u_k}\right] = 1 + u^*.$$

It follows that $\left(\sum_{k=1}^n u_k\right) \in \mathcal{GS}(1+u^*)$, and thus that $(\gamma_n) \in \mathcal{GS}(-1)$.

The following proposition gives the MISE of the estimator f_n.

Proposition 3
Let Assumptions $(A1) - (A3)$ hold, and assume that $f^{(2)}$ is continuous and integrable.

1. *If $a < \alpha/5$, then*

$$MISE = \frac{1}{4(1-2a\xi)^2} h_n^4 \mu_2^2 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx + o\left(h_n^4\right).$$

2. *If $a = \alpha/5$, then*

$$MISE = \frac{1}{4(1-2a\xi)^2} h_n^4 \mu_2^2 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx + \frac{1}{2-(1-a)\xi} \frac{\gamma_n}{h_n} \int_{\mathbb{R}} K^2(z)\, dz + o\left(h_n^4 + \frac{\gamma_n}{h_n}\right).$$

3. *If $a > \alpha/5$, then*

$$MISE = \frac{1}{2-(1-a)\xi} \frac{\gamma_n}{h_n} \int_{\mathbb{R}} K^2(z)\, dz + o\left(\frac{\gamma_n}{h_n}\right).$$

Following the same lines as for the proof of Corollary 1, we deduce the following corollary from Proposition 3.

Corollary 2
Let Assumptions $(A1) - (A3)$ hold, and assume that $f^{(2)}$ is continuous and integrable. To minimize the MISE of f_n, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1$, the bandwidth (h_n) must equal

$$\left(\left[\frac{3}{10} \frac{\int_{\mathbb{R}} K^2(z)\, dz}{\mu_2^2 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx}\right]^{\frac{1}{5}} \gamma_n^{\frac{1}{5}}\right),$$

and we then have

$$MISE = n^{-\frac{4}{5}} \frac{5^{\frac{11}{5}}}{4^{\frac{7}{5}} 3^{\frac{6}{5}}} \mu_2^{\frac{2}{5}} \left[\int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx\right]^{\frac{1}{5}} \left[\int_{\mathbb{R}} K^2(z)\, dz\right]^{\frac{4}{5}} [1 + o(1)].$$

2.2.2 Choice of the optimal stepsize according to the estimation by confidence intervals point of view

Let us first state the following theorem, which gives the weak convergence rate of the estimator f_n defined in (2.1), and which is proved in Section 2.5.4.

Theorem 1 (Weak convergence rate)
Let Assumptions $(A1)-(A3)$ hold, and assume that $f(x)>0$ and that $f^{(2)}$ is continuous at x.

1. *If there exists $c\geq 0$ such that $\gamma_n^{-1}h_n^5\to c$, then*

$$\sqrt{\gamma_n^{-1}h_n}\left(f_n\left(x\right)-f\left(x\right)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(\frac{c^{\frac{1}{2}}}{2\left(1-2a\xi\right)}f^{(2)}\left(x\right)\mu_2,\frac{1}{\left(2-\left(1-a\right)\xi\right)}f\left(x\right)\int_{\mathbb{R}}K^2\left(z\right)dz\right).$$

2. *If $\gamma_n^{-1}h_n^5\to\infty$, then*

$$\frac{1}{h_n^2}\left(f_n\left(x\right)-f\left(x\right)\right)\stackrel{\mathbb{P}}{\to}\frac{1}{2\left(1-2a\xi\right)}f^{(2)}\left(x\right)\mu_2,$$

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

As mentioned in the introduction, Hall (1992) shows that, to minimize the coveraged error of probability density confidence intervals, avoiding bias estimation by a slight undersmoothing is more efficient than bias correction. In the framework of undersmoothing, the bias is negligible in front of the variance, so that the quantity to minimize is no more the MSE, but the variance.

When undersmoothing, we have $\lim_{n\to\infty}\gamma_n^{-1}h_n^5=0$, and thus $a\geq\alpha/5$. In this case, the variance of f_n is given by (2.9) in Proposition 2, and we have

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

To minimize the variance of f_n, the stepsize (γ_n) must thus belong to $\mathcal{GS}\left(-1\right)$ and satisfy $\lim_{n\to\infty}n\gamma_n=\gamma_0\in\left]\left(1-a\right)/2,+\infty\right[$. For such a choice, $\xi=\gamma_0^{-1}$, so that

$$Var\left(f_n\left(x\right)\right)=\frac{\gamma_0}{2-\left(1-a\right)\gamma_0^{-1}}\frac{1}{nh_n}f\left(x\right)\int_{\mathbb{R}}K^2\left(z\right)dz+o\left(\frac{1}{nh_n}\right).$$

The function $\gamma_0\mapsto\gamma_0\left[2-\left(1-a\right)\gamma_0^{-1}\right]^{-1}$ reaching its minimum at the point $\gamma_0=1-a$, we can state the following corollary.

Corollary 3
Let Assumptions $(A1)-(A3)$ hold with $a\geq\alpha/5$, and assume that $f(x)>0$ and that $f^{(2)}$ is continuous at x. To minimize the variance of f_n, α must be chosen equal to 1, (γ_n) must satisfy $\lim_{n\to\infty}n\gamma_n=1-a$, and we then have

$$Var\left[f_n\left(x\right)\right]=\frac{1-a}{nh_n}f\left(x\right)\int_{\mathbb{R}}K^2\left(z\right)dz+o\left(\frac{1}{nh_n}\right).$$

Moreover, in this case, if (h_n) satisfies $\lim_{n\to\infty}nh_n^5=0$, then

$$\sqrt{nh_n}\left(f_n\left(x\right)-f\left(x\right)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(0,(1-a)f\left(x\right)\int_{\mathbb{R}}K^2\left(z\right)dz\right). \tag{2.13}$$

A typical example of stepsize belonging to $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1-a$ is $(\gamma_n) = \left((1-a)\,n^{-1}\right)$. On the other hand, for a given bandwidth $(h_n) \in \mathcal{GS}(-a)$, set $\gamma_n = \left(h_n \left[\sum_{k=1}^n h_k\right]^{-1}\right)$, and note that (2.1) then gives

$$\begin{aligned}
\left[\sum_{k=1}^n h_k\right] f_n(x) &= \left(\left[\sum_{k=1}^n h_k\right] - h_n\right) f_{n-1}(x) + K\left(\frac{x-X_n}{h_n}\right) \\
&= \left[\sum_{k=1}^{n-1} h_k\right] f_{n-1}(x) + K\left(\frac{x-X_n}{h_n}\right) \\
&= \sum_{k=1}^{n} K\left(\frac{x-X_k}{h_k}\right).
\end{aligned}$$

For this choice of stepsize, the estimator f_n defined by the stochastic approximation algorithm (2.1) can thus be rewritten as (2.3) ; in other words, it then equals the recursive kernel estimator introduced by Deheuvels (1973) and studied by Duflo (1997). Moreover, in view of (2.12), this stepsize satisfies the property $\lim_{n\to\infty} n\gamma_n = 1-a$, and, in view of the development below (2.12), it belongs to $\mathcal{GS}(-1)$. The estimator introduced in Deheuvels (1973) thus belongs to the subclass of recursive estimators whose variance is minimum.

Let us recall that the variance of the nonrecursive Rosenblatt estimator $\tilde{f}_n$ defined in (2.5) equals

$$Var\left[\tilde{f}_n(x)\right] = \frac{1}{nh_n} f(x) \int_{\mathbb{R}} K^2(z)\,dz + o\left(\frac{1}{nh_n}\right),$$

and that, if (h_n) is chosen such that $\lim_{n\to\infty} nh_n^5 = 0$, then

$$\sqrt{nh_n}\left(\tilde{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, f(x)\int_{\mathbb{R}} K^2(z)\,dz\right). \tag{2.14}$$

The variance of Rosenblatt estimator is greater than the one of Deheuvels estimator (or than the one of the estimator f_n defined in (2.1) with the stepsize $(\gamma_n) = ([1-a]n^{-1})$. To construct confidence intervals for the density, it is thus advised to use one of these two recursive estimators (and Property (2.13)) rather than Rosenblatt estimator (and the central limit theorem (2.14)). The simulations results we present in Section 2.4 corroborate our theoritical results.

2.3 Auxiliary results

The aim of this section is to state two results, which will be widely applied in the third chapter. The first one gives the exact pointwise strong convergence rate of $f_n - f$, the second one gives an upper bound of the uniform strong convergence rate of $f_n - f$ on any bounded interval of $\mathbb{R}$. They are proved in Subsections 2.5.5 and 2.5.6 respectively.

Theorem 2 (Strong pointwise convergence rate)
Let Assumptions $(A1)-(A3)$ hold, and assume that $f^{(2)}$ is continuous at x.

1. *If there exists $c_1 \geq 0$ such that $\gamma_n^{-1}h_n^5/\left(\ln \sum_{k=1}^n \gamma_k\right) \to c_1$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{\gamma_n^{-1}h_n}{2\ln\sum_{k=1}^n \gamma_k}}\left(f_n(x) - f(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\frac{1}{2(1-2a\xi)}\sqrt{\frac{c_1}{2}}f^{(2)}(x)\mu_2-\sqrt{\frac{f(x)}{(2-(1-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz},\right.$$
$$\left.\frac{1}{2(1-2a\xi)}\sqrt{\frac{c_1}{2}}f^{(2)}(x)\mu_2+\sqrt{\frac{f(x)}{(2-(1-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz}\right].$$

2. *If* $\gamma_n^{-1}h_n^5/\left(\ln\sum_{k=1}^n\gamma_k\right)\to\infty$*, then, with probability one,*

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(f_n(x)-f(x)\right)=\frac{1}{2(1-2a\xi)}f^{(2)}(x)\mu_2.$$

To establish a uniform strong upper bound of the convergence rate of f_n, we need the following additional assumption.

$(A4)$ K is Lipschitz-continuous.

Theorem 3 (Strong uniform convergence rate)
Assume Assumptions $(A1)-(A4)$ *hold, let* I *be a bounded interval of* $\mathbb{R}$*, and assume that* $f^{(2)}$ *is uniformly continuous on* I*.*

1. *If there exists* $c\geq 0$ *such that* $\gamma_n^{-1}h_n^5/(\ln n)^2\to c$*, then*

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

2. *If* $\gamma_n^{-1}h_n^5/(\ln n)^2\to\infty$*, then*

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

2.4 Simulations

Set

$$I_n=\left[g_n(x)-1.96\sqrt{\frac{g_n(x)\int_{\mathbb{R}}K^2(z)\,dz}{nh_n}},\,g_n(x)+1.96\sqrt{\frac{g_n(x)\int_{\mathbb{R}}K^2(z)\,dz}{nh_n}}\right].$$

In the case $g_n(x)=\tilde{f}_n(x)$ is the Rosenblatt estimator defined in (2.5), I_n is a confidence interval with asymptotic level $2\Phi(1.96)-1=0.95$, where Φ is the distribution function of the standard normal. In the case $g_n(x)$ is the recursive estimator $f_n(x)$ defined by the algorithm (2.1) with the stepsize $(\gamma_n)=\left(n^{-1}\right)$ (that is, $g_n(x)$ is the Wolverton-Wagner estimator) the asymptotic level of I_n is $2\Phi\left(1.96\sqrt{1+a}\right)-1>0.95$. In the case $g_n(x)$ is the recursive estimator $f_n(x)$ defined by the algorithm (2.1) with the stepsize $(\gamma_n)=\left((1-a)n^{-1}\right)$, the asymptotic level of I_n is $2\Phi\left(1.96/\sqrt{1-a}\right)-1>2\Phi\left(1.96\sqrt{1+a}\right)-1$. To compare the performance of these three estimators, we consider :

- three choices of bandwidth (h_n) :
  - Choice A : $h_n=n^{-1/5}/\log(n)$ (and thus $a=1/5$),
  - Choice B : $h_n=n^{-1/4}$ (and thus $a=1/4$),
  - Choice C : $h_n=n^{-1/3}$ (and thus $a=1/3$),
- three sample sizes : $n=50$, $n=100$, and $n=200$,

- four densities f : standard normal, normal mixture, standard Cauchy, and Student with 6 degrees of freedom,
- three points : $x = 0$, $x = 0.5$, and $x = 1$.

In each case the number of simulations is $N = 5000$; the tables give the empirical levels $\# \{f(x) \in I_n\} / N$. For each choice of bandwidth A, B and C, the first line corresponds to the Rosenblatt estimator, the second one to the recursive estimator with $(\gamma_n) = \left(n^{-1}\right)$, and the third one to the recursive estimator with $(\gamma_n) = \left((1-a)\, n^{-1}\right)$. The last column CL gives the theoritical levels (the theoritical level is always 95% for the Rosenblatt estimator, but depends on the parameter a for the recursive estimators).

The empirical results confirm our theoritical results : the recursive estimator defined by the algorithm (2.1) with the choice of stepsize $(\gamma_n) = \left((1-a)\, n^{-1}\right)$ performs better than the two other estimators. Let us underline that, among the choices of bandwidth we have considered, the choice A gives the smallest width of the confidence interval I_n ; we can see that it leads to good empirical levels.

Estimation of normal standard distribution

x=0			x=0.5			x=1			CL
$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	
				A					
94.38%	94.68%	95.08%	93.94%	94%	94.48%	92.14%	92.84%	94.18%	95%
97.34%	97.56%	98.14%	97.38%	97.52%	97.8%	97%	97.36%	96.96%	96.76%
98.66%	98.48%	98.66%	98.56%	98.6%	98.7%	98.76%	98.3%	98.3%	97.14%
				B					
96.68%	96.64%	96.64%	97.14%	96.76%	96.68%	97.38%	97.06%	96.18%	95%
97.14%	96.62%	96.78%	98%	97.52%	97.78%	98.7%	98.36%	98.34%	97.14%
98.5%	97.7%	97.62%	99.02%	98.5%	98.38%	99.56%	99.36%	99.1%	97.62%
				C					
97.04%	96.76%	96.6%	96.6%	96.26%	96.16%	96.78%	96.78%	97.06%	95%
97.86%	98.12%	98.10%	98.08%	98.02%	98.24%	97.86%	98.08%	97.98%	97.62%
99.62%	99.32%	99.10%	99.64%	99.3%	99.22%	99.58%	99.18%	99.06%	98.36%

Estimation of normal mixture $\mathbf{1/2}\mathcal{N}\left(\mathbf{-1/2, 1}\right) + \mathbf{1/2}\mathcal{N}\left(\mathbf{1/2, 1}\right)$

x=0			x=0.5			x=1			CL
$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	
				A					
93.8%	94.58%	94.56%	92.86%	93.48%	94%	92.14%	92.96%	93.8%	95%
97.32%	97.42%	97.72%	96.78%	97.16%	97.7%	96.6%	97.36%	97.18%	96.76%
98.46%	98.4%	98.64%	98.44%	98.1%	98.46%	98.44%	98.44%	98.16%	97.14%
				B					
96.9%	97.22%	97.42%	96.9%	97.08%	97.04%	96.48%	96.44%	97.44%	95%
97.58%	97.78%	97.96%	98.04%	98.26%	98%	98.14%	98.06%	98.62%	97.14%
98.98%	98.78%	98.52%	99.26%	99.1%	98.9%	99.36%	98.98%	99.36%	97.62%
				C					
96.18%	96.66%	96.4%	96.22%	96.3%	96.2%	95.58%	95.76%	95.78%	95%
98.02%	98.26%	98.44%	98.2%	98.22%	98.16%	97.9%	98.18%	97.96%	97.62%
99.62%	99.42%	99.36%	99.8%	99.54%	99.44%	99.64%	99.66%	99.38%	98.36%

Estimation of standard Cauchy

x=0			x=0.5			x=1			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				A					
92.94%	94.3%	94.3%	92.5%	93.46%	94.28%	89.68%	91.48%	92.56%	95%
95.68%	97.4%	97.4%	96.5%	97.06%	97.56%	95.82%	96.42%	96.76%	96.76%
97.22%	98.22%	98.26%	98.18%	98.36%	98.44%	98.3%	98.26%	98.02%	97.14%
				B					
92.8%	93.1%	93.82%	95.84%	96.16%	96.18%	96.76%	96.78%	96.74%	95%
91.74%	91.14%	91.9%	96.9%	97.28%	97.4%	98.88%	98.88%	98.56%	97.14%
94.98%	93.58%	93.02%	98.78%	98.5%	98.5%	99.48%	99.36%	98.9%	97.62%
				C					
94.4%	94.52%	95.4%	94.94%	95.6%	95.96%	94.88%	95.5%	95.74%	95%
95.28%	95.64%	96.64%	97.1%	97.76%	98.04%	98.16%	98.5%	98.34%	97.62%
98.44%	98.22%	98.12%	99.64%	99.44%	99.24%	99.52%	99.62%	99.38%	98.36%

Estimation of Student with 6 degrees of freedom

x=0			x=0.5			x=1			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				A					
93.92%	94.88%	94.5%	93.92%	93.66%	94.68%	91.36%	93.04%	93.88%	95%
96.38%	97.48%	97.24%	96.62%	96.66%	97.46%	94.64%	95.9%	96.64%	96.76%
97.34%	98.06%	97.98%	97.5%	97.3%	98.14%	96.18%	96.7%	97.54%	97.14%
				B					
96.04%	96.24%	95.9%	96.34%	96.74%	96.78%	97.32%	97.12%	97.14%	95%
96.18%	96.26%	96.24%	97.2%	97.68%	97.64%	99%	98.86%	98.68%	97.14%
98.1%	97.5%	97.04%	99.08%	98.78%	98.64%	99.7%	99.6%	99.36%	97.62%
				C					
95.88%	97.36%	96.12%	96.02%	96.4%	96.06%	95.72%	96.16%	95.58%	95%
97.48%	98.44%	98.04%	97.96%	98.3%	97.06%	98.42%	98.4%	98.44%	97.62%
99.52%	99.48%	99.06%	99.72%	99.58%	99.12%	99.84%	99.42%	99.46%	98.36%

2.5 Proof of the results

Throughout this section we use the following notations :

$$\begin{aligned}
\Pi_n &= \prod_{j=1}^{n} \left(1-\gamma_j\right), \\
s_n &= \sum_{k=1}^{n} \gamma_k, \\
Z_n\left(x\right) &= \frac{1}{h_n} K\left(\frac{x-X_n}{h_n}\right). \qquad (2.15)
\end{aligned}$$

Let us first state the following technical lemma.

Lemma 1
Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$ and $m > 0$, such that $m - v^*\xi > 0$ where ξ is defined in (2.6), then

$$\lim_{n\to+\infty} v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\xi}. \tag{2.16}$$

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

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + \delta \right] = 0. \tag{2.17}$$

Lemma 1 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption $(A2)iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity. Let us mention that, in particular, to prove (2.9), Lemma 1 is applied with $m = 2$ and $(v_n) = (\gamma_n^{-1} h_n)$ (and thus $v^* = \alpha - a$) ; the stepsize (γ_n) must thus fulfill the condition $\lim_{n\to\infty} (n\gamma_n) > (\alpha - a)/2$. Now, since $\lim_{n\to\infty} (n\gamma_n) < \infty$ if and only if $\alpha = 1$, the condition $\lim_{n\to\infty} (n\gamma_n) \in] \min\{2a, (1-a)/2\}, \infty]$ in $(A2)iii)$ is equivalent to the condition $\lim_{n\to\infty} (n\gamma_n) \in] \min\{2a, (\alpha-a)/2\}, \infty]$, which appears throughout our proofs. Similarly, since $\xi \neq 0$ if and only if $\alpha = 1$, the equality $2 - (\alpha - a)\xi = 2 - (1 - a)\xi$ holds in all cases. Lemma 1 for such m and (v_n) equals the factor $[2 - (1 - a)\xi]^{-1}$ that stands in the statement of our main results.

Our proofs are now organized as follows. Lemma 1 is proved in Section 2.5.1, Propositions 2 and 3 in Sections 2.5.2 and 2.5.3 respectively, Theorems 1, 2, and 3 in Sections 2.5.4, 2.5.5, and 2.5.6 respectively.

2.5.1 Proof of Lemma 1

We first prove (2.17). Set

$$Q_n = v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \gamma_k v_k^{-1} \alpha_k + \delta \right].$$

We have

$$Q_n = \frac{v_n}{v_{n-1}} (1 - \gamma_n)^m Q_{n-1} + \gamma_n \alpha_n$$

with, since $(v_n) \in \mathcal{GS}(v^*)$ and in view of (2.6),

$$\begin{aligned}
\frac{v_n}{v_{n-1}} (1 - \gamma_n)^m &= \left(1 + \frac{v^*}{n} + o\left(\frac{1}{n}\right)\right) (1 - m\gamma_n + o(\gamma_n)) \\
&= (1 + v^*\xi\gamma_n + o(\gamma_n)) (1 - m\gamma_n + o(\gamma_n)) \\
&= 1 - (m - v^*\xi)\gamma_n + o(\gamma_n). \qquad (2.18)
\end{aligned}$$

Set $A \in]0, m - v^*\xi[$; for n large enough, we obtain

$$Q_n \leq (1 - A\gamma_n) Q_{n-1} + \gamma_n \alpha_n$$

and (2.17) follows straightforwardly from the application of Lemma 4.I.1 in Duflo (1996). Now, let C denote a positive generic constant that may vary from line to line ; we have

$$v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \gamma_k v_k^{-1} - (m - v^*\xi)^{-1} = v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \gamma_k v_k^{-1} - (m - v^*\xi)^{-1} P_n \right]$$

with, in view of (2.18),

$$\begin{aligned} P_n &= v_n^{-1}\Pi_n^{-m} \\ &= \sum_{k=2}^{n}\left(v_k^{-1}\Pi_k^{-m} - v_{k-1}^{-1}\Pi_{k-1}^{-m}\right) + C \\ &= \sum_{k=2}^{n} v_k^{-1}\Pi_k^{-m}\left[1 - \frac{v_k}{v_{k-1}}\left(1-\gamma_k\right)^m\right] + C \\ &= \sum_{k=2}^{n} v_k^{-1}\Pi_k^{-m}\left[\left(m - v^*\xi\right)\gamma_k + o\left(\gamma_k\right)\right] + C. \end{aligned}$$

It follows that

$$v_n\Pi_n^m\sum_{k=1}^{n}\Pi_k^{-m}\gamma_k v_k^{-1} - \left(m - v^*\xi\right)^{-1} = v_n\Pi_n^m\left[\sum_{k=1}^{n}\Pi_k^{-m}v_k^{-1}o\left(\gamma_k\right) + C\right],$$

and (2.16) follows from the application of (2.17).

2.5.2 Proof of Proposition 2

In view of (2.1) and (2.15), we have

$$\begin{aligned} & f_n(x) - f(x) \\ &= \left(1-\gamma_n\right)\left(f_{n-1}(x) - f(x)\right) + \gamma_n\left(Z_n(x) - f(x)\right) \\ &= \sum_{k=1}^{n-1}\left[\prod_{j=k+1}^{n}\left(1-\gamma_j\right)\right]\gamma_k\left(Z_k(x) - f(x)\right) + \gamma_n\left(Z_n(x) - f(x)\right) + \left[\prod_{j=1}^{n}\left(1-\gamma_j\right)\right]\left(f_0(x) - f(x)\right) \\ &= \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\left(Z_k(x) - f(x)\right) + \Pi_n\left(f_0(x) - f(x)\right). \end{aligned} \tag{2.19}$$

It follows that

$$\mathbb{E}\left(f_n(x)\right) - f(x) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\left(\mathbb{E}\left(Z_k(x)\right) - f(x)\right) + \Pi_n\left(f_0(x) - f(x)\right).$$

Taylor with remainder integral ensures that

$$\begin{aligned} \mathbb{E}\left[Z_k(x)\right] - f(x) &= \int_{\mathbb{R}} K(z)\left[f\left(x - zh_k\right) - f(x)\right]dz \\ &= \frac{1}{2}\mu_2 f^{(2)}(x)\,h_k^2 + h_k^2\delta_k(x) \end{aligned} \tag{2.20}$$

with

$$\delta_k(x) = \int_{\mathbb{R}}\int_0^1(1-s)\,z^2K(z)\left(f^{(2)}\left(x - zh_k s\right) - f^{(2)}(x)\right)dsdz,$$

and, since $f^{(2)}$ is bounded and continuous at x, we have $\lim_{k\to\infty}\delta_k(x) = 0$. In the case $a \leq \alpha/5$, we have $\lim_{n\to\infty}\left(n\gamma_n\right) > 2a$; the application of Lemma 1 then gives

$$\begin{aligned} \mathbb{E}\left[f_n(x)\right] - f(x) &= \frac{1}{2}\mu_2 f^{(2)}(x)\,\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k h_k^2[1+o(1)] + \Pi_n\left(f_0(x) - f(x)\right) \\ &= \frac{1}{2(1-2a\xi)}\mu_2 f^{(2)}(x)\left[h_n^2 + o(1)\right], \end{aligned}$$

and (2.7) follows. In the case $a > \alpha/5$, we have $h_n^2 = o\left(\sqrt{\gamma_n h_n^{-1}}\right)$; since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$, Lemma 1 then ensures that

$$\begin{aligned}\mathbb{E}\left[f_n(x)\right] - f(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k o\left(\sqrt{\gamma_k h_k^{-1}}\right) + O(\Pi_n) \\ &= o\left(\sqrt{\gamma_n h_n^{-1}}\right),\end{aligned}$$

which gives (2.8). Now, we have

$$\begin{aligned}Var\left[f_n(x)\right] &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var\left[Z_k(x)\right] \\ &= \Pi_n^2 \sum_{k=1}^{n} \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[\int_{\mathbb{R}} K^2(z) f(x - zh_k)\,dz - h_k\left(\int_{\mathbb{R}} K(z) f(x - zh_k)\,dz\right)^2\right] \\ &= \Pi_n^2 \sum_{k=1}^{n} \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[f(x)\int_{\mathbb{R}} K^2(z)\,dz + \nu_k(x) - h_k\tilde{\nu}_k(x)\right]\end{aligned}$$

with

$$\begin{aligned}\nu_k(x) &= \int_{\mathbb{R}} K^2(z)\left[f(x - zh_k) - f(x)\right]dz, \\ \tilde{\nu}_k(x) &= \left(\int_{\mathbb{R}} K(z) f(x - zh_k)\,dz\right)^2.\end{aligned}$$

Since f is bounded and continuous, we have $\lim_{k\to\infty}\nu_k(x) = 0$ and $\lim_{k\to\infty} h_k\tilde{\nu}_k(x) = 0$. In the case $a \geq \alpha/5$, we have $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$, and the application of Lemma 1 gives

$$\begin{aligned}Var\left[f_n(x)\right] &= \Pi_n^2 \sum_{k=1}^{n} \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[f(x)\int_{\mathbb{R}} K^2(z)\,dz + o(1)\right] \\ &= \frac{1}{(2 - (\alpha - a)\xi)}\frac{\gamma_n}{h_n}\left[f(x)\int_{\mathbb{R}} K^2(z)\,dz + o(1)\right],\end{aligned}$$

which proves (2.9). In the case $a < \alpha/5$, we have $\gamma_n h_n^{-1} = o\left(h_n^4\right)$; since $\lim_{n\to\infty}(n\gamma_n) > 2a$, Lemma 1 then ensures that

$$\begin{aligned}Var\left[f_n(x)\right] &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k o\left(h_k^4\right) \\ &= o\left(h_n^4\right),\end{aligned}$$

which gives (2.10).

2.5.3 Proof of Proposition 3

Let us first note that, in view of (2.20), we have

$$\begin{aligned}
&\int_{\mathbb{R}} \left\{ \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left[\mathbb{E}\left(Z_k(x)\right) - f(x) \right] \right\}^2 dx \\
&\quad = \frac{1}{4}\mu_2^2 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx \left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2\right]^2 + \int_{\mathbb{R}} \left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 \delta_k(x)\right]^2 dx \\
&\qquad + \mu_2 \left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2\right)\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 \int_{\mathbb{R}} f^{(2)}(x)\,\delta_k(x)\,dx\right).
\end{aligned}$$

Since $f^{(2)}$ is continuous, bounded, and integrable, the application of Lebesgue's convergence theorem ensures that $\lim_{k\to+\infty}\int_{\mathbb{R}} \delta_k^2(x)\,dx = 0$ and $\lim_{k\to+\infty}\int_{\mathbb{R}} f^{(2)}(x)\,\delta_k(x)\,dx = 0$. Moreover, Jensen's inequality gives

$$\begin{aligned}
\int_{\mathbb{R}} \left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 \delta_k(x)\right]^2 dx &\leq \left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2\right)\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 \int_{\mathbb{R}} \delta_k^2(x)\,dx\right) \\
&\leq \left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2\right)\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k o\left(h_k^2\right)\right),
\end{aligned}$$

so that we get

$$\begin{aligned}
&\int_{\mathbb{R}} \left\{ \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left[\mathbb{E}\left(Z_k(x)\right) - f(x) \right] \right\}^2 dx \\
&\quad = \frac{1}{4}\mu_2^2 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx \left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2\right]^2 + O\left(\left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2\right]\left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k o(h_k^2)\right]\right).
\end{aligned}$$

• Let us first consider the case $a \leq \alpha/5$. In this case, $\lim_{n\to\infty}(n\gamma_n) > 2a$, and the application of Lemma 1 gives

$$\int_{\mathbb{R}} \left\{ \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left[\mathbb{E}\left(Z_k(x)\right) - f(x) \right] \right\}^2 dx = \frac{\mu_2^2}{4(1-2a\xi)^2} h_n^4 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx + o\left(h_n^4\right),$$

and ensures that $\Pi_n^2 = o(h_n^4)$. In view of (2.19), we then deduce that

$$\int_{\mathbb{R}} \left\{\mathbb{E}\left(f_n(x)\right) - f(x)\right\}^2 dx = \frac{\mu_2^2}{4(1-2a\xi)^2} h_n^4 \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 dx + o\left(h_n^4\right). \tag{2.21}$$

• Let us now consider the case $a > \alpha/5$. In this case, we have $h_k^2 = o(\sqrt{\gamma_k h_k^{-1}})$ and $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$. The application of Lemma 1 then gives

$$\begin{aligned}
\int_{\mathbb{R}} \left\{ \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left[\mathbb{E}\left(Z_k(x)\right) - f(x) \right] \right\}^2 dx &= O\left(\left[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k o\left(\sqrt{\gamma_k h_k^{-1}}\right)\right]^2\right) \\
&= o\left(\gamma_n h_n^{-1}\right),
\end{aligned}$$

and ensures that $\Pi_n^2 = o(\gamma_n h_n^{-1})$. In view of (2.19), we then deduce that

$$\int_{\mathbb{R}} \{\mathbb{E}(f_n(x)) - f(x)\}^2 dx = o\left(\gamma_n h_n^{-1}\right). \tag{2.22}$$

On the other hand, we note that

$$\begin{aligned}
&\int_{\mathbb{R}} Var[f_n(x)]\, dx \\
&= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2}\gamma_k^2 \int_{\mathbb{R}} Var[Z_k(x)]\, dx \\
&= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2}\gamma_k^2 \left[\frac{1}{h_k}\int_{\mathbb{R}}\int_{\mathbb{R}} K^2(z) f(x - zh_k)\, dz dx - \int_{\mathbb{R}}\left(\int_{\mathbb{R}} K(z) f(x - zh_k)\, dz\right)^2 dx\right]
\end{aligned}$$

with

$$\begin{aligned}
\int_{\mathbb{R}}\int_{\mathbb{R}} K^2(z) f(x - zh_k)\, dz dx &= \int_{\mathbb{R}} K^2(z)\left(\int_{\mathbb{R}} f(x - zh_k)\, dx\right) dz \\
&= \int_{\mathbb{R}} K^2(z)\, dz
\end{aligned}$$

and

$$\begin{aligned}
\int_{\mathbb{R}}\left(\int_{\mathbb{R}} K(z) f(x - zh_k)\, dz\right)^2 dx &= \int_{\mathbb{R}^3} K(z) K(z') f(x - zh_k) f(x - z'h_k)\, dz dz' dx \\
&\leq \|f\|_\infty \|K\|_1^2.
\end{aligned}$$

• In the case $a \geq \alpha/5$, we have $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$, and Lemma 1 ensures that

$$\begin{aligned}
\int_{\mathbb{R}} Var[f_n(x)]\, dx &= \Pi_n^2 \sum_{k=1}^n \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[\int_{\mathbb{R}} K^2(z)\, dz + o(1)\right] \\
&= \frac{\gamma_n}{h_n}\frac{1}{(2 - (\alpha - a)\xi)}\int_{\mathbb{R}} K^2(z)\, dz + o\left(\frac{\gamma_n}{h_n}\right).
\end{aligned} \tag{2.23}$$

• In the case $a < \alpha/5$, we have $\gamma_n h_n^{-1} = o(h_n^4)$ and $\lim_{n\to\infty}(n\gamma_n) > 2a$, so that Lemma 1 gives

$$\begin{aligned}
\int_{\mathbb{R}} Var[f_n(x)]\, dx &= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2}\gamma_k o\left(h_k^4\right) \\
&= o\left(h_n^4\right).
\end{aligned} \tag{2.24}$$

Part 1 of Proposition 3 follows from the combination of (2.21) and (2.24), Part 2 from the one of (2.21) and (2.23), and Part 3 from the one of (2.22) and (2.23).

2.5.4 Proof of Theorem 1

Let us at first assume that, if $a \geq \alpha/5$, then

$$\sqrt{\gamma_n^{-1}h_n}\left(f_n(x) - \mathbb{E}[f_n(x)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{(2 - (\alpha - a)\xi)} f(x)\int_{\mathbb{R}} K^2(z)\, dz\right). \tag{2.25}$$

In the case when $a > \alpha/5$, Part 1 of Theorem 1 follows from the combination of (2.8) and (2.25). In the case when $a = \alpha/5$, Parts 1 and 2 of Theorem 1 follow from the combination of (2.7) and (2.25). In the case $a < \alpha/5$, (2.10) implies that

$$h_n^{-2}\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right) \xrightarrow{\mathbb{P}} 0,$$

and the application of (2.7) gives Part 2 of Theorem 1.

We now prove (2.25). In view of (2.1), we have

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

Set

$$Y_k(x) = \Pi_k^{-1}\gamma_k\left(Z_k(x) - \mathbb{E}\left(Z_k(x)\right)\right). \tag{2.26}$$

The application of Lemma 1 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^{n} Var\left(Y_k(x)\right) \\ &= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var\left(Z_k(x)\right) \\ &= \sum_{k=1}^{n} \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[f(x)\int_{\mathbb{R}} K^2(z)\,dz + o(1)\right] \\ &= \frac{1}{\Pi_n^2}\frac{\gamma_n}{h_n}\left[\frac{1}{2-(\alpha-a)\xi}f(x)\int_{\mathbb{R}} K^2(z)\,dz + o(1)\right]. \end{aligned} \tag{2.27}$$

On the other hand, we have, for all $p > 0$,

$$\mathbb{E}\left[\left|Z_k(x)\right|^{2+p}\right] = O\left(\frac{1}{h_k^{1+p}}\right), \tag{2.28}$$

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

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

and we thus obtain

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

The CLT (2.25) then follows from the application of Lyapounov Theorem.

2.5.5 Proof of Theorem 2

Set

$$S_n(x) = \sum_{k=1}^{n} Y_k(x)$$

where Y_k is defined in (2.26), and set $\gamma_0 = h_0 = 1$.

• Let us first consider the case $a \geq \alpha/5$ (in which case $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$). We set $H_n^2 = \Pi_n^2 \gamma_n^{-1} h_n$, and note that, since $\left(\gamma_n^{-1} h_n\right) \in \mathcal{GS}(\alpha - a)$, we have

$$\begin{aligned}
\ln\left(H_n^{-2}\right) &= -2\ln(\Pi_n) + \ln\left(\prod_{k=1}^{n} \frac{\gamma_{k-1}^{-1} h_{k-1}}{\gamma_k^{-1} h_k}\right) \\
&= -2\sum_{k=1}^{n} \ln(1-\gamma_k) + \sum_{k=1}^{n} \ln\left(1 - \frac{\alpha - a}{k} + o\left(\frac{1}{k}\right)\right) \\
&= \sum_{k=1}^{n} (2\gamma_k + o(\gamma_k)) - \sum_{k=1}^{n} ((\alpha - a)\xi\gamma_k + o(\gamma_k)) \\
&= (2 - \xi(\alpha - a)) s_n + o(s_n).
\end{aligned} \tag{2.29}$$

Since $2 - \xi(\alpha - a) > 0$, it follows in particular that $\lim_{n\to+\infty} H_n^{-2} = \infty$. Moreover, we clearly have $\lim_{n\to+\infty} H_n^2/H_{n-1}^2 = 1$, and by (2.27)

$$\lim_{n\to+\infty} H_n^2 \sum_{k=1}^{n} Var[Y_k(x)] = \frac{1}{2 - (\alpha - a)\xi} f(x) \int_{\mathbb{R}} K^2(z)\, dz.$$

Now, in view of (2.28), $\mathbb{E}\left[|Y_k(x)|^3\right] = O\left(\Pi_k^{-3}\gamma_k^3 h_k^{-2}\right)$ and, since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$, the application of Lemma 1 and of (2.29) gives

$$\begin{aligned}
\frac{1}{n\sqrt{n}} \sum_{k=1}^{n} \mathbb{E}\left(|H_n Y_k(x)|^3\right) &= O\left(\frac{H_n^3}{n\sqrt{n}} \sum_{k=1}^{n} \Pi_k^{-3}\gamma_k^3 h_k^{-2}\right) \\
&= O\left(\frac{H_n^3}{n\sqrt{n}} \sum_{k=1}^{n} \Pi_k^{-3}\gamma_k o\left(\gamma_k^{\frac{3}{2}} h_k^{-\frac{3}{2}}\right)\right) \\
&= o\left(\frac{H_n^3}{n\sqrt{n}} \Pi_n^{-3}\gamma_n^{\frac{3}{2}} h_n^{-\frac{3}{2}}\right) \\
&= o\left(\frac{1}{n\sqrt{n}}\right) \\
&= o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right).
\end{aligned}$$

The application of Theorem 1 of Mokkadem and Pelletier (2006) then ensures that, with probability one, the sequence

$$\left(\frac{H_n S_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right) = \left(\frac{\sqrt{\gamma_n^{-1} h_n}\left(f_n(x) - \mathbb{E}(f_n(x))\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\frac{f(x)}{(2-(\alpha-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz},\sqrt{\frac{f(x)}{(2-(\alpha-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz}\right]. \tag{2.30}$$

In view of (2.29), we have $\lim_{n\to\infty}\ln\ln\left(H_n^{-2}\right)/\ln s_n = 1$. It follows that, with probability one, the sequence $\left(\sqrt{\gamma_n^{-1}h_n}\left(f_n(x)-\mathbb{E}(f_n(x))\right)/\sqrt{2\ln s_n}\right)$ is relatively compact, and its limit set is the interval given in (2.30). The application of (2.7) (respectively (2.8)) concludes the proof of Theorem 2 in the case $a=\alpha/5$ (respectively $a>\alpha/5$).

• Let us now consider the case $a<\alpha/5$ (in which case $\lim_{n\to\infty}(n\gamma_n)>2a$). Set $H_n^{-2}=\Pi_n^{-2}h_n^4\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{-1}$ and note that, since $\left(h_n^{-4}\right)\in\mathcal{GS}(4a)$, we have

$$\begin{aligned}
\ln\left(\Pi_n^{-2}h_n^4\right) &= -2\ln(\Pi_n)+\ln\left(\prod_{k=1}^{n}\frac{h_{k-1}^{-4}}{h_k^{-4}}\right)\\
&= -2\sum_{k=1}^{n}\ln(1-\gamma_k)+\sum_{k=1}^{n}\ln\left(1-\frac{4a}{k}+o\left(\frac{1}{k}\right)\right)\\
&= \sum_{k=1}^{n}(2\gamma_k+o(\gamma_k))-\sum_{k=1}^{n}(4a\xi\gamma_k+o(\gamma_k))\\
&= (2-4a\xi)\,s_n+o(s_n).
\end{aligned} \tag{2.31}$$

Since $2-4a\xi>0$, it follows in particular that $\lim_{n\to\infty}\Pi_n^{-2}h_n^4=\infty$, and thus $\lim_{n\to\infty}H_n^{-2}=\infty$. Moreover, we clearly have $\lim_{n\to\infty}H_n^2/H_{n-1}^2=1$. Set $\epsilon\in\,]0,\alpha-5a[$ such that $\lim_{n\to\infty}(n\gamma_n)>2a+\epsilon/2$; in view of (2.27), and applying Lemma 1, we get

$$\begin{aligned}
H_n^2\sum_{k=1}^{n}Var\left[Y_k(x)\right] &= O\left(\Pi_n^2h_n^{-4}\ln\ln\left(\Pi_n^{-2}h_n^4\right)\sum_{k=1}^{n}\frac{\Pi_k^{-2}\gamma_k^2}{h_k}\right)\\
&= O\left(\Pi_n^2h_n^{-4}\ln\ln\left(\Pi_n^{-2}h_n^4\right)\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k o\left(h_k^4k^{-\epsilon}\right)\right)\\
&= o\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)n^{-\epsilon}\right)\\
&= o(1).
\end{aligned}$$

Moreover, applying (2.28), Lemma 1, and (2.31), we obtain

$$\begin{aligned}
\frac{1}{n\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\left(\left|H_nY_k(x)\right|^3\right) &= O\left(\frac{\Pi_n^3h_n^{-6}}{n\sqrt{n}}\left[\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right]^{\frac{3}{2}}\left(\sum_{k=1}^{n}\Pi_k^{-3}\gamma_k^3h_k^{-2}\right)\right)\\
&= O\left(\frac{\Pi_n^3h_n^{-6}}{n\sqrt{n}}\left[\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right]^{\frac{3}{2}}\left(\sum_{k=1}^{n}\Pi_k^{-3}\gamma_k o\left(h_k^6\right)\right)\right)\\
&= o\left(\frac{\Pi_n^3h_n^{-6}}{n\sqrt{n}}\Pi_n^{-3}h_n^6\left[\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right]^{\frac{3}{2}}\right)\\
&= o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right).
\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier (2006) then ensures that, with probability one,

$$\lim_{n\to\infty}\frac{H_nS_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}=\lim_{n\to\infty}h_n^{-2}\frac{\sqrt{\ln\ln\left(\Pi_n^{-2}h_n^4\right)}}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\left(f_n(x)-\mathbb{E}(f_n(x))\right)=0.$$

Noting that (2.31) ensures that $\lim_{n\to\infty} \ln\ln\left(H_n^{-2}\right)/\ln\ln\left(\Pi_n^{-2}h_n^4\right) = 1$, we deduce that

$$\lim_{n\to\infty} h_n^{-2}\left[T_n(x) - \mathbb{E}(T_n(x))\right] = 0 \quad a.s.,$$

and Theorem 2 in the case $a < \alpha/5$ follows from (2.7).

2.5.6 Proof of Theorem 3

Theorem 3 is proved by showing that

- if $a \geq \alpha/5$, then $\sup_{x\in I}\left|f_n(x) - \mathbb{E}(f_n(x))\right| = O\left(\sqrt{\gamma_n h_n^{-1}}\ln n\right) \quad a.s.$ (2.32)
- if $a > \alpha/5$, then $\sup_{x\in I}\left|\mathbb{E}(f_n(x)) - f(x)\right| = o\left(\sqrt{\gamma_n h_n^{-1}}\ln n\right),$ (2.33)
- if $a < \alpha/5$, then $\sup_{x\in I}\left|f_n(x) - \mathbb{E}(f_n(x))\right| = o\left(h_n^2\right) \quad a.s.$ (2.34)
- if $a \leq \alpha/5$, then $\sup_{x\in I}\left|\mathbb{E}(f_n(x)) - f(x)\right| = O\left(h_n^2\right).$ (2.35)

As a matter of fact, Theorem 3 follows from the combination of (2.32) and (2.33) in the case $a > \alpha/5$, from the one of (2.32) and (2.35) in the case $a = \alpha/5$, and from the one of (2.34) and (2.35) in the case $a < \alpha/5$.

The proof of (2.33) and (2.35) is similar to the one of (2.7) and (2.8) and is omitted. To prove simultaneously (2.32) and (2.34), we introduce the sequence (v_n) defined as

$$(v_n) = \begin{cases} \left(\sqrt{\gamma_n^{-1}h_n}\right) & \text{if} \quad a \geq \alpha/5, \\ \left(h_n^{-2}\left[\ln n\right]^2\right) & \text{if} \quad a < \alpha/5. \end{cases}$$

As a matter of fact, $(v_n) \in \mathcal{GS}(v^*)$ with $v^* = \min\left\{\frac{\alpha-a}{2}, 2a\right\}$, and to prove that (2.32) and (2.34), it is sufficient to prove that

$$\sup_{x\in I}\left|f_n(x) - \mathbb{E}(f_n(x))\right| = O\left(v_n^{-1}\ln n\right) \quad a.s. \tag{2.36}$$

Let us first assume that the following lemma holds.

Lemma 2 *There exists $s > 0$ such that, for all $C > 0$,*

$$\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left|f_n(x) - \mathbb{E}(f_n(x))\right| \geq C\right] = O\left(n^{-C}\right).$$

We first show how (2.36) can be deduced from Lemma 2, and then prove Lemma 2. Set $\rho_n = h_n^2 v_n^{-1}$. One can choose $N(n)$ intervals $I_i^{(n)}$, $i \in \{1,\ldots,N(n)\}$, of length ρ_n and such that $\cup_{i=1}^{N(n)} I_i^{(n)} = I$. For all i set $x_i^{(n)} \in I_i^{(n)}$. We have

$$\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|f_n(x) - \mathbb{E}(f_n(x))\right| \geq C\right] \leq \sum_{i=1}^{N(n)}\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I_i^{(n)}}\left|f_n(x) - \mathbb{E}(f_n(x))\right| \geq C\right].$$

Since K is Lipschitz-continuous, and by application of Lemma 1, there exist $k^*, c^* > 0$, such that, for all $x, y \in \mathbb{R}$, satisfying $|x-y| \leq \rho_n$,

$$\begin{aligned} |f_n(x) - f_n(y)| &= \left| \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-1} \left[K\left(\frac{x-X_k}{h_k}\right) - K\left(\frac{y-X_k}{h_k}\right)\right]\right| \\ &\leq k^* \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-2} \rho_n \\ &\leq c^* h_n^{-2} \rho_n. \end{aligned}$$

It follows that, for all $x \in I_i^{(n)}$, we have

$$\begin{aligned} &|f_n(x) - \mathbb{E}(f_n(x))| \\ &\leq \left|f_n(x) - f_n\left(x_i^{(n)}\right)\right| + \left|f_n\left(x_i^{(n)}\right) - \mathbb{E}\left(f_n\left(x_i^{(n)}\right)\right)\right| + \left|\mathbb{E}\left(f_n\left(x_i^{(n)}\right)\right) - \mathbb{E}(f_n(x))\right| \\ &\leq 2c^* h_n^{-2}\rho_n + \left|f_n\left(x_i^{(n)}\right) - \mathbb{E}\left(f_n\left(x_i^{(n)}\right)\right)\right|, \end{aligned}$$

so that, for all $C > 0$,

$$\begin{aligned} &\mathbb{P}\left[\frac{v_n}{\ln n} \sup_{x\in I} |f_n(x) - \mathbb{E}(f_n(x))| \geq C\right] \\ &\leq \sum_{i=1}^{N(n)} \mathbb{P}\left[\frac{v_n}{\ln n}\left|f_n\left(x_i^{(n)}\right) - \mathbb{E}\left(f_n\left(x_i^{(n)}\right)\right)\right| + 2c^* \frac{v_n}{\ln n} h_n^{-2}\rho_n \geq C\right]. \end{aligned}$$

Now, for n large enough, we have $2c^* v_n (\ln n)^{-1} h_n^{-2} \rho_n \leq C/2$. Applying Lemma 2 and noting that $N(n) = O\left(\rho_n^{-1}\right)$, we obtain, for n large enough,

$$\begin{aligned} \mathbb{P}\left[\frac{v_n}{\ln n} \sup_{x\in I} |f_n(x) - \mathbb{E}(f_n(x))| \geq C\right] &\leq N(n) \sup_{x\in I} \mathbb{P}\left[\frac{v_n}{\ln n} |f_n(x) - \mathbb{E}(f_n(x))| \geq \frac{C}{2}\right] \\ &= O\left(\rho_n^{-1} n^{-\frac{C}{2}}\right). \end{aligned}$$

Since $\left(\rho_n^{-1}\right) \in \mathcal{GS}(v^* + 2a)$, it follows that, for C large enough,

$$\sum_{n\geq 2} \mathbb{P}\left[\frac{v_n}{\ln n} \sup_{x\in I} |f_n(x) - \mathbb{E}(f_n(x))| \geq C\right] < \infty,$$

and the application of Borel-Cantelli Lemma gives

$$\sup_{x\in I} |f_n(x) - \mathbb{E}(f_n(x))| = O\left(v_n^{-1} \ln n\right) \quad a.s.,$$

which proves (2.36).

Proof of Lemma 2 For all $x \in I$, we have

$$\begin{aligned} \mathbb{P}\left[\frac{v_n}{\ln n}(f_n(x) - \mathbb{E}(f_n(x))) \geq C\right] &= \mathbb{P}\left[\exp\left[v_n(f_n(x) - \mathbb{E}(f_n(x)))\right] \geq n^C\right] \\ &\leq n^{-C} \mathbb{E}\left(\exp\left[v_n(f_n(x) - \mathbb{E}(f_n(x)))\right]\right) \\ &\leq n^{-C} \prod_{k=1}^{n} \mathbb{E}\left(\exp\left(V_{k,n}(x)\right)\right) \end{aligned}$$

with

$$V_{k,n}(x) = v_n \Pi_n \Pi_k^{-1} \gamma_k h_k^{-1} \left[K\left(\frac{x - X_k}{h_k}\right) - \mathbb{E}\left(K\left(\frac{x - X_k}{h_k}\right)\right)\right].$$

Note that

$$|V_{k,n}(x)| \leq 2\|K\|_\infty \frac{\Pi_n v_n}{\Pi_k v_k} v_k \gamma_k h_k^{-1}.$$

Since $(v_n) \in \mathcal{GS}(v^*)$ with $v^* = \min\left\{\frac{\alpha - a}{2}, 2a\right\}$, we have $1 - v^*\xi > 0$, and thus

$$\begin{aligned}
\frac{\Pi_n}{\Pi_{n-1}} \frac{v_n}{v_{n-1}} &= (1 - \gamma_n)\left(1 + \frac{v^*}{n} + o\left(\frac{1}{n}\right)\right) \\
&= (1 - \gamma_n)\left(1 + v^*\xi\gamma_n + o(\gamma_n)\right) \\
&= 1 - (1 - v^*\xi)\gamma_n + o(\gamma_n) \\
&\leq 1 \qquad \text{for } n \text{ large enough.}
\end{aligned}$$

Writing

$$\frac{v_n \Pi_n}{v_k \Pi_k} = \prod_{i=k}^{n-1} \frac{v_{i+1}\Pi_{i+1}}{v_i \Pi_i},$$

we obtain

$$\sup_n \sup_{k \leq n} \frac{v_n \Pi_n}{v_k \Pi_k} < \infty.$$

Since the sequence $\left(v_k \gamma_k h_k^{-1}\right)$ is bounded, it follows that there exists $M > 0$ such that, for all $x \in \mathbb{R}$, $|V_{k,n}(x)| \leq M$. We deduce that, for all $x \in I$,

$$\begin{aligned}
\mathbb{E}\left(\exp\left[V_{k,n}(x)\right]\right) &\leq 1 + \mathbb{E}\left[V_{k,n}^2(x)\right]\exp[M] \\
&\leq 1 + v_n^2 \Pi_n^2 \Pi_k^{-2} \gamma_k^2 h_k^{-2} Var\left[K\left(\frac{x - X_k}{h_k}\right)\right]\exp[M].
\end{aligned}$$

Noting that

$$\begin{aligned}
Var\left(K\left(\frac{x - X_k}{h_k}\right)\right) &\leq h_k \int_{\mathbb{R}} K^2(z) f(x - zh_k)\, dz + h_k^2 \left(\int_{\mathbb{R}} K(z) f(x - zh_k)\, dz\right)^2 \\
&\leq h_k \|f\|_\infty \int_{\mathbb{R}} K^2(z)\, dz + h_k^2 \|f\|_\infty^2 \|K\|_1^2,
\end{aligned}$$

and since the sequence $\left(\gamma_k h_k^{-1} v_k^2\right)$ is bounded, we deduce that there exist $M_1^*, M_2^* > 0$ such that, for all $x \in I$,

$$\begin{aligned}
\mathbb{E}\left(\exp\left[V_{k,n}(x)\right]\right) &\leq 1 + M_1^* v_n^2 \Pi_n^2 \Pi_k^{-2} \gamma_k^2 h_k^{-1} \\
&\leq \exp\left[M_1^* v_n^2 \Pi_n^2 \Pi_k^{-2} \gamma_k v_k^{-2}\left(\gamma_k h_k^{-1} v_k^2\right)\right] \\
&\leq \exp\left[M_2^* v_n^2 \Pi_n^2 \Pi_k^{-2} \gamma_k v_k^{-2}\right].
\end{aligned}$$

Applying Lemma 1, it follows that, for all $C > 0$,

$$\begin{aligned}
\sup_{x \in I} \mathbb{P}\left[\frac{v_n}{\ln n}\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right) \geq C\right] &\leq n^{-C} \exp\left[M_2^* v_n^2 \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k v_k^{-2}\right] \\
&= O\left(n^{-C}\right).
\end{aligned}$$

We establish in the same way that

$$\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left(\mathbb{E}\left(f_n\left(x\right)\right)-f_n\left(x\right)\right)\geq C\right] \quad = \quad O\left(n^{-C}\right),$$

which concludes the proof of Lemma 2.

Bibliographie

[1] Blum, J.R. (1954), Multidimensional stochastic approximation methods, *Ann. Math. Statist.,* **25**, 737-744

[2] Bojanic, R. and Seneta, E. (1973), A unified theory of regularly varying sequences, *Math.* Z., **134**, 91-106.

[3] Chen, H. (1988), Lower rate of convergence for locating a maximum of a function, *Ann. Statist.,* **16**, 1330-1334

[4] Chen, H.F., Duncan, T.E., and Pasik-Duncan, B. (1999), A Kiefer-Wolfowitz algorithm with randomized differences, *IEEE Trans. Automat. Control,* **44**, 442-453

[5] Davies, H.L. (1973), Strong consistency of a sequential estimator of a probability density function. *Bull. Math. Statist.* **15**, pp. 49-54.

[6] Deheuvels, P. (1973), *Sur l'estimation séquentielle de la densité, C. R. Acad. Sci. Paris Ser. A-B* **276**, 1119-1121.

[7] Devroye, L. (1979), On the pointwise and integral convergence of recursive kernel estimates of probability densities. *Utilitas Math.* **15**, pp. 113-128.

[8] Dippon, J. and Renz J. (1997), Weighted means in stochastic approximation of minima, *SIAM J. Control Optim.,* **35**, 1811-1827

[9] Dippon, J. (2003), Accelerated randomized stochastic optimization, *Ann. Statist.,* **31**, 1260-1281

[10] Duflo, M. (1996), *Algorithmes stochastiques*, Collection Applications of mathematics, Springer.

[11] Duflo, M. (1997), *Random Iterative Models*, Collection Applications of mathematics, Springer.

[12] Fabian, V. (1967), Stochastic approximation of minima with improved asymptotic speed, *Ann. Math. Statist.,* **38**, 191-200

[13] Galambos, J. and Seneta, E. (1973), Regularly varying sequences, *Proc. Amer. Math. Soc.*, **41**, 110-116.

[14] Hall, P. and Heyde, C.C. (1980), *Martingale limit theory and its application,* Academic Press, Inc., New York-London

[15] Hall, P. (1992), Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Ann. Statist.* **20**, 675-694

[16] Kiefer, J. and Wolfowitz, J. (1952), Stochastic approximation of the maximum of a regression functions, *Ann. Math. Statist.,* **23**, 462-466

[17] Kushner H.J. and Clark, D.S. (1978) *Stochastic approximation methods for constrained and unconstrained systems,* Springer, New York

[18] 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**, pp. 73-82.

[19] Mokkadem, A. and Pelletier, M. (2004), A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *submitted*

[20] Mokkadem, A. and Pelletier, M. (2006), Compact law of the iterated logarithm for matrix-normalized sums of random vectors, *to appear in Theory Probab. Appl.*

[21] Parzen, E. (1962). On Estimation of a Probability Density and Mode, *Ann. Math. Statist.*, **33**, 1065-1076.

[22] Polyak, B.T. and Tsybakov, A.B. (1990), Optimal orders of accuracy for search algorithms of stochastic optimization, *Problems Inform. Transmission,* **26**, 126-133.

[23] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119-126.

[24] Rosenblatt, M. (1956) . Remarks on Some Nonparametric Estimates of a Density Function, *Ann. Math. Statist.*, **27**, 832-837.

[25] 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. *J. of Nonparam. Statist.* **3**, pp.171-195.

[26] Ruppert, D. (1982), Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise, *Ann. of Probab.,* **10**, 178-187.

[27] Spall, J.C. (1988), A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting. In *Proc. Conference on Decision and Control*, 1544-1548. IEEE, New York.

[28] Spall, J.C. (1997), A one-measurement form of simultaneous perturbation stochastic approximation, *Automatica J. IFAC,* **33**, 109-112.

[29] Tsybakov, A.B. (1990), Recurrent estimation of the mode of a multidimensional distribution, *Problems Inform. Transmission* **26**, 31-37.

[30] Wegman, E.J. and Davies, H. I. (1979), Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7**, pp. 316-327.

[31] Wertz, W. (1985), Sequential and recursive estimators of the probability density. *Statistics* **16**, pp. 277-295.

[32] Wolverton, C.T. and Wagner, T.J. (1969), Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory* **15**, pp. 258-265.

[33] Yamato, H. (1971), Sequential estimation of a continuous probability density function and mode. *Bull. Math. Satist.* **14**, pp. 1-12.

Chapitre 3

Application of the averaging principle of stochastic approximation in the estimation of a regression function

3.1 Introduction

The use of stochastic approximation algorithms in the framework of regression estimation has been introduced by Kiefer and Wolfowitz (1952). The famous Kiefer and Wolfowitz algorithm allows the approximation of the point at which a regression function reaches its maximum. This pioneer work has been widely discussed and extended in many directions (see, among many others, Blum (1954), Fabian (1967), Kushner and Clark (1978), Hall and Heyde (1980), Ruppert (1982), Chen (1988), Spall (1988), Polyak and Tsybakov (1990), Dippon and Renz (1997), Spall (1997), Chen, Duncan and Pasik-Duncan (1999), Dippon (2003), and Mokkadem and Pelletier (2004)).

The question of applying the Robbins-Monro procedure to construct a stochastic approximation algorithm, which allows the estimation of a regression function at a given point (instead of approximating its mode) has been introduced by Révész (1973).

Let us recall that the Robbins-Monro procedure consists in building up stochastic approximation algorithms, which allow the search of the zero z^* of an unknown function $h : \mathbb{R} \to \mathbb{R}$. These algorithms are constructed in the following way : (i) $Z_0 \in \mathbb{R}$ is arbitrarily chosen ; (ii) the sequence (Z_n) is recursively defined by setting

$$Z_n = Z_{n-1} + \gamma_n \mathcal{W}_n$$

where $\mathcal{W}_n$ is an observation of the function h at the point Z_{n-1}, and where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero.

Let $((X_1, Y_1), \ldots, (X_n, Y_n))$ be independent, identically distributed pairs of random variables, and let f denote the probability density of X. 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 that $f(x) \neq 0$, Révész (1973) defines an algorithm, which approximates the zero of the function $h : y \mapsto f(x)r(x) - f(x)y$. Following the Robbins-Monro procedure, this algorithm is defined by setting $r_0(x) \in \mathbb{R}$ and, for $n \geq 1$,

$$r_n(x) = r_{n-1}(x) + \frac{1}{n}\mathcal{W}_n(x)$$

where $\mathcal{W}_n(x)$ is an "observation" of the function h at the point $r_{n-1}(x)$. To define $\mathcal{W}_n(x)$, Révész (1973) introduces a kernel K (that is, a function satisfying $\int_{\mathbb{R}} K(x)dx = 1$) and a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and sets

$$\mathcal{W}_n(x) = h_n^{-1} Y_n K(h_n^{-1}[x - X_n]) - h_n^{-1} K(h_n^{-1}[x - X_n]) r_{n-1}(x). \tag{3.1}$$

Révész (1977) chooses the bandwidth (h_n) equal to (n^{-a}) with $a \in]1/2, 1[$, and establishes a central limit theorem for $r_n(x) - r(x)$ under the assumption $f(x) > (1-a)/2$, as well as an upper bound of the uniform strong convergence rate of r_n on any bounded interval I on which $\inf_{x \in I} f(x) > (1-a)/2$. The two drawbacks of his approach are the following. First, the condition $a > 1/2$ on the bandwidth leads to a convergence rate of r_n slower than $n^{1/4}$, whereas the optimal convergence rate of the kernel estimator of a regression function introduced by Nadaraya (1964) and Watson (1964) is $n^{2/5}$ (and obtained by choosing $a = 1/5$). Then, the condition $f(x) > (1-a)/2$ (or $\inf_{x \in I} f(x) > (1-a)/2$) is stronger than the condition $f(x) > 0$ (or $\inf_{x \in I} f(x) > 0$) usually required to establish the convergence rate of regression's estimators. To understand why this second drawback is inherent in the definition of Révész's estimator, let us come back on the Robbins-Monro algorithm.

The convergence rate of the Robbins-Monro algorithm, and, more generally, of stochastic approximation algorithms used for the search of the zero z^* of an unknown function h, has been widely studied (see, among many others, Nevels'on and Has'minskii (1976), Kushner and Clark (1978), Ljung, Pflug, and Walk (1992), and Duflo (1996)). It is now well known that the convergence rate of such algorithms is obtained under the condition that the limit of the sequence $(n\gamma_n)$ as n goes to infinity is larger than a quantity, which involves the differential of h at the point z^*. In the case the stepsize (γ_n) is chosen such that $\lim_{n\to\infty} n\gamma_n = \gamma^* \in]0,\infty[$, this condition is seen, in the framework of stochastic approximation algorithms, as a tedious condition on the parameter γ^* ; in the case the stepsize is chosen such that $\lim_{n\to\infty} n\gamma_n = \infty$, this condition is automatically fulfilled, but the convergence rate of the corresponding stochastic approximation algorithm is then slower than for the previous choice of stepsize.

Let us underline that the stepsize (γ_n) used by Révész (1973, 1977) is (n^{-1}), so that $\gamma^* = 1$; moreover, the differential of the function $h\ : y \mapsto f(x)r(x) - f(x)y$ at the point $y^* = r(x)$ equals $-f(x)$. Consequently, the tedious condition which involves the parameter γ^* and the differential of h in the framework of stochastic approximation algorithms, comes down, in Révész's framework, to a tedious condition on the probability density f.

Now, the famous approach to obtain optimal convergence rates for stochastic approximation algorithms without tedious condition on the stepsize is to use the averaging principle independently introduced by Ruppert (1988) and Polyak (1990). Their averaging procedure, which has been widely discussed and extended (see, among many others, Yin (1991), Delyon and Juditsky (1992), Polyak and Juditsky (1992), Kushner and Yang (1993), Le Breton (1993), Le Breton and Novikov (1995), Dippon and Renz (1996, 1997), and Pelletier (2000)) allows to obtain asymptotically efficient algorithms, that is, algorithms which not only converge at the optimal rate, but which also have an optimal asymptotic variance. This procedure consists in : (i) running the approximation algorithm by using slower stepsizes ; (ii) computing a suitable average of the approximations obtained in (i).

Our aim in this chapter is to introduce the averaging principle of stochastic approximation algorithms in the framework of regression estimation. To this end, we first need to generalize the definition of Révész's estimator. More precisely, we define the estimator r_n by setting $r_0(x) \in \mathbb{R}$ and, for $n \geq 1$,

$$r_n(x) = r_{n-1}(x) + \gamma_n \mathcal{W}_n(x) \tag{3.2}$$

where $\mathcal{W}_n(x)$ is defined in (3.1), and where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero. Then, to apply the averaging principle, we proceed as follows. We first run the algorithm (3.2) by using a stepsize such that $\lim_{n\to\infty} n\gamma_n = \infty$, and then define a new estimator $\overline{r}_n$ as an average of the r_k.

Let us dwell on the fact that, although Révész's estimator has been constructed by following the Robbins-Monro procedure, the algorithm (3.2) is not "of Robbins-Monro type". As a matter of fact, the use of the kernel K and of the bandwidth (h_n) to construct the "observations" $\mathcal{W}_n(x)$ is, in a certain sense, very similar to the approximation of the differential of the regression function made in the construction of the Kiefer and Wolfowitz algorithm, and the behaviour of the algorithm (3.2) that it induces is very comparable with the behaviour of the Kiefer and Wolfowitz algorithm. Consequently, the average of the r_k, which leads to an estimator $\overline{r}_n$ with a minimal asymptotic variance, is not the arithmetic average (as it is the case for the Robbins-Monro algorithm), but a weighted average (as it is the case for the Kiefer-Wolfowitz algorithm). To define the averaged estimator $\overline{r}_n$, we thus introduce a sequence of positive real numbers (q_n) such that $\sum q_n = \infty$, and set

$$\overline{r}_n = \frac{1}{\sum_{k=1}^{n} q_k} \sum_{k=1}^{n} q_k r_k. \tag{3.3}$$

We first establish the asymptotic behaviour of the generalized Révész's estimator defined by (3.2), and then the one of the averaged estimator $\overline{r}_n$ defined by (3.3).

For the study of the generalized Révész's estimator, the technic we use, which is totally different from the one employed by Révész (1977), is the following. Noting that the approximation algorithm

(3.2) can be rewritten as

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

we approximate the algorithm (3.2) by the unobservable sequence (ρ_n) recursively defined by

$$\rho_n(x) = (1 - \gamma_n f(x))\, \rho_{n-1}(x) + \gamma_n \left[f(x) - h_n^{-1} K\left(\frac{x - X_n}{h_n}\right)\right] r(x) + \gamma_n h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right). \quad (3.4)$$

The asymptotic properties (pointwise weak and strong convergence rate, upper bound of the uniform strong convergence rate) of the approximating algorithm (3.4) are established by applying different results on the sums of independent variables and on the martingales. To show that the asymptotic properties of the approximating algorithm (3.4) are also satisfied by the generalized Révész's estimator, we use a technic of successive upper bounds.

In the case the stepsize (γ_n) in (3.2) is chosen such that $\lim_{n\to\infty} n\gamma_n = \gamma^* \in]0, \infty[$, we require a condition, which involves γ^* and the probability density f; in the case (γ_n) is such that $\lim_{n\to\infty} n\gamma_n = \infty$, this tedious condition disappears.

To study the asymptotic behaviour of the averaged estimator $\overline{r}_n$, we first establish the asymptotic properties (pointwise weak and strong convergence rate, upper bound of the uniform strong convergence rate) of the averaged sequence $\overline{\rho}_n$ defined as

$$\overline{\rho}_n = \frac{1}{\sum_{k=1}^n q_k} \sum_{k=1}^n q_k \rho_k,$$

and then show how the asymptotic behaviour of $\overline{r}_n$ can be deduced from the one of the approximating sequence $\overline{\rho}_n$.

The condition we require on the density f to prove the pointwise (respectively uniform) convergence rate of $\overline{r}_n$ is the usual condition $f(x) > 0$ (respectively $\inf_{x\in I} f(x) > 0$). Concerning the bandwidth, we allow the choice $(h_n) = (n^{-1/5})$, which leads to the optimal convergence rate $n^{2/5}$. Finally, we show that to construct confidence intervals by slightly undersmoothing, it is preferable to use the averaged Révész estimator $\overline{r}_n$ (with an adequate choice of weights (q_n)) rather than the Nadaraya-Watson estimator, since the asymptotic variance of this lattest estimator is larger than the one of $\overline{r}_n$.

Our chapter is organized as follows. Our assumptions and main results are stated in Section 3.2, the outlines of the proofs given in Section 3.3, whereas Section 3.4 is devoted to the proof of several lemmas.

3.2 Assumptions and main results

Let us first define the class of positive sequences that will be used in the statement of our assumptions.

Definition 2 *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*

$$\lim_{n\to\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma \quad (3.5)$$

Condition (3.5) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1973)). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\log n)^{b}$, $n^{\gamma}(\log\log n)^{b}$.

Let $g(s,t)$ denote the density of the couple (X,Y) (in particular $f(x) = \int_{\mathbb{R}} g(x,t)\,dt$), and set $a(x) = r(x) f(x)$.
The assumptions to which we will refer for our pointwise results are the following ones.

(A1) $K : \mathbb{R} \to \mathbb{R}$ is a nonnegative, continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\int_{\mathbb{R}} z^2 K(z)\,dz < \infty$.

(A2) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in \left]\frac{3}{4}, 1\right]$; moreover the limit of $(n\gamma_n)^{-1}$ as n goes to infinity exists.
ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in \left]\frac{1-\alpha}{4}, \frac{\alpha}{3}\right[$.

(A3) *i)* $g(s,t)$ is two times continuously differentiable with respect to s.
ii) For $q \in \{0,1,2\}$, $s \mapsto \int_{\mathbb{R}} t^q g(s,t)\,dt$ is a bounded function continuous at $s = x$.
For $q \in [2,3]$, $s \mapsto \int_{\mathbb{R}} |t|^q g(s,t)\,dt$ is a bounded function.
iii) For $q \in \{0,1\}$, $\int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt < \infty$, and $s \mapsto \int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial s^2}(s,t)\,dt$ is a bounded function continuous at $s = x$.

For our uniform results, we will also need the following additional assumption.

$(A4)$ *i)*K is Lipschitz-continuous.
ii) There exists $t^* > 0$ such that $\mathbb{E}(\exp(t^*|Y|)) < \infty$.
iii) $a \in]1-\alpha, \alpha - 2/3[$.
iv) For $q \in \{0,1\}$, $x \mapsto \int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt$ is bounded on the set $\{x, f(x) > 0\}$.

Throughout this chapter we will use the following notation :

$$\xi = \lim_{n\to\infty} (n\gamma_n)^{-1}, \tag{3.6}$$

and, for $f(x) \neq 0$,

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

The asymptotic properties of the generalized Révész estimator defined in (3.2) are stated in Section 3.2.1, the ones of the averaged estimator defined in (3.3) in Section 3.2.2.

3.2.1 Asymptotic behaviour of the generalized Révész estimator

The aim of this section is to state the asymptotic properties of the generalized Révész estimator defined in (3.2). Theorems 4, 5, and 6 below give its weak pointwise convergence rate, its strong pointwise convergence rate, and an upper bound of its strong uniform convergence rate, respectively.

Theorem 4 (Weak pointwise convergence rate of r_n)
Let Assumptions $(A1)-(A3)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $\gamma_n^{-1} h_n^5 \to c$, and if $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2f(x))$, then*

$$\sqrt{\gamma_n^{-1}h_n}\,(r_n(x) - r(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}f(x)\,m^{(2)}(x)}{f(x)-2a\xi}, \frac{Var[Y|X=x]\,f(x)}{2f(x)-(1-a)\xi}\int_{\mathbb{R}} K^2(z)\,dz\right).$$

2. If $\gamma_n^{-1}h_n^5 \to \infty$, and if $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x)$, then

$$\frac{1}{h_n^2}\left(r_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} \frac{f(x)\, m^{(2)}(x)}{(f(x) - 2a\xi)},$$

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

Theorem 5 (Strong pointwise convergence rate of r_n)
Let Assumptions $(A1)-(A3)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $\gamma_n^{-1}h_n^5/\ln\left(\sum_{k=1}^n \gamma_k\right) \to c$, and if $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2f(x))$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{\gamma_n^{-1}h_n}{2\ln\left(\sum_{k=1}^n \gamma_k\right)}}\left(r_n(x) - r(x)\right)\right)$$

is relatively compact and its limit set is the interval

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

2. *If $\gamma_n^{-1}h_n^5/\ln\left(\sum_{k=1}^n \gamma_k\right) \to \infty$, and if $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x)$, then, with probability one,*

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(r_n(x) - r(x)\right) = \frac{f(x)\, m^{(2)}(x)}{f(x) - 2a\xi}.$$

Theorem 6 (Strong uniform convergence rate of r_n)
Let I be a bounded open interval of $\mathbb{R}$ on which $\varphi = \inf_{x\in I} f(x) > 0$, and let Assumptions $(A1)-(A4)$ hold for all $x \in I$.

1. *If the sequence $\left(\gamma_n^{-1}h_n^5/[\ln n]^2\right)$ is bounded and if $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2\varphi)$, then*

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

2. *If $\lim_{n\to\infty}\left(\gamma_n^{-1}h_n^5/[\ln n]^2\right) = \infty$ and if $\lim_{n\to\infty}(n\gamma_n) > 2a/\varphi$, then*

$$\sup_{x\in I}\left|r_n(x) - r(x)\right| = O\left(h_n^2\right) \quad a.s.$$

Part 1 of Theorems 4 and 6 have been obtained by Révész (1977) for the choices $(\gamma_n) = (n^{-1})$ and $(h_n) = (n^{-a})$ with $a \in]1/2, 1[$. Let us underline that, for this choice of stepsize, the conditions $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2f(x))$ and $\lim_{n\to\infty}(n\gamma_n) > (1-a)/(2\inf_{x\in I} f(x))$ come down to the following conditions on the unknown density f : $f(x) > (1-a)/2$ and $\inf_{x\in I} f(x) > (1-a)/2$. Let us also mention that our assumption (A2) implies $a \in]0, 1/3[$, so that our results on the generalized Révész estimator do not include the results given in Révész (1977). However, our assumptions include the choice $(\gamma_n) = (n^{-1})$ and $a = 1/5$, which leads to the optimal weak convergence rate

$n^{2/5}$, whereas the condition on the bandwidth required by Révész leads to a convergence rate of r_n slower than $n^{1/4}$.

Although the optimal convergence rate we obtain for the generalized Révész estimator r_n has the same order as the one of the Nadaraya-Watson estimator, this estimator has a main drawback : to make r_n converge with its optimal rate, one must set $a = 1/5$ and choose (γ_n) such that $\lim_{n\to\infty} n\gamma_n = \gamma^* \in]0, \infty[$ with $\gamma^* > 2/[5f(x)]$ whereas the density f is unknown. This tedious condition disappears as soon as the stepsize is chosen such that $\lim_{n\to\infty} n\gamma_n = \infty$ (for instance when $(\gamma_n) = ((\ln n)^b n^{-1})$ with $b > 0$), but the optimal convergence rate $n^{2/5}$ is not reached any more.

3.2.2 Asymptotic behaviour of the averaged Révész estimator

To state the asymptotic properties of the averaged Révész estimator defined in (3.3), we need the following additional assumptions.

(A5) $\lim_{n\to\infty} n\gamma_n \left(\ln\left(\sum_{k=1}^n \gamma_k\right)\right)^{-1} = \infty$ and $a \in]1-\alpha, (4\alpha-3)/2[$.

(A6) $(q_n) \in \mathcal{GS}(-q)$ with $q < \min\{1-2a, (1+a)/2\}$.

Theorem 7 (Weak pointwise convergence rate of $\bar{r}_n$)
Let Assumptions $(A1)-(A3)$, $(A5)$ and $(A6)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $nh_n^5 \to c$, then*

$$\sqrt{nh_n}\left(\bar{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(c^{\frac{1}{2}}\frac{1-q}{1-q-2a}m^{(2)}(x), \frac{(1-q)^2}{1+a-2q}\frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz\right).$$

2. *If $nh_n^5 \to \infty$, then*

$$\frac{1}{h_n^2}\left(\bar{r}_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} \frac{1-q}{1-q-2a}m^{(2)}(x).$$

Let us note that, whatever the choices of the stepsize (γ_n) and of the weight (q_n) are, the convergence rate of the averaged Révész estimator has the same order as the one of the generalized Révész estimator when the stepsize in (3.2) is chosen belonging to $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = \gamma^* < \infty$, but, this time, there is no need to add any condition on the marginal density f.

Hall (1992) shows that, to construct confidence intervals, slightly undersmoothing is more efficient than bias estimation. To undersmooth, we choose (h_n) such that $\lim_{n\to\infty} nh_n^5 = 0$ (and thus $a \geq 1/5$). Moreover, to construct a confidence interval for $r(x)$, it is advised to choose the weight (q_n), which minimizes the asymptotic variance of $\bar{r}_n$. For a given a, the function $q \mapsto (1-q)^2/(1+a-2q)$ reaching its minimum at the point $q = a$, we can state the following corollary.

Corollary 4
Let Assumptions $(A1)-(A3)$, $(A5)$ and $(A6)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$, and with $a \geq 1/5$. To minimize the asymptotic variance of $\bar{r}_n$, q must be chosen equal to a. Moreover, if $\lim_{n\to\infty} nh_n^5 = 0$, we then have

$$\sqrt{nh_n}\left(\bar{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1-a)\frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz\right).$$

Let us recall that, when the bandwidth (h_n) is chosen such that $\lim_{n\to\infty} nh_n^5 = 0$, the Nadaraya-Watson estimator, which is defined as

$$\tilde{r}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(h_n^{-1}(x - X_i)\right)}{\sum_{i=1}^n K\left(h_n^{-1}(x - X_i)\right)},$$

satisfies the central limit theorem

$$\sqrt{nh_n}\left(\tilde{r}_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{Var\left[Y|X=x\right]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz\right).$$

To construct confidence intervals for $r(x)$, it is thus clearly better to use the averaged Révész estimator, rather than the Nadaraya-Watson estimator.

We now state the strong pointwise convergence rate of the averaged Révész estimator, as well as an upper bound of its strong uniform convergence rate.

Theorem 8 (Strong pointwise convergence rate of $\bar{r}_n$)
Let Assumptions $(A1)-(A3)$, $(A5)$ and $(A6)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c_1 \geq 0$ such that $nh_n^5/\ln\ln n \to c_1$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{nh_n}{2\ln\ln n}}\left(\bar{r}_n(x) - r(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[c_1^{\frac{1}{2}}\frac{1-q}{1-q-2a}m^{(2)}(x) - \sqrt{\frac{(1-q)^2}{1+a-2q}\frac{Var\left[Y/X=x\right]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz},\right.$$
$$\left. c_1^{\frac{1}{2}}\frac{1-q}{1-q-2a}m^{(2)}(x) + \sqrt{\frac{(1-q)^2}{1+a-2q}\frac{Var\left[Y/X=x\right]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz}\right].$$

2. *If $nh_n^5/\ln\ln n \to \infty$, then*

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(\bar{r}_n(x) - r(x)\right) = \frac{1-q}{1-q-2a}m^{(2)}(x) \quad a.s.$$

Theorem 9 (Strong uniform convergence rate of $\bar{r}_n$)
Let I be a bounded open interval of $\mathbb{R}$ on which $\inf_{x\in I} f(x) > 0$, and let Assumptions $(A1)-(A6)$ hold for all $x \in I$.

1. *If the sequence $\left(nh_n^5/[\ln n]^2\right)$ is bounded, and if $\alpha > (3a+3)/4$, then*

$$\sup_{x\in I}\left|\bar{r}_n(x) - r(x)\right| = O\left(\sqrt{n^{-1}h_n^{-1}}\ln n\right) \quad a.s.$$

2. *If $\lim_{n\to\infty}\left(nh_n^5/[\ln n]^2\right) = \infty$, and if, in the case $a \in [\alpha/5, 1/5]$, $\alpha > (4a+1)/2$, then*

$$\sup_{x\in I}\left|\bar{r}_n(x) - r(x)\right| = O\left(h_n^2\right) \quad a.s.$$

3.3 Outlines of the proofs

From now on, we set $n_0 \geq 3$ such that $\forall k \geq n_0$, $\gamma_k \leq (2\|f\|_\infty)^{-1}$ and $\gamma_k h_k^{-1}\|K\|_\infty \leq 1$. Moreover, we introduce the following notations :

$$\begin{aligned} s_n &= \sum_{k=n_0}^{n} \gamma_k, \\ Z_n(x) &= h_n^{-1} K\left(\frac{x - X_n}{h_n}\right), \\ W_n(x) &= h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right), \end{aligned}$$

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 (m_n) and $(\tilde{m}_n)$ by setting

$$(m_n) = \begin{cases} \left(\sqrt{\gamma_n h_n^{-1}}\right) & \text{if } \lim_{n\to\infty}\left(\gamma_n h_n^{-5}\right) = \infty, \\ \left(h_n^2\right) & \text{otherwise.} \end{cases} \tag{3.7}$$

$$(\tilde{m}_n) = \begin{cases} \left(\sqrt{\gamma_n h_n^{-1}} \ln n\right) & \text{if } \lim_{n\to\infty}\left(\gamma_n h_n^{-5} \ln n\right) = \infty, \\ \left(h_n^2\right) & \text{otherwise.} \end{cases} \tag{3.8}$$

Note that the sequences (m_n) and $(\tilde{m}_n)$ belong to $\mathcal{GS}(-m^*)$ with

$$m^* = \min\left\{\frac{\alpha - a}{2}, 2a\right\}. \tag{3.9}$$

Before giving the outlines of the proofs, we state the following technical lemma, which is proved in the first chapter, and which will be used throughout the demonstrations.

Lemma 3
Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha > 0$, and set $m > 0$. If $ms - v^*\xi > 0$ (where ξ is defined in (3.6)), then

$$\lim_{n\to\infty} v_n \Pi_n^m(s) \sum_{k=n_0}^{n} \Pi_k^{-m}(s) \frac{\gamma_k}{v_k} = \frac{1}{ms - v^*\xi}.$$

Moreover, for all positive sequence (α_n) such that $\lim_{n\to\infty} \alpha_n = 0$, and all C,

$$\lim_{n\to\infty} v_n \Pi_n^m(s) \left[\sum_{k=n_0}^{n} \Pi_k^{-m}(s) \frac{\gamma_k}{v_k} \alpha_k + C\right] = 0.$$

As explained in the introduction, we note that the stochastic approximation algorithm (3.2) can be rewritten as :

$$
\begin{aligned}
r_n(x) &= \left(1-\gamma_n Z_n(x)\right) r_{n-1}(x)+\gamma_n W_n(x) && (3.10)\\
&= \left(1-\gamma_n f(x)\right) r_{n-1}(x)+\gamma_n\left(f(x)-Z_n(x)\right) r_{n-1}(x)+\gamma_n W_n(x). && (3.11)
\end{aligned}
$$

To establish the asymptotic behaviour of (r_n) and $(\bar{r}_n)$, 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 $n\geq n_0$,

$$
\rho_n(x) = \left(1-\gamma_n f(x)\right)\rho_{n-1}(x)+\gamma_n\left(f(x)-Z_n(x)\right) r(x)+\gamma_n W_n(x). \tag{3.12}
$$

We first give the asymptotic behaviour of (ρ_n) and of $(\bar{\rho}_n)$ in Section 3.3.1 and 3.3.2 respectively (and refer to Section 3.4 for the proof of the different lemmas). Then, we show in Section 3.3.3 how the asymptotic behaviour of (r_n) and $(\bar{r}_n)$ can be deduced from the one of (ρ_n) and $(\bar{\rho}_n)$ respectively.

3.3.1 Asymptotic behaviour of ρ_n

The aim of this section is to give the outlines of the proof of the three following lemmas.

Lemma 4 (Weak pointwise convergence rate of ρ_n)
Let Assumptions $(A1)-(A3)$ hold for $x\in\mathbb{R}$ such that $f(x)\neq 0$.

1. *If there exists $c\geq 0$ such that $\gamma_n^{-1}h_n^5\to c$, and if $\lim_{n\to\infty}(n\gamma_n)>(1-a)/(2f(x))$, then*

$$
\sqrt{\gamma_n^{-1}h_n}\left(\rho_n(x)-r(x)\right) \overset{\mathcal{D}}{\to}\mathcal{N}\left(\frac{\sqrt{c}f(x)m^{(2)}(x)}{f(x)-2a\xi},\frac{Var[Y|X=x]f(x)}{2f(x)-(\alpha-a)\xi}\int_{\mathbb{R}}K^2(z)dz\right). \tag{3.13}
$$

2. *If $\gamma_n^{-1}h_n^5\to\infty$, and if $\lim_{n\to\infty}(n\gamma_n)>2a/f(x)$, then*

$$
\frac{1}{h_n^2}\left(\rho_n(x)-r(x)\right)\overset{\mathbb{P}}{\to}\frac{f(x)m^{(2)}(x)}{(f(x)-2a\xi)}. \tag{3.14}
$$

Lemma 5 (Strong pointwise convergence rate of ρ_n)
Let Assumptions $(A1)-(A3)$ hold for $x\in\mathbb{R}$ such that $f(x)\neq 0$.

1. *If there exists $c\geq 0$ such that $\gamma_n^{-1}h_n^5/\ln(s_n)\to c$, and if $\lim_{n\to\infty}(n\gamma_n)>(1-a)/(2f(x))$, then, with probability one, the sequence*

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

is relatively compact and its limit set is the interval

$$
\left[\sqrt{\frac{c}{2}}\frac{f(x)m^{(2)}(x)}{f(x)-2a\xi}-\sqrt{\frac{Var[Y|X=x]f(x)\int_{\mathbb{R}}K^2(z)dz}{2f(x)-(\alpha-a)\xi}},\right.
\left.\sqrt{\frac{c}{2}}\frac{f(x)m^{(2)}(x)}{f(x)-2a\xi}+\sqrt{\frac{Var[Y|X=x]f(x)\int_{\mathbb{R}}K^2(z)dz}{2f(x)-(\alpha-a)\xi}}\right].
$$

2. *If* $\gamma_n^{-1}h_n^5/\ln(s_n)\to\infty$, *and if* $\lim_{n\to\infty}(n\gamma_n)>2a/f(x)$ *then, with probability one,*

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(\rho_n(x)-r(x)\right)=\frac{f(x)\,m^{(2)}(x)}{f(x)-2a\xi}.$$

Lemma 6 (Strong uniform convergence rate of ρ_n)
Let I be a bounded open interval on which $\varphi=\inf_{x\in I}f(x)>0$, and let Assumptions $(A1)-(A4)$ hold for all $x\in I$. If $\lim_{n\to\infty}n\gamma_n>\min\{(1-a)/(2\varphi),2a/\varphi\}$, then

$$\sup_{x\in I}|\rho_n(x)-r(x)|=O\left(\max\left\{\sqrt{\gamma_nh_n^{-1}}\ln n,h_n^2\right\}\right)\quad a.s.$$

To prove Lemmas 4 and 5, we first remark that, in view of (3.12), we have, for $n\geq n_0$,

$$\begin{aligned}
\rho_n(x)-r(x) &= (1-\gamma_nf(x))\left(\rho_{n-1}(x)-r(x)\right)+\gamma_n\left(W_n(x)-r(x)Z_n(x)\right)\\
&= \Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k\left(W_k(x)-r(x)Z_k(x)\right)\\
&\quad+\Pi_n(f(x))\left(\rho_{n_0-1}(x)-r(x)\right)\\
&= \tilde{T}_n(x)+\tilde{R}_n(x),
\end{aligned}\tag{3.15}$$

with, since $\rho_{n_0-1}=r_{n_0-1}$,

$$\begin{aligned}
\tilde{T}_n(x) &= \sum_{k=n_0}^{n}U_{k,n}(f(x))\gamma_k\left(W_k(x)-r(x)Z_k(x)\right),\\
\tilde{R}_n(x) &= \Pi_n(f(x))\left(r_{n_0-1}(x)-r(x)\right).
\end{aligned}$$

Noting that $|r_{n_0-1}(x)-r(x)|=O(1)$ a.s. and applying Lemma 3, we get

$$\begin{aligned}
\left|\tilde{R}_n(x)\right| &= O\left(\Pi_n(f(x))\right)\quad a.s.\\
&= o(m_n)\quad a.s.
\end{aligned}$$

Lemmas 4 and 5 are thus straightforward consequences of the following lemmas, which are proved in Sections 3.4.1 and 3.4.2, respectively.

Lemma 7 *The two parts of Lemma 4 hold when $\rho_n(x)-r(x)$ is replaced by $\tilde{T}_n(x)$.*

Lemma 8 *The two parts of Lemma 5 hold when $\rho_n(x)-r(x)$ is replaced by $\tilde{T}_n(x)$.*

In the same way, we remark that

$$\begin{aligned}
\sup_{x\in I}\left|\tilde{R}_n(x)\right| &= O\left(\sup_{x\in I}\Pi_n(f(x))\right)\quad a.s.\\
&= O\left(\Pi_n(\varphi)\right)\quad a.s.\\
&= o(m_n)\quad a.s.,
\end{aligned}$$

so that Lemma 6 is a straightforward consequence of the following lemma, which is proved in Section 3.4.3.

Lemma 9 *Lemma 6 holds when ρ_n-r is replaced by $\tilde{T}_n$.*

3.3.2 Asymptotic behaviour of $\bar{\rho}_n$

The purpose of this section is to give the outlines of the proof of the three following lemmas.

Lemma 10 (Weak pointwise convergence rate of $\bar{\rho}_n$)
Let Assumptions $(A1)-(A3)$, $(A5)$ and $(A6)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $nh_n^5 \to c$, then*

$$\sqrt{nh_n}\left(\bar{\rho}_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(c^{\frac{1}{2}}\frac{1-q}{1-q-2a}m^{(2)}(x), \frac{(1-q)^2}{1+a-2q}\frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}}K^2(z)\,dz\right).$$

2. *If $nh_n^5 \to \infty$, then*

$$h_n^{-2}\left(\bar{\rho}_n(x) - r(x)\right) \stackrel{\mathbb{P}}{\to} \frac{1-q}{1-q-2a}m^{(2)}(x).$$

Lemma 11 (Strong pointwise convergence rate of $\bar{\rho}_n$)
Let Assumptions $(A1)-(A3)$, $(A5)$ and $(A6)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c_1 \geq 0$ such that $nh_n^5/\ln\ln n \to c_1$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{nh_n}{2\ln\ln n}}\left(\bar{\rho}_n(x) - r(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\frac{1-q}{1-q-2a}\sqrt{\frac{c_1}{2}}m^{(2)}(x) - \sqrt{\frac{(1-q)^2}{1+a-2q}\frac{Var[Y/X=x]}{f(x)}\int_{\mathbb{R}}K^2(z)\,dz},\right.$$
$$\left.\frac{1-q}{1-q-2a}\sqrt{\frac{c_1}{2}}m^{(2)}(x) + \sqrt{\frac{(1-q)^2}{1+a-2q}\frac{Var[Y/X=x]}{f(x)}\int_{\mathbb{R}}K^2(z)\,dz}\right].$$

2. *If $nh_n^5/\ln\ln n \to \infty$, then*

$$\lim_{n\to\infty} h_n^{-2}\left(\bar{\rho}_n(x) - r(x)\right) = \frac{1-q}{1-q-2a}m^{(2)}(x) \quad a.s.$$

Lemma 12 (Strong uniform convergence rate of $\bar{\rho}_n$)
Let I be a bounded open interval on which $\varphi = \inf_{x\in I} f(x) > 0$ and let Assumptions $(A1)-(A6)$ hold for all $x \in I$. We have

$$\sup_{x\in I}\left|\bar{\rho}_n(x) - r(x)\right| = O\left(\max\left\{\sqrt{n^{-1}h_n^{-1}}\ln n, h_n^2\right\}\right) \quad a.s.$$

To prove Lemmas 10-12, we note that (3.12) gives, for $n \geq n_0$,

$$\rho_n(x) - \rho_{n-1}(x) = -\gamma_n f(x)\left[\rho_{n-1}(x) - r(x)\right] + \gamma_n\left[W_n(x) - r(x)Z_n(x)\right],$$

and thus

$$\rho_{n-1}(x) - r(x) = \frac{1}{f(x)}\left[W_n(x) - r(x)Z_n(x)\right] - \frac{1}{\gamma_n f(x)}\left[\rho_n(x) - \rho_{n-1}(x)\right].$$

It follows that

$$\begin{aligned}\bar{\rho}_n(x) - r(x) &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=1}^n q_k\left[\rho_k(x) - r(x)\right] \\ &= \frac{1}{f(x)}T_n(x) - \frac{1}{f(x)}R_n^{(0)}(x) \qquad (3.16)\end{aligned}$$

with

$$\begin{aligned}T_n(x) &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^n q_k\left[W_{k+1}(x) - r(x)Z_{k+1}(x)\right], \\ R_n^{(0)}(x) &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^n \frac{q_k}{\gamma_{k+1}}\left[\rho_{k+1}(x) - \rho_k(x)\right].\end{aligned}$$

Let us note that $R_n^{(0)}$ can be rewritten as

$$\begin{aligned}R_n^{(0)}(x) &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^n \frac{q_k}{\gamma_{k+1}}\left[\left(\rho_{k+1}(x) - r(x)\right) - \left(\rho_k(x) - r(x)\right)\right] \\ &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0}^n\left(\frac{q_{k-1}}{\gamma_k} - \frac{q_k}{\gamma_{k+1}}\right)\left(\rho_k(x) - r(x)\right) \\ &\quad + \frac{1}{\sum_{k=1}^n q_k}\frac{q_n}{\gamma_{n+1}}\left(\rho_{n+1}(x) - r(x)\right) - \frac{1}{\sum_{k=1}^n q_k}\frac{q_{n_0-1}}{\gamma_{n_0}}\left(\rho_{n_0-1}(x) - r(x)\right) \\ &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0}^n\frac{q_{k-1}}{\gamma_k}\left[1 - \frac{q_{k-1}^{-1}\gamma_k}{q_k^{-1}\gamma_{k+1}}\right]\left(\rho_k(x) - r(x)\right) \\ &\quad + \frac{1}{\sum_{k=1}^n q_k}\frac{q_n}{\gamma_{n+1}}\left(\rho_{n+1}(x) - r(x)\right) - \frac{1}{\sum_{k=1}^n q_k}\frac{q_{n_0-1}}{\gamma_{n_0}}\left(\rho_{n_0-1}(x) - r(x)\right).\end{aligned}$$

Since $\left(q_{k-1}^{-1}\gamma_k\right) \in \mathcal{GS}\left(q - \alpha\right)$, we have

$$\begin{aligned}\left[1 - \frac{q_{k-1}^{-1}\gamma_k}{q_k^{-1}\gamma_{k+1}}\right] &= 1 - \left(1 - \frac{(q-\alpha)}{k} + o\left(\frac{1}{k}\right)\right) \\ &= O\left(k^{-1}\right),\end{aligned}$$

and thus

$$\begin{aligned}\left|R_n^{(0)}(x)\right| &= O\left(\frac{1}{\sum_{k=1}^n q_k}\left[\sum_{k=n_0}^n k^{-1}q_{k-1}\gamma_k^{-1}\left|\rho_k(x) - r(x)\right| + \frac{q_n}{\gamma_{n+1}}\left|\rho_{n+1}(x) - r(x)\right|\right.\right. \\ &\quad \left.\left. + \frac{q_{n_0-1}}{\gamma_{n_0}}\left|\rho_{n_0-1}(x) - r(x)\right|\right]\right). \qquad (3.17)\end{aligned}$$

The application of Lemma 5 ensures that

$$\begin{aligned}\left|R_n^{(0)}(x)\right| &= O\left(\frac{1}{\sum_{k=1}^n q_k}\left[\sum_{k=n_0}^n\left(k^{-1}q_k\gamma_k^{-1}\right)\left(\left(\gamma_k h_k^{-1}\ln(s_k)\right)^{\frac{1}{2}} + h_k^2\right)\right.\right. \\ &\quad \left.\left. + \frac{q_n}{\gamma_{n+1}}\left(\left(\gamma_n h_n^{-1}\ln(s_n)\right)^{\frac{1}{2}} + h_n^2\right) + 1\right]\right) \quad a.s.\end{aligned}$$

Now, let us recall that, if $(u_n) \in \mathcal{GS}(-u^*)$ with $u^* < 1$, then we have, for any fixed $k_0 \geq 1$,

$$\lim_{n\to\infty} \frac{nu_n}{\sum_{k=k_0}^n u_k} = 1 - u^*, \tag{3.18}$$

and, if $u^* \geq 1$, then for all $\epsilon > 0$, $u_n = O\left(n^{-1+\epsilon}\right)$ and thus

$$\sum_{k=1}^n u_k = O\left(n^{\epsilon}\right). \tag{3.19}$$

Now, set $\epsilon \in]0, \min\{1-q-2a, (1+a)/2 - q\}[$ (the existence of such an ϵ being ensured by $(A6)$) ; in view of $(A5)$, we get

$$\begin{aligned}
\left|R_n^{(0)}(x)\right| &= O\left(\frac{1}{nq_n}\left(n^{\epsilon} + q_n\gamma_n^{-\frac{1}{2}}h_n^{-\frac{1}{2}}\left(\ln(s_n)\right)^{\frac{1}{2}} + q_n\gamma_n^{-1}h_n^2\right) + \frac{1}{n\gamma_n}\left(\left(\gamma_n h_n^{-1}\ln(s_n)\right)^{\frac{1}{2}} + h_n^2\right)\right) a.s. \\
&= O\left(\frac{n^{\epsilon}}{nq_n} + \frac{\sqrt{n^{-1}h_n^{-1}}}{\sqrt{n\gamma_n\left(\ln(s_n)\right)^{-1}}} + \frac{h_n^2}{n\gamma_n}\right) a.s. \\
&= o\left(\sqrt{n^{-1}h_n^{-1}} + h_n^2\right) a.s.
\end{aligned}$$

In view of (3.16), Lemmas 10 and 11 are thus straightforward consequences of the two following lemmas, which are proved in Sections 3.4.4 and 3.4.5 respectively.

Lemma 13 (Weak pointwise convergence rate of T_n)
The two parts of Lemma 10 hold when $\bar{\rho}_n(x) - r(x)$ is replaced by $[f(x)]^{-1}T_n(x)$.

Lemma 14 (Strong pointwise convergence rate of T_n)
The two parts of Lemma 11 hold when $\bar{\rho}_n(x) - r(x)$ is replaced by $[f(x)]^{-1}T_n(x)$.

Now, in view of (3.17), the application of Lemma 6 ensures that

$$\begin{aligned}
&\sup_{x\in I}\left|R_n^{(0)}(x)\right| \\
&= O\left(\frac{1}{\sum_{k=1}^n q_k}\left[\sum_{k=n_0}^n k^{-1}q_k\gamma_k^{-1}\sup_{x\in I}\left|\rho_k(x) - r(x)\right| + \frac{q_n}{\gamma_{n+1}}\sup_{x\in I}\left|\rho_{n+1}(x) - r(x)\right|\right]\right) \quad a.s. \\
&= O\left(\frac{1}{\sum_{k=1}^n q_k}\left[\sum_{k=n_0}^n \left(k^{-1}q_k\gamma_k^{-1}\right)\left(\left(\gamma_k h_k^{-1}\right)^{\frac{1}{2}}\ln k + h_k^2\right) + \frac{q_n}{\gamma_{n+1}}\left(\left(\gamma_n h_n^{-1}\right)^{\frac{1}{2}}\ln n + h_n^2\right)\right]\right) \quad a.s.
\end{aligned}$$

Setting $\epsilon \in]0, \min\{1-q-2a, (1+a)/2 - q\}[$ again, we get, in view of $(A5)$,

$$\begin{aligned}
\sup_{x\in I}\left|R_n^{(0)}(x)\right| &= O\left(\frac{1}{nq_n}\left(n^{\epsilon} + q_n\gamma_n^{-\frac{1}{2}}h_n^{-\frac{1}{2}}\ln n + q_n\gamma_n^{-1}h_n^2\right) + \frac{1}{n\gamma_n}\left(\left(\gamma_n h_n^{-1}\right)^{\frac{1}{2}}\ln n + h_n^2\right)\right) \quad a.s. \\
&= O\left(\frac{n^{\epsilon}}{nq_n} + \frac{\sqrt{n^{-1}h_n^{-1}}\ln n}{\sqrt{n\gamma_n}} + \frac{h_n^2}{n\gamma_n}\right) a.s. \\
&= o\left(\sqrt{n^{-1}h_n^{-1}}\ln n + h_n^2\right) \quad a.s.
\end{aligned}$$

In view of (3.16), Lemma 12 is thus a straightforward consequence of the following lemma, which is proved in Section 3.4.6.

Lemma 15 (Strong uniform convergence rate of T_n)
Lemma 12 holds when $\bar{\rho}_n - r$ is replaced by T_n .

3.3.3 How to deduce the asymptotic behaviour of r_n and $\bar{r}_n$ from the one of ρ_n and $\bar{\rho}_n$

Set

$$\Delta_n(x) = r_n(x) - \rho_n(x)$$

and

$$\bar{\Delta}_n(x) = \frac{1}{\sum_{k=1}^n q_k} \sum_{k=1}^n q_k \Delta_k(x) = \bar{r}_n(x) - \bar{\rho}_n(x).$$

To deduce the asymptotic behaviour of r_n (respectively $\bar{r}_n$) from the one of ρ_n (respectively $\bar{\rho}_n$), we prove that Δ_n (respectively $\bar{\Delta}_n$) is negligible in front of ρ_n (respectively $\bar{\rho}_n$). Note that, in view of (3.11) and (3.12), and since $\rho_{n_0-1}(x) = r_{n_0-1}(x)$, we have, for $n \geq n_0$,

$$\begin{aligned} \Delta_n(x) &= (1-\gamma_n f(x))\,\Delta_{n-1}(x) + \gamma_n (f(x) - Z_n(x))\,(r_{n-1}(x) - r(x)) \\ &= \sum_{k=n_0}^n U_{k,n}(f(x))\,(f(x) - Z_k(x))\,(r_{k-1}(x) - r(x)). \end{aligned} \tag{3.20}$$

The difficulty which appears here is that Δ_n is expressed in function of the terms $r_k - r$, so that an upper bound of $r_n - r$ is necessary for the obtention of an upper bound of Δ_n. Now, the key to overcome this difficulty is the following property $(\mathcal{P})$: if $(r_n - r)$ is known to be bounded almost surely by a sequence (w_n), then it can be shown that (Δ_n) is bounded almost surely by a sequence (w'_n) such that $\lim_{n\to\infty} w'_n w_n^{-1} = 0$, which may allow to upper bound $r_n - r$ by a sequence smaller than (w_n). To deduce the asymptotic behaviour of r_n (respectively $\bar{r}_n$) from the one of ρ_n (respectively $\bar{\rho}_n$), we thus proceed as follows. We first establish a rudimentary upper bound of $(r_n - r)$. Then, applying Property $(\mathcal{P})$ several times, we successively improve our upper bound of $(r_n - r)$, and this until we obtain an upper bound, which allows to prove that Δ_n (respectively $\bar{\Delta}_n$) is negligible in front of ρ_n (respectively $\bar{\rho}_n$).

We first establish the pointwise results on r_n and $\bar{\rho}_n$ (that is, Theorems 4, 5, 7, and 8) in Section 3.3.3.a, and then the uniform ones (that is, Theorems 6 and 9) in Section 3.3.3.b.

3.3.3.a Proof of Theorems 4, 5, 7 and 8

The proof of Theorems 4, 5, 7 and 8 relies on the repeted application of the following lemma, which is proved in Section 3.4.7.

Lemma 16 *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.*

1. *If the sequence $(n\gamma_n)$ is bounded, if $\lim_{n\to\infty} n\gamma_n > \min\{(1-a)/(2f(x)), 2a/f(x)\}$, and if $w^* \geq 0$, then, for all $\delta > 0$,*

$$|\Delta_n(x)| = O\left(m_n w_n (\ln n)^{\frac{(1+\delta)}{2}}\right) + o(m_n) \quad a.s.$$

2. *If $\lim_{n\to\infty}(n\gamma_n) = \infty$, then, for all $\delta > 0$,*

$$|\Delta_n(x)| = O\left(m_n w_n \left(n^{1+\delta}\gamma_n\right)^{\frac{(1+\delta)}{2}}\right) \quad a.s.$$

We first establish a preliminary upper bound for $r_n(x) - r(x)$. Then, we successively prove Theorems 4 and 5 in the case $(n\gamma_n)$ is bounded, Theorems 4 and 5 in the case $\lim_{n\to\infty}(n\gamma_n) = \infty$, and finally Theorems 7 and 8.

Preliminary 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 (3.10) that, for $n \geq n_0$,

$$\begin{aligned} |r_n(x)| &\leq |r_{n-1}(x)| + \gamma_n |Y_n| h_n^{-1} \|K\|_\infty \\ &\leq |r_{n_0-1}(x)| + \left(\sup_{k\leq n} |Y_k|\right) \|K\|_\infty \sum_{k=1}^{n} \gamma_k h_k^{-1}. \end{aligned} \tag{3.21}$$

Since

$$\mathbb{P}\left(\sup_{k\leq n} |Y_k| > n^2\right) \leq n\mathbb{P}\left(|Y| > n^2\right) \leq n^{-3}\mathbb{E}\left(|Y|^2\right),$$

we have $\sup_{k\leq n}|Y_k| \leq n^2$ a.s. Moreover, since $\left(\gamma_n h_n^{-1}\right) \in \mathcal{GS}(-\alpha + a)$ with $1 - \alpha + a > 0$, we note that $\sum_{k=1}^{n} \gamma_k h_k^{-1} = O\left(n\gamma_n h_n^{-1}\right)$. We thus deduce that

$$|r_n(x) - r(x)| = O\left(n^3 \gamma_n h_n^{-1}\right) \quad a.s. \tag{3.22}$$

Proof of Theorems 4 and 5 in the case the sequence $(n\gamma_n)$ is bounded.
In this case, $\alpha = 1$, and Lemmas 5 and 16 imply that :

- $|\rho_n(x) - r(x)| = O(m_n \ln n) \quad a.s.$ (3.23)
- If there exists $(w_n) \in \mathcal{GS}(w^*)$, $w^* \geq 0$, such that $|r_n(x) - r(x)| = O(w_n)$ a.s., then $|\Delta_n(x)| = O(m_n w_n \ln n) + o(m_n) \quad a.s.$ (3.24)

Set $p_0 = \max\{p \text{ such that } - m^* p + 2 + a \geq 0\}$, set $j \in \{0, 1, \ldots, p_0 - 1\}$, and assume that

$$|r_n(x) - r(x)| = O\left(m_n^j \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^j\right) \quad a.s. \tag{3.25}$$

Since the sequence $(w_n) = \left(m_n^j \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^j\right)$ belongs to $\mathcal{GS}(-m^* j + 2 + a)$ with $-m^* j + 2 + a > 0$, the application of (3.24) implies that

$$|\Delta_n(x)| = O\left(m_n^{j+1} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{j+1}\right) + o(m_n) \quad a.s.$$

Since $\left(m_n^{j+1} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{j+1}\right) \in \mathcal{GS}(-m^*(j+1) + 2 + a)$ with $-m^*(j+1) + 2 + a \geq 0$, whereas $(m_n) \in \mathcal{GS}(-m^*)$ with $-m^* < 0$, it follows that

$$|\Delta_n(x)| = O\left(m_n^{j+1} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{j+1}\right) \quad a.s.,$$

and the application of (3.23) leads to

$$\begin{aligned} |r_n(x) - r(x)| &\leq |\rho_n(x) - r(x)| + |\Delta_n(x)| \\ &= O\left(m_n^{j+1} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{j+1}\right) \quad a.s. \end{aligned}$$

Since (3.22) ensures that (3.25) is satisfied for $j = 0$, we have proved by induction that

$$|r_n(x) - r(x)| = O\left(m_n^{p_0} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{p_0}\right) \quad a.s.$$

Applying (3.24) with $(w_n) = \left(m_n^{p_0} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{p_0}\right)$ and then (3.23), we obtain

$$|r_n(x) - r(x)| = O\left(m_n^{p_0+1} \left(n^3 \gamma_n h_n^{-1}\right) (\ln n)^{p_0+1}\right) + O(m_n \ln n) \quad a.s.$$

Since the sequences $\left(m_n^{p_0+1}\left(n^3\gamma_n h_n^{-1}\right)(\ln n)^{p_0+1}\right)$ and $(m_n \ln n)$ are in $\mathcal{GS}\left(-m^*\left(p_0+1\right)+2+a\right)$ with $-m^*\left(p_0+1\right)+2+a<0$, and $\mathcal{GS}\left(-m^*\right)$ with $-m^*<0$ respectively, it follows that

$$|r_n(x)-r(x)| = O\left((\ln n)^{-2}\right) \quad a.s.$$

Applying once more (3.24) with $(w_n)=\left((\ln n)^{-2}\right)\in\mathcal{GS}(0)$, we get

$$|\Delta_n(x)| = O\left(m_n(\ln n)^{-1}\right)+o(m_n) = o(m_n) \quad a.s.$$

Theorem 4 (respectively Theorem 5) in the case $(n\gamma_n)$ is bounded then follows from the application of Lemma 4 (respectively Lemma 5).

Proof of Theorems 4 and 5 in the case $\lim_{n\to\infty}(n\gamma_n)=\infty$
In this case, Lemmas 5 and 16 imply that, for all $\delta>0$,

- $|\rho_n(x)-r(x)| = O(\tilde{m}_n) \quad a.s.$ (3.26)
- If there exists $(w_n)\in\mathcal{GS}(w^*)$ such that $|r_n(x)-r(x)|=O(w_n)\,a.s.$, then $|\Delta_n(x)| = O\left(m_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}w_n\right) \quad a.s.$ (3.27)

Now, set $\delta>0$ such that $c(\delta)=-m^*+(1+\delta)(1+\delta-\alpha)/2<0$ (the existence of such a δ being ensured by $(A2)$). In view of (3.22), the application of (3.27) with $(w_n)=\left(n^3\gamma_n h_n^{-1}\right)$ ensures that

$$|\Delta_n(x)| \quad = \quad O\left(m_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}} n^3\gamma_n h_n^{-1}\right) \quad a.s.,$$

and, in view of (3.26), it follows that

$$|r_n(x)-r(x)| \quad = \quad O(\tilde{m}_n)+O\left(m_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}} n^3\gamma_n h_n^{-1}\right) \quad a.s.$$

Set $p\geq 1$, and assume that

$$|r_n(x)-r(x)| \quad = \quad O(\tilde{m}_n)+O\left(m_n^p\left(n^{1+\delta}\gamma_n\right)^{p\left(\frac{1+\delta}{2}\right)} n^3\gamma_n h_n^{-1}\right) \quad a.s.$$

The application of (3.27) with $(w_n)=(\tilde{m}_n)$ and with $(w_n)=\left(m_n^p\left(n^{1+\delta}\gamma_n\right)^{p\left(\frac{1+\delta}{2}\right)} n^3\gamma_n h_n^{-1}\right)$ ensures that

$$|\Delta_n(x)| \quad = \quad O\left(m_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\tilde{m}_n\right)+O\left(m_n^{p+1}\left(n^{1+\delta}\gamma_n\right)^{(p+1)\frac{1+\delta}{2}} n^3\gamma_n h_n^{-1}\right) \quad a.s.$$

The sequence $\left(m_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right)$ being in $\mathcal{GS}(c(\delta))$ with $c(\delta)<0$, it follows that

$$|\Delta_n(x)| \quad = \quad o(\tilde{m}_n)+O\left(m_n^{p+1}\left(n^{1+\delta}\gamma_n\right)^{(p+1)\frac{1+\delta}{2}} n^3\gamma_n h_n^{-1}\right) \quad a.s.$$

and, in view of (3.26), we obtain

$$|r_n(x)-r(x)| \quad = \quad O(\tilde{m}_n)+O\left(m_n^{p+1}\left(n^{1+\delta}\gamma_n\right)^{(p+1)\frac{1+\delta}{2}} n^3\gamma_n h_n^{-1}\right) \quad a.s.$$

We have thus proved by induction that, for all $p \geq 1$,

$$|r_n(x) - r(x)| = O(\tilde{m}_n) + O\left(m_n^p \left(n^{1+\delta}\gamma_n\right)^{p\left(\frac{1+\delta}{2}\right)} n^3\gamma_n h_n^{-1}\right) \quad a.s.$$

By setting p large enough, we deduce that

$$|r_n(x) - r(x)| = O(\tilde{m}_n) \quad a.s.$$

Applying once more (3.27) with $(w_n) = (\tilde{m}_n)$, we get

$$\begin{aligned} |\Delta_n(x)| &= O\left(m_n \left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}} \tilde{m}_n\right) \quad a.s. \\ &= o(m_n) \quad a.s. \end{aligned} \tag{3.28}$$

Theorem 4 (respectively Theorem 5) in the case $\lim_{n\to\infty}(n\gamma_n) = \infty$ then follows from the application of Lemma 4 (respectively Lemma 5).

Proof of Theorems 7 and 8

In view of (3.28), and applying (3.18) and (3.19), we get, for all $\delta > 0$,

$$\begin{aligned} \left|\bar{\Delta}_n(x)\right| &= O\left(\frac{1}{\sum_{k=1}^n q_k}\sum_{k=1}^n q_k\tilde{m}_k^2\left(k^{1+\delta}\gamma_k\right)^{\frac{1+\delta}{2}}\right) \quad a.s. \\ &= O\left(\frac{1}{nq_n}\left[n^\delta + nq_n\tilde{m}_n^2\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right]\right) \quad a.s. \\ &= O\left(n^{\delta-1}q_n^{-1} + \tilde{m}_n^2\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) \quad a.s. \end{aligned} \tag{3.29}$$

• Let us first consider the case when the sequence $\left(nh_n^5\right)$ is bounded. In this case, we have $a \geq 1/5$, so that $a \geq \alpha/5$ and $m^* = (\alpha - a)/2$. Noting that $(A2)$ implies $a < 3\alpha - 2$, and applying $(A6)$, we can set $\delta > 0$ such that

$$\delta - \frac{(1+a)}{2} + q < 0 \text{ and } \frac{(1-a)}{2} - 2m^* + \frac{(1+\delta)}{2}(1+\delta-\alpha) < 0.$$

In view of (3.29), we then obtain

$$\sqrt{nh_n}\left|\bar{\Delta}_n(x)\right| = o(1) \quad a.s.$$

The first part of Theorem 7 (respectively of Theorem 8) then follows from the application of the first part of Lemma 10 (respectively of Lemma 11).

• Let us now consider the case when $\lim_{n\to\infty}\left(nh_n^5\right) = \infty$. Noting that $(A2)$ then ensures that $6a < 3\alpha - 1$, and applying $(A6)$, we can set $\delta > 0$ such that

$$2a + \delta - 1 + q < 0 \text{ and } 2a - 2m^* + \frac{(1+\delta)}{2}(1+\delta-\alpha) < 0.$$

It then follows from (3.29) that

$$h_n^{-2}\left|\bar{\Delta}_n(x)\right| = o(1) \quad a.s.$$

The second part of Theorem 7 (respectively of Theorem 8) then follows from the application of the second part of Lemma 10 (respectively of Lemma 11).

3.3.3.b Proof of Theorems 6 and 9

Set

$$B_n = n\gamma_n h_n^{-1} \ln n. \tag{3.30}$$

The proof of Theorems 6 and 9 relies on the repeted application of the following lemma, which is proved in Section 3.4.8.

Lemma 17 *Let I be a bounded open interval on which $\varphi = \inf_{x\in I} f(x) > 0$, let Assumptions $(A1)-(A4)$ hold for all $x \in I$, and assume that there exists $(w_n) \in \mathcal{GS}(w^*)$ such that $\sup_{x\in I} |r_n(x) - r(x)| = O(w_n)$ a.s. Moreover,*
- *in the case when $(n\gamma_n)$ is bounded, assume that $\lim_{n\to\infty} (n\gamma_n) > m^*/\varphi$ and that $w^* \geq 0$;*
- *in the case when $\lim_{n\to\infty} (n\gamma_n) = \infty$, assume that the sequence $\left(w_n^{-1} B_n \sqrt{\gamma_n h_n^{-1} \ln n}\right)$ is bounded.*

Then, we have

$$\sup_{x\in I} |\Delta_n(x)| = O\left(m_n w_n \sqrt{\ln n}\right) \quad a.s.$$

We first establish a preliminary upper bound of $\sup_{x\in I}(|r_n(x) - r(x)|)$ (which is better than the pointwise upper bound (3.22) since the random variables Y_k are assumed to have a finite exponential moment in Theorems 6 and 9). Then, we successively prove Theorem 6 in the case when $(n\gamma_n)$ is bounded, Theorem 6 in the case when $\lim_{n\to\infty}(n\gamma_n) = \infty$, and finally Theorem 9.

Preliminary upper bound.
Proceeding as for the proof of (3.22), we note that, for all $n \geq n_0$,

$$\sup_{x\in I} |r_n(x)| \leq \sup_{x\in I} |r_{n_0-1}(x)| + \left(\sup_{k\leq n} |Y_k|\right) \|K\|_\infty \sum_{k=1}^{n} \gamma_k h_k^{-1},$$

with, this time, in view of $(A4)$,

$$\mathbb{P}\left[\sup_{k\leq n} |Y_k| > \frac{3}{t^*} \ln n\right] \leq n\mathbb{P}\left[\exp(t^*|Y|) > n^3\right] \leq n^{-2}\mathbb{E}\left(\exp(t^*|Y|)\right).$$

We deduce that

$$\sup_{x\in I} |r_n(x) - r(x)| = O(B_n) \quad a.s.$$

Proof of Theorem 6 in the case $(n\gamma_n)$ is bounded
In this case, we have $\alpha = 1$, $(B_n) \in \mathcal{GS}(a)$ (with $a > 0$); the application of Lemma 17 with $(w_n) = (B_n)$ ensures that

$$\sup_{x\in I} |\Delta_n(x)| = O\left(m_n B_n \sqrt{\ln n}\right) \quad a.s.$$

Applying Lemma 6, we get

$$\begin{aligned}\sup_{x\in I} |r_n(x) - r(x)| &\leq \sup_{x\in I} |\rho_n(x) - r(x)| + \sup_{x\in I} |\Delta_n(x)| \\ &= O\left(m_n B_n \sqrt{\ln n}\right) \quad a.s.\end{aligned}$$

Since $\left(m_n B_n \sqrt{\ln n}\right) \in \mathcal{GS}\left(-m^* + a\right)$ (with $-m^* + a < 0$), it follows that

$$\sup_{x\in I} |r_n(x) - r(x)| \quad = \quad O\left([\ln n]^{-1}\right) \quad a.s.$$

Applying once more Lemma 17 with $(w_n) = \left([\ln n]^{-1}\right)$, we get

$$\begin{aligned}\sup_{x\in I} |\Delta_n(x)| &= O\left(m_n (\ln n)^{-\frac{1}{2}}\right) \quad a.s.\\ &= o(\tilde{m}_n) \quad a.s.\end{aligned}$$

Theorem 6 in the case when $(n\gamma_n)$ is bounded then follows from the application of Lemma 6.

Proof of Theorem 6 in the case $\lim_{n\to\infty}(n\gamma_n) = \infty$
The sequence $\left(\sqrt{\gamma_n h_n^{-1} \ln n}\right)$ being clearly bounded, we can apply Lemma 17 with $(w_n) = (B_n)$; we then obtain

$$\sup_{x\in I} |\Delta_n(x)| = O\left(m_n B_n \sqrt{\ln n}\right) \quad a.s.$$

The application of Lemma 6 then ensures that

$$\begin{aligned}\sup_{x\in I} |r_n(x) - r(x)| &= O(\tilde{m}_n) + O\left(m_n B_n \sqrt{\ln n}\right) \quad a.s.\\ &= O\left(m_n B_n \sqrt{\ln n}\right) \quad a.s.\end{aligned}$$

Since $\left(m_n B_n \sqrt{\ln n}\right)^{-1} B_n \sqrt{\gamma_n h_n^{-1} \ln n} = m_n^{-1} \sqrt{\gamma_n h_n^{-1}} = O(1)$, we can apply once more Lemma 17 with $(w_n) = \left(m_n B_n \sqrt{\ln n}\right)$; we get

$$\sup_{x\in I} |\Delta_n(x)| \quad = \quad O\left(m_n^2 B_n \ln n\right) \quad a.s. \tag{3.31}$$

Noting that $(m_n B_n \ln n) \in \mathcal{GS}(-m^* + 1 - \alpha + a)$ with, in view of $(A4)\, iii)$, $-m^* + 1 - \alpha + a < 0$, it follows that

$$\sup_{x\in I} |\Delta_n(x)| \quad = \quad o(m_n) \quad a.s.$$

Theorem 6 in the case when $\lim_{n\to\infty}(n\gamma_n) = \infty$ then follows from the application of Lemma 6.

Proof of Theorem 9

– In the case when the sequence $\left(nh_n^5/\ln n\right)$ is bounded, we have, in view of (3.31),

$$\sqrt{nh_n}\,(\ln n)^{-1} \sup_{x\in I} |\Delta_n(x)| = O\left(\sqrt{nh_n}\, m_n^2 B_n\right) \quad a.s.$$

Now, in this case, we have $a \geq 1/5 \geq \alpha/5$ and thus $m^* = (\alpha - a)/2$. It follows that $\left(\sqrt{nh_n}\, m_n^2 B_n\right) \in \mathcal{GS}(3(1+a)/2 - 2\alpha)$ with $3(1+a)/2 - 2\alpha < 0$, and thus

$$\sup_{x\in I} |\Delta_n(x)| = O\left(\sqrt{n^{-1} h_n^{-1}}\, \ln n\right) \quad a.s.$$

The first part of Theorem 9 then follows from the application of Lemma 12.

– In the case when $\lim_{n\to\infty}\left(nh_n^5/\ln n\right)=\infty$, (3.31) ensures that

$$h_n^{-2}\sup_{x\in I}|\Delta_n(x)|=O\left(h_n^{-2}m_n^2B_n\ln n\right)\quad a.s.,$$

with $\left(h_n^{-2}m_n^2B_n\ln n\right)\in\mathcal{GS}(3a-2m^*+1-\alpha)$. Noting that the assumptions of Theorem 9 ensure that $3a-2m^*+1-\alpha<0$, we deduce that

$$\sup_{x\in I}|\Delta_n(x)|=O\left(h_n^2\right)\quad a.s.$$

The second part of Theorem 9 then follows from the application of Lemma 12.

3.4 Proof of Lemmas

3.4.1 Proof of Lemma 7

We establish that, under the condition $\lim_{n\to\infty}(n\gamma_n)>(\alpha-a)/(2f(x))$,

- if $a\geq\alpha/5$, then
$$\sqrt{\gamma_n^{-1}h_n}\left(\tilde{T}_n(x)-\mathbb{E}\left(\tilde{T}_n(x)\right)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\frac{Var[Y|X=x]f(x)}{2f(x)-(\alpha-a)\xi}\int_{\mathbb{R}}K^2(z)\,dz\right),\tag{3.32}$$
- if $a>\alpha/5$, then $\sqrt{\gamma_n^{-1}h_n}\mathbb{E}\left(\tilde{T}_n(x)\right)\to 0,$ (3.33)

and prove that, under the condition $\lim_{n\to\infty}(n\gamma_n)>2a/f(x)$,

- if $a\leq\alpha/5$, then $h_n^{-2}\mathbb{E}\left(\tilde{T}_n(x)\right)\to\dfrac{f(x)\,m^{(2)}(x)}{f(x)-2a\xi}$, (3.34)
- if $a<\alpha/5$, then $h_n^{-2}\left(\tilde{T}_n(x)-\mathbb{E}\left(\tilde{T}_n(x)\right)\right)\xrightarrow{\mathbb{P}}0.$ (3.35)

As a matter of fact the combination of (3.32) and (3.33) (respectively of (3.32) and (3.34)) gives Part 1 of Lemma 7 in the case $a>\alpha/5$ (respectively $a=\alpha/5$), the one of (3.34) and (3.35) (respectively of (3.32) and (3.34)) gives Part 2 of Lemma 7 in the case $a<\alpha/5$ (respectively $a=\alpha/5$). We prove (3.32), (3.35), (3.34), and (3.33) successively.

Proof of (3.32) Set

$$\tilde{\eta}_k(x)=\Pi_k^{-1}(f(x))\,\gamma_k\left[W_k(x)-r(x)\,Z_k(x)\right],\tag{3.36}$$

so that $\tilde{T}_n(x)-\mathbb{E}\left(\tilde{T}_n(x)\right)=\Pi_n(f(x))\sum_{k=n_0}^n\left[\tilde{\eta}_k(x)-\mathbb{E}(\tilde{\eta}_k(x))\right]$. We have

$$Var(\tilde{\eta}_k(x))\;=\;\Pi_k^{-2}(f(x))\,\gamma_k^2\left[Var(W_k(x))+r^2(x)\,Var(Z_k(x))-2r(x)\,Cov(W_k(x),Z_k(x))\right].$$

In view of $(A3)$, classical computations give

$$Var(W_k(x))\;=\;\frac{1}{h_k}\left[\mathbb{E}\left[Y^2|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz+o(1)\right],\tag{3.37}$$

$$Var(Z_k(x))\;=\;\frac{1}{h_k}\left[f(x)\int_{\mathbb{R}}K^2(z)\,dz+o(1)\right],\tag{3.38}$$

$$Cov(W_k(x),Z_k(x))\;=\;\frac{1}{h_k}\left[r(x)\,f(x)\int_{\mathbb{R}}K^2(z)\,dz+o(1)\right].\tag{3.39}$$

It follows that

$$Var(\tilde{\eta}_k(x)) = \frac{\Pi_k^{-2}(f(x))\gamma_k^2}{h_k}\left[Var[Y|X=x]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right], \qquad (3.40)$$

and, since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/(2f(x))$, Lemma 3 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=n_0}^{n} Var(\tilde{\eta}_k(x)) \\ &= \sum_{k=n_0}^{n} \frac{\Pi_k^{-2}(f(x))\gamma_k^2}{h_k}\left[Var[Y|X=x]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right] \\ &= \frac{1}{\Pi_n^2(f(x))}\frac{\gamma_n}{h_n}\frac{1}{2f(x)-(\alpha-a)\xi}\left[Var[Y|X=x]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right]. \qquad (3.41) \end{aligned}$$

For all $p \in]0,1]$ and in view of $(A3)$, we have

$$\begin{aligned} &\mathbb{E}\left(|Y_k - r(x)|^{2+p} K^{2+p}\left(\frac{x - X_k}{h_k}\right)\right) \\ &= h_k \int_{\mathbb{R}^2} |y - r(x)|^{2+p} K^{2+p}(s)\, g(x - h_k s, y)\, dyds \\ &\leq 2^{1+p} h_k \int_{\mathbb{R}}\left\{\int_{\mathbb{R}} |y|^{2+p} g(x - h_k s, y)\, dy + |r(x)|^{2+p}\int_{\mathbb{R}} g(x - h_k s, y)\, dy\right\} K^{2+p}(s)\, ds \\ &= O(h_k). \qquad (3.42) \end{aligned}$$

Now, set $p \in]0,1]$ such that $\lim_{n\to\infty}(n\gamma_n) > (1+p)(\alpha - a)/((2+p)f(x))$. Applying Lemma 3, we get

$$\begin{aligned} \sum_{k=n_0}^{n} \mathbb{E}\left[|\tilde{\eta}_k(x)|^{2+p}\right] &= O\left(\sum_{k=n_0}^{n} \frac{\Pi_k^{-2-p}(f(x))\gamma_k^{2+p}}{h_k^{2+p}}\mathbb{E}\left(|Y_k - r(x)|^{2+p} K^{2+p}\left(\frac{x - X_k}{h_k}\right)\right)\right) \\ &= O\left(\sum_{k=n_0}^{n} \frac{\Pi_k^{-2-p}(f(x))\gamma_k^{2+p}}{h_k^{1+p}}\right) \\ &= O\left(\frac{1}{\Pi_n^{2+p}(f(x))}\frac{\gamma_n^{1+p}}{h_n^{1+p}}\right). \qquad (3.43) \end{aligned}$$

Using (3.41), we deduce that

$$\begin{aligned} \frac{1}{v_n^{2+p}}\sum_{k=n_0}^{n} \mathbb{E}\left[|\tilde{\eta}_k(x)|^{2+p}\right] &= O\left(\left(\frac{\gamma_n}{h_n}\right)^{\frac{p}{2}}\right) \\ &= o(1), \end{aligned}$$

and (3.32) follows by application of Lyapounov Theorem.

Proof of (3.35) In view of (3.40), and since $a<\alpha/5$ and $\lim_{n\to\infty}(n\gamma_n)>2a/f(x)$, the application of Lemma 3 ensures that

$$\begin{aligned} Var\left(\tilde{T}_n(x)\right) &= \Pi_n^2(f(x))\sum_{k=n_0}^{n}\frac{\Pi_k^{-2}(f(x))\gamma_k^2}{h_k}\left[Var[Y|X=x]f(x)\int_{\mathbb{R}}K^2(z)\,dz+o(1)\right] \\ &= \Pi_n^2(f(x))\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k o\left(h_k^4\right) \\ &= o\left(h_n^4\right), \end{aligned}$$

which gives (3.35).

Proof of (3.34) We have

$$\mathbb{E}\left(\tilde{T}_n(x)\right) = \Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k\left[(\mathbb{E}(W_k(x))-a(x))-r(x)(\mathbb{E}(Z_k(x))-f(x))\right].$$

In view of $(A3)$ we obtain

$$\mathbb{E}(W_k(x))-a(x) = \frac{1}{2}h_k^2\int_{\mathbb{R}}y\frac{\partial^2 g}{\partial x^2}(x,y)\,dy\,[1+o(1)]\int_{\mathbb{R}}z^2K(z)\,dz, \tag{3.44}$$

$$\mathbb{E}(Z_k(x))-f(x) = \frac{1}{2}h_k^2\int_{\mathbb{R}}\frac{\partial^2 g}{\partial x^2}(x,y)\,dy\,[1+o(1)]\int_{\mathbb{R}}z^2K(z)\,dz. \tag{3.45}$$

Since $\lim_{n\to\infty}(n\gamma_n)>2a/f(x)$, it follows from the application of Lemma 3 that

$$\begin{aligned} \mathbb{E}\left(\tilde{T}_n(x)\right) &= \Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k h_k^2 \\ &\quad \left[\frac{1}{2}\left(\int_{\mathbb{R}}y\frac{\partial^2 g}{\partial x^2}(x,y)\,dy-r(x)\int_{\mathbb{R}}\frac{\partial^2 g}{\partial x^2}(x,y)\,dy\right)+o(1)\right]\int_{\mathbb{R}}z^2K(z)\,dz \\ &= \Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k h_k^2\left[m^{(2)}(x)f(x)+o(1)\right] \\ &= \frac{1}{f(x)-2a\xi}h_n^2\left[m^{(2)}(x)f(x)+o(1)\right], \end{aligned}$$

which gives (3.34).

Proof of (3.33) Since $a>\alpha/5$ and $\lim_{n\to\infty}(n\gamma_n)>(1-a)/(2f(x))$, we have

$$\begin{aligned} \mathbb{E}\left(\tilde{T}_n(x)\right) &= \Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k o\left(\sqrt{\gamma_k h_k^{-1}}\right) \\ &= o\left(\sqrt{\gamma_n h_n^{-1}}\right), \end{aligned}$$

which gives (3.33).

3.4.2 Proof of Lemma 8

Set

$$S_n(x) = \sum_{k=n_0}^{n} \left[\tilde{\eta_k}(x) - \mathbb{E}(\tilde{\eta_k}(x))\right].$$

where $\tilde{\eta_k}$ is defined in (3.36).

• Let us first consider the case $a \geq \alpha/5$ (in which case $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/(2f(x))$). We set $H_n^2(f(x)) = \Pi_n^2(f(x))\gamma_n^{-1}h_n$, and note that, since $\left(\gamma_n^{-1}h_n\right) \in \mathcal{GS}(\alpha - a)$, we have

$$\begin{aligned}
\ln\left(H_n^{-2}(f(x))\right) &= -2\ln(\Pi_n(f(x))) + \ln\left(\prod_{k=n_0}^{n} \frac{\gamma_{k-1}^{-1}h_{k-1}}{\gamma_k^{-1}h_k}\right) + \ln\left(\gamma_{n_0-1}h_{n_0-1}^{-1}\right) \\
&= -2\sum_{k=n_0}^{n}\ln(1 - f(x)\gamma_k) + \sum_{k=n_0}^{n}\ln\left(1 - \frac{\alpha - a}{k} + o\left(\frac{1}{k}\right)\right) + \ln\left(\gamma_{n_0-1}h_{n_0-1}^{-1}\right) \\
&= \sum_{k=n_0}^{n}(2f(x)\gamma_k + o(\gamma_k)) - \sum_{k=n_0}^{n}((\alpha - a)\xi\gamma_k + o(\gamma_k)) + \ln\left(\gamma_{n_0-1}h_{n_0-1}^{-1}\right) \\
&= (2f(x) - \xi(\alpha - a))s_n + o(s_n).
\end{aligned} \tag{3.46}$$

Since $2f(x) - \xi(\alpha - a) > 0$, it follows in particular that $\lim_{n\to\infty} H_n^{-2}(f(x)) = \infty$. Moreover, we clearly have $\lim_{n\to\infty} H_n^2(f(x))/H_{n-1}^2(f(x)) = 1$, and by (3.41),

$$\lim_{n\to\infty} H_n^2(f(x))\sum_{k=n_0}^{n} Var\left[\tilde{\eta_k}(x)\right] = \left[2f(x) - (\alpha - a)\xi\right]^{-1} Var\left[Y|X = x\right] f(x)\int_{\mathbb{R}} K^2(z)\,dz,$$

and, in view of (3.42), $\mathbb{E}\left[|\tilde{\eta_n}(x)|^3\right] = O\left(\Pi_n^{-3}(f(x))\gamma_n^3h_n^{-2}\right)$. Now, since $\left(\gamma_n^{-1}h_n\right) \in \mathcal{GS}(\alpha - a)$, applying Lemma 3 and (3.46), we get

$$\begin{aligned}
\frac{1}{n\sqrt{n}}\sum_{k=n_0}^{n}\mathbb{E}\left(|H_n(f(x))\tilde{\eta_k}(x)|^3\right) &= O\left(\frac{H_n^3(f(x))}{n\sqrt{n}}\left(\sum_{k=n_0}^{n}\frac{\Pi_k^{-3}(f(x))\gamma_k^3}{h_k^2}\right)\right) \\
&= O\left(\frac{\Pi_n^3(f(x))\gamma_n^{-\frac{3}{2}}h_n^{\frac{3}{2}}}{n\sqrt{n}}\left(\sum_{k=n_0}^{n}\Pi_k^{-3}(f(x))\gamma_k o\left(\left(\gamma_k h_k^{-1}\right)^{\frac{3}{2}}\right)\right)\right) \\
&= o\left(\frac{1}{n\sqrt{n}}\right) \\
&= o\left(\left[\ln\left(H_n^{-2}(f(x))\right)\right]^{-1}\right).
\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier (2006) then ensures that, with probability one, the sequence

$$\left(\frac{H_n(f(x))S_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}(f(x))\right)}}\right) = \left(\frac{\sqrt{\gamma_n^{-1}h_n}\left(\tilde{T}_n(x) - \mathbb{E}\left(\tilde{T}_n(x)\right)\right)}{\sqrt{2\ln\ln\left(H_n^{-2}(f(x))\right)}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\frac{Var\left[Y|X = x\right]f(x)}{2f(x) - (\alpha - a)\xi}\int_{\mathbb{R}}K^2(z)\,dz}, \sqrt{\frac{Var\left[Y|X = x\right]f(x)}{2f(x) - (\alpha - a)\xi}\int_{\mathbb{R}}K^2(z)\,dz}\right]. \tag{3.47}$$

In view of (3.46), we have $\lim_{n\to\infty} \ln\ln\left(H_n^{-2}(f(x))\right)/\ln s_n = 1$. It follows that, with probability one, the sequence $\left(\sqrt{\gamma_n^{-1}h_n}\left(\tilde{T}_n(x) - \mathbb{E}\left(\tilde{T}_n(x)\right)\right)/\sqrt{2\ln s_n}\right)$ is relatively compact, and its limit set is the interval given in (3.47). The application of (3.33) (respectively of (3.34)) concludes the proof of Lemma 8 in the case $a > \alpha/5$ (respectively $a = \alpha/5$).

• Let us now consider the case $a < \alpha/5$ (in which case $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x)$). Set $H_n^{-2}(f(x)) = \Pi_n^{-2}(f(x))\,h_n^4\left(\ln\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right)\right)^{-1}$, and note that, since $\left(h_n^{-4}\right) \in \mathcal{GS}(4a)$, we have

$$
\begin{aligned}
\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right) &= -2\ln\left(\Pi_n(f(x))\right) + \ln\left(\prod_{k=n_0}^{n} \frac{h_{k-1}^{-4}}{h_k^{-4}}\right) + \ln\left(h_{n_0-1}^4\right) \\
&= -2\sum_{k=n_0}^{n} \ln\left(1-\gamma_k f(x)\right) + \sum_{k=n_0}^{n} \ln\left(1 - \frac{4a}{k} + o\left(\frac{1}{k}\right)\right) + \ln\left(h_{n_0-1}^4\right) \\
&= \sum_{k=n_0}^{n} \left(2\gamma_k f(x) + o(\gamma_k)\right) - \sum_{k=n_0}^{n}\left(4a\xi\gamma_k + o(\gamma_k)\right) + \ln\left(h_{n_0-1}^4\right) \\
&= \left(2f(x) - 4a\xi\right)s_n + o(s_n). \qquad (3.48)
\end{aligned}
$$

Since $2f(x) - 4a\xi > 0$, it follows in particular that $\lim_{n\to\infty} \Pi_n^{-2}(f(x))\,h_n^4 = \infty$, and thus $\lim_{n\to\infty} H_n^{-2}(f(x)) = \infty$. Moreover, we clearly have $\lim_{n\to\infty} H_n^2(f(x))/H_{n-1}^2(f(x)) = 1$. Now, set $\epsilon \in\,]0, \alpha - 5a[$ such that $\lim_{n\to\infty}(n\gamma_n) > 2a/f(x) + \epsilon/2$; in view of (3.41), and applying Lemma 3, we get

$$
\begin{aligned}
H_n^2(f(x))\sum_{k=n_0}^{n} Var\left[\tilde{\eta}_k(x)\right] &= O\left(\Pi_n^2(f(x))\,h_n^{-4}\ln\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right)\sum_{k=n_0}^{n}\frac{\Pi_k^{-2}(f(x))\,\gamma_k^2}{h_k}\right) \\
&= O\left(\Pi_n^2(f(x))\,h_n^{-4}\ln\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right)\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\,\gamma_k o\left(h_k^4 k^{-\epsilon}\right)\right) \\
&= o(1).
\end{aligned}
$$

Moreover, in view of (3.42) we have $\mathbb{E}\left[\left|\tilde{\eta}_n(x)\right|^3\right] = O\left(\Pi_n^{-3}(f(x))\,\gamma_n^3 h_n^{-2}\right)$, and thus in view of (3.48), we get

$$
\begin{aligned}
&\frac{1}{n\sqrt{n}}\sum_{k=n_0}^{n}\mathbb{E}\left(\left|H_n(f(x))\,\tilde{\eta}_k(x)\right|^3\right) \\
&= O\left(\frac{\Pi_n^3(f(x))\,h_n^{-6}}{n\sqrt{n}}\left(\ln\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right)\right)^{\frac{3}{2}}\left(\sum_{k=n_0}^{n}\frac{\Pi_k^{-3}(f(x))\,\gamma_k^3}{h_k^2}\right)\right) \\
&= O\left(\frac{\Pi_n^3(f(x))\,h_n^{-6}}{n\sqrt{n}}\left(\ln\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right)\right)^{\frac{3}{2}}\left(\sum_{k=n_0}^{n}\Pi_k^{-3}(f(x))\,\gamma_k o\left(h_k^6\right)\right)\right) \\
&= o\left(\frac{\left(\ln\ln\left(\Pi_n^{-2}(f(x))\,h_n^4\right)\right)^{\frac{3}{2}}}{n\sqrt{n}}\right) \\
&= o\left(\left[\ln\left(H_n^{-2}(f(x))\right)\right]^{-1}\right).
\end{aligned}
$$

The application of Theorem 1 in Mokkadem and Pelletier (2006) then ensures that, with probability one,

$$\lim_{n\to\infty}\frac{H_n\left(f\left(x\right)\right)S_n\left(x\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\left(f\left(x\right)\right)\right)}}=\lim_{n\to\infty}h_n^{-2}\frac{\sqrt{\ln\ln\left(\Pi_n^{-2}\left(f\left(x\right)\right)h_n^4\right)}}{\sqrt{2\ln\ln\left(H_n^{-2}\left(f\left(x\right)\right)\right)}}\left(\tilde{T}_n\left(x\right)-\mathbb{E}\left(\tilde{T}_n\left(x\right)\right)\right)=0.$$

Noting that (3.48) ensures that $\lim_{n\to\infty}\ln\ln\left(H_n^{-2}\left(f\left(x\right)\right)\right)/\ln\ln\left(\Pi_n^{-2}\left(f\left(x\right)\right)h_n^4\right)=1$, we get

$$\lim_{n\to\infty}h_n^{-2}\left[\tilde{T}_n\left(x\right)-\mathbb{E}\left(\tilde{T}_n\left(x\right)\right)\right]=0\quad a.s.,$$

and Lemma 8 in the case $a<\alpha/5$ follows from (3.34).

3.4.3 Proof of Lemma 9

Let us write

$$\tilde{T}_n\left(x\right)=\tilde{T}_n^{(1)}\left(x\right)-r\left(x\right)\tilde{T}_n^{(2)}\left(x\right)$$

with

$$\begin{aligned}\tilde{T}_n^{(1)}\left(x\right)&=\sum_{k=n_0}^{n}U_{k,n}\left(f\left(x\right)\right)\gamma_kh_k^{-1}Y_kK\left(\frac{x-X_k}{h_k}\right)\\ \tilde{T}_n^{(2)}\left(x\right)&=\sum_{k=n_0}^{n}U_{k,n}\left(f\left(x\right)\right)\gamma_kh_k^{-1}K\left(\frac{x-X_k}{h_k}\right).\end{aligned}$$

Lemma 9 is proved by showing that, for $i\in\{1,2\}$, under the condition $\lim_{n\to\infty}\left(n\gamma_n\right)>\left(\alpha-a\right)/\left(2\varphi\right)$,

- if $a\geq\alpha/5$, then $\sup_{x\in I}\left|\tilde{T}_n^{(i)}\left(x\right)-\mathbb{E}\left(\tilde{T}_n^{(i)}\left(x\right)\right)\right|=O\left(\sqrt{\gamma_nh_n^{-1}}\ \ln n\right)\quad a.s.,$ (3.49)
- if $a>\alpha/5$, then $\sup_{x\in I}\left|\mathbb{E}\left(\tilde{T}_n\left(x\right)\right)\right|=o\left(\sqrt{\gamma_nh_n^{-1}}\ \ln n\right),$ (3.50)

and by proving that, under the condition $\lim_{n\to\infty}\left(n\gamma_n\right)>2a/\varphi$,

- if $a<\alpha/5$, then $\sup_{x\in I}\left|\tilde{T}_n^{(i)}\left(x\right)-\mathbb{E}\left(\tilde{T}_n^{(i)}\left(x\right)\right)\right|=o\left(h_n^2\right)\quad a.s.,$ (3.51)
- if $a\leq\alpha/5$, then $\sup_{x\in I}\mathbb{E}\left(\tilde{T}_n\left(x\right)\right)=O\left(h_n^2\right).$ (3.52)

As a matter of fact, Lemma 9 follows from the combination of (3.49) and (3.50) in the case $a>\alpha/5$, from the one of (3.49) and (3.52) in the case $a=\alpha/5$, and from the one of (3.51) and (3.52) in the case $a<\alpha/5$.

The proof of (3.50) and (3.52) is similar to the one of (3.33) and (3.34), and is omitted. Moreover the proof of (3.49) and (3.51) for $i=2$ is similar to the one for $i=1$, and is omitted too. To prove simultaneously (3.49) and (3.51) for $i=1$, we introduce the sequence (v_n) defined as

$$\left(v_n\right)=\begin{cases}\left(\sqrt{\gamma_n^{-1}h_n}\right) & \text{if } a\geq\alpha/5,\\ \left(h_n^{-2}\left[\ln n\right]^2\right) & \text{if } a<\alpha/5.\end{cases}\qquad(3.53)$$

As a matter of fact, (3.49) and (3.51) are proved for $i=1$ by establishing that

$$\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x)-\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| = O\left(v_n^{-1}\ln n\right) \quad a.s. \tag{3.54}$$

To this end we first state the following lemma.

Lemma 18 *There exists $s>0$ such that, for all $C>0$,*

$$\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left|\tilde{T}_n^{(1)}(x)-\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right|\geq C\right]=O\left(n^{-\frac{C}{s}}\right).$$

We first show how (3.54) can be deduced from Lemma 18. Set $(M_n)\in\mathcal{GS}(\tilde{m})$ with $\tilde{m}>0$, and note that, for all $C>0$, we have

$$\begin{aligned}
&\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x)-\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right|\geq C\right]\\
&\leq \mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x)-\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right|\geq C \text{ and } \sup_{k\leq n}|Y_k|\leq M_n\right]\\
&\quad +\mathbb{P}\left[\sup_{k\leq n}|Y_k|\geq M_n\right].
\end{aligned}\tag{3.55}$$

Let (d_n) be a positive sequence such that $d_n<1$ for all n, and such that $\lim_{n\to\infty}\gamma_n^{-1}d_n=0$. Let $I_i^{(n)}$ be $N(n)$ intervals of length d_n such that $\cup_{i=1}^{N(n)}I_i^{(n)}=I$, and for all $i\in\{1,\dots,N(n)\}$, set $x_i^{(n)}\in I_i^{(n)}$. We have

$$\begin{aligned}
&\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x)-\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right|\geq C \text{ and } \sup_{k\leq n}|Y_k|\leq M_n\right]\\
&\leq \sum_{i=1}^{N(n)}\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I_i^{(n)}}\left|\tilde{T}_n^{(1)}(x)-\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right|\geq C \text{ and } \sup_{k\leq n}|Y_k|\leq M_n\right].
\end{aligned}$$

Let us prove that there exists c^* such that, for all $x,y\in I$ such that $|x-y|\leq d_n$, and on $\left\{\sup_{k\leq n}|Y_k|\leq M_n\right\}$,

$$\left|\tilde{T}_n^{(1)}(x)-\tilde{T}_n^{(1)}(y)\right|\leq c^*M_nh_n^{-1}\gamma_n^{-1}d_n. \tag{3.56}$$

To this end, we write

$$\left|\tilde{T}_n^{(1)}(x)-\tilde{T}_n^{(1)}(y)\right| \leq A_{n,1}(x,y)+A_{n,2}(x,y)$$

with

$$\begin{aligned}
A_{n,1}(x,y) &= \left|\sum_{k=n_0}^{n}U_{k,n}(f(x))\gamma_kh_k^{-1}Y_k\left[K\left(\frac{x-X_k}{h_k}\right)-K\left(\frac{y-X_k}{h_k}\right)\right]\right|,\\
A_{n,2}(x,y) &= \left|\sum_{k=n_0}^{n}U_{k,n}(f(y))\gamma_kh_k^{-1}Y_kK\left(\frac{y-X_k}{h_k}\right)\left[\frac{U_{k,n}(f(x))}{U_{k,n}(f(y))}-1\right]\right|.
\end{aligned}$$

• Since K is Lipschitz-continuous, and by application of Lemma 3, there exist $k^*, c_1^* > 0$ such that, for all $x, y \in \mathbb{R}$ satisfying $|x-y| \leq d_n$ and on $\left\{\sup_{k\leq n}|Y_k| \leq M_n\right\}$, we have :

$$\begin{aligned} A_{n,1}(x,y) &\leq k^* M_n \sum_{k=n_0}^{n} U_{k,n}(\varphi)\gamma_k h_k^{-2} d_n \\ &\leq c_1^* M_n h_n^{-2} d_n. \end{aligned} \tag{3.57}$$

• Now, let c_i^* be positive constants ; for $j \geq n_0$ and for $p \in \{1,2\}$ we have

$$\begin{aligned} \left(\frac{1-\gamma_j f(x)}{1-\gamma_j f(y)}\right)^p &= \left(1+\frac{\gamma_j\left(f(y)-f(x)\right)}{1-\gamma_j f(y)}\right)^p \\ &\leq \left(1+\frac{c_2^*\gamma_j d_n}{1-\gamma_j\|f\|_\infty}\right)^p \\ &\leq \left(1+2c_2^*\gamma_j d_n\right)^p \\ &\leq 1+c_3^*\gamma_j d_n. \end{aligned} \tag{3.58}$$

We deduce that, for k and n such that $n_0 \leq k \leq n$,

$$\begin{aligned} \left[\frac{U_{k,n}^p(f(x))}{U_{k,n}^p(f(y))}-1\right] &= \left(\prod_{j=k+1}^{n}\frac{1-\gamma_j f(x)}{1-\gamma_j f(y)}\right)^p - 1 \\ &\leq \left(\prod_{j=k+1}^{n}\exp\left(c_3^*\gamma_j d_n\right)\right) - 1 \\ &\leq \exp\left[c_3^* d_n \gamma_n^{-1}\sum_{j=k+1}^{n}\gamma_j\gamma_n\right] - 1 \\ &\leq \exp\left[c_4^* d_n\gamma_n^{-1}\sum_{n\geq 1}\gamma_n^2\right] - 1 \\ &\leq \exp\left[c_5^* d_n\gamma_n^{-1}\right] - 1 \\ &\leq c_5^* d_n\gamma_n^{-1}\exp\left[c_5^* d_n\gamma_n^{-1}\right] \\ &\leq c_6^* d_n\gamma_n^{-1}. \end{aligned} \tag{3.59}$$

The application of (3.59) with $p = 1$ and of Lemma 3 ensures that, for all $x, y \in I$ satisfying $|x-y| \leq d_n$, and on $\left\{\sup_{k\leq n}|Y_k| \leq M_n\right\}$, we have :

$$\begin{aligned} A_{n,2}(x,y) &\leq \|K\|_\infty M_n \sum_{k=n_0}^{n} U_{k,n}(\varphi)\gamma_k h_k^{-1}\left(c_6^* d_n\gamma_n^{-1}\right) \\ &\leq c_7^* M_n h_n^{-1}\gamma_n^{-1} d_n. \end{aligned} \tag{3.60}$$

The upper bound (3.56) follows from the combination of (3.57) and (3.60).
Now, it follows from (3.56) that, for all $x \in I_i^{(n)}$ and on $\left\{\sup_{k\leq n}|Y_k| \leq M_n\right\}$, we have

$$\begin{aligned} &\left|\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| \\ &\leq \left|\tilde{T}_n^{(1)}(x) - \tilde{T}_n^{(1)}\left(x_i^{(n)}\right)\right| + \left|\tilde{T}_n^{(1)}\left(x_i^{(n)}\right) - \mathbb{E}\left(\tilde{T}_n^{(1)}\left(x_i^{(n)}\right)\right)\right| + \left|\mathbb{E}\left(\tilde{T}_n^{(1)}\left(x_i^{(n)}\right)\right) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| \\ &\leq 2c^* M_n h_n^{-1}\gamma_n^{-1} d_n + \left|\tilde{T}_n^{(1)}\left(x_i^{(n)}\right) - \mathbb{E}\left(\tilde{T}_n^{(1)}\left(x_i^{(n)}\right)\right)\right|. \end{aligned}$$

In view of (3.55), we obtain, for all $C > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| \geq C\right] \\
&\leq \sum_{i=1}^{N(n)} \mathbb{P}\left[\frac{v_n}{\ln n}\left|T_n^{(1)}\left(x_i^{(n)}\right) - \mathbb{E}\left(T_n^{(1)}\left(x_i^{(n)}\right)\right)\right| + 2c^* M_n \frac{v_n}{\ln n} h_n^{-1}\gamma_n^{-1} d_n \geq C\right] \\
&\quad + n\mathbb{P}\left[|Y| \geq M_n\right].
\end{aligned}$$

Now, note that, in view of (3.53), $(v_n) \in \mathcal{GS}(m^*)$ where m^* is defined in (3.9). Set $(d_n) \in \mathcal{GS}(-(\tilde{m} + m^* + a + \alpha))$ such that, for all n, $2c^* M_n v_n (\ln n)^{-1} h_n^{-1}\gamma_n^{-1} d_n \leq C/2$; in view of Lemma 18 and Assumption $(A4)\, ii)$, there exists $s > 0$ such that

$$\begin{aligned}
&\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| \geq C\right] \\
&\leq N(n)\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left|\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| \geq \frac{C}{2}\right] + n\exp\left(-t^* M_n\right)\mathbb{E}\left(\exp\left(t^*|Y|\right)\right) \\
&= O\left(d_n^{-1} n^{-\frac{C}{2s}} + n\exp\left(-t^* M_n\right)\right).
\end{aligned}$$

Since $(M_n) \in \mathcal{GS}(\tilde{m})$ with $\tilde{m} > 0$, and since $\left(d_n^{-1}\right) \in \mathcal{GS}(\tilde{m} + m^* + a + \alpha)$, we can choose C large enough so that

$$\sum_{n\geq 0}\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right| \geq C\right] < \infty,$$

which gives (3.54).
It remains to prove Lemma 18. For all $x \in I$ and all $s > 0$, we have

$$\begin{aligned}
\mathbb{P}\left[\frac{v_n}{\ln n}\left(\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right) \geq C\right] &= \mathbb{P}\left[\exp\left[s^{-1} v_n\left(\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right)\right] \geq n^{\frac{C}{s}}\right] \\
&\leq n^{-\frac{C}{s}}\mathbb{E}\left(\exp\left[s^{-1} v_n\left(\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right)\right]\right) \\
&\leq n^{-\frac{C}{s}}\prod_{k=n_0}^{n}\mathbb{E}\left(\exp\left(s^{-1} V_{k,n}(x)\right)\right) \qquad (3.61)
\end{aligned}$$

with

$$V_{k,n}(x) = v_n U_{k,n}(f(x))\gamma_k h_k^{-1}\left[Y_k K\left(\frac{x - X_k}{h_k}\right) - \mathbb{E}\left(Y_k K\left(\frac{x - X_k}{h_k}\right)\right)\right].$$

For k and n such that $n_0 \leq k \leq n$, set

$$\alpha_{k,n} = v_n U_{k,n}(\varphi)\gamma_k h_k^{-1}.$$

We have, for all $x \in I$,

$$\begin{aligned}
&\mathbb{E}\left(\exp\left[s^{-1} V_{k,n}(x)\right]\right) \\
&\leq 1 + \frac{1}{2}\mathbb{E}\left[s^{-2} V_{k,n}^2(x)\right] + \mathbb{E}\left(s^{-3}\left|V_{k,n}^3(x)\right|\right)\exp\left[|V_{k,n}(x)|\right] \\
&\leq 1 + \frac{1}{2}s^{-2}\alpha_{k,n}^2 Var\left[Y_k K\left(\frac{x - X_k}{h_k}\right)\right] \\
&\quad + s^{-3}\alpha_{k,n}^3\|K\|_\infty^3\mathbb{E}\left[\left(|Y_k|^3 + \left(\mathbb{E}\left(|Y_k|\right)\right)^3\right)\exp\left(s^{-1}\alpha_{k,n}\|K\|_\infty\left(|Y_k| + \mathbb{E}\left(|Y_k|\right)\right)\right)\right].
\end{aligned}$$

Now, note that $\alpha_{k,n}$ can be rewritten as :

$$\alpha_{k,n} = \frac{v_n \Pi_n(\varphi)}{v_k \Pi_k(\varphi)} v_k \gamma_k h_k^{-1}.$$

Since $(v_n) \in \mathcal{GS}(m^*)$ with $\varphi - m^* \xi > 0$ (where ξ is defined in (3.6)), we have

$$\begin{aligned}\frac{\Pi_n(\varphi)}{\Pi_{n-1}(\varphi)} \frac{v_n}{v_{n-1}} &= (1-\gamma_n \varphi)\left(1 + m^* \frac{1}{n} + o\left(\frac{1}{n}\right)\right) \\ &= (1-\gamma_n\varphi)(1 + m^*\xi\gamma_n + o(\gamma_n)) \\ &= 1 - (\varphi - m^*\xi)\gamma_n + o(\gamma_n) \\ &\leq 1 \qquad \text{for } n \text{ large enough.}\end{aligned}$$

Writing

$$\frac{v_n \Pi_n(\varphi)}{v_k \Pi_k(\varphi)} = \prod_{i=k}^{n-1} \frac{v_{i+1}\Pi_{i+1}(\varphi)}{v_i \Pi_i(\varphi)},$$

we obtain

$$\sup_{n_0 \leq k \leq n} \frac{v_n \Pi_n(\varphi)}{v_k \Pi_k(\varphi)} < \infty,$$

and, since $\lim_{n\to\infty} v_k \gamma_k h_k = 0$, we deduce that $\sup_{n_0 \leq k \leq n} \alpha_{k,n} < \infty$. Thus, in view of Assumption $(A4)\, ii)$, there exist $s > 0$ and $c^* > 0$ such that, for all k and n such that $n_0 \leq k \leq n$,

$$\mathbb{E}\left[\left(|Y_k|^3 + (\mathbb{E}(|Y_k|))^3\right) \exp\left(s^{-1}\alpha_{k,n}\|K\|_\infty (|Y_k| + \mathbb{E}(|Y_k|))\right)\right] \leq c^*.$$

From classical computations, we have $\sup_{x\in I} Var\left[Y_k K\left((x - X_k) h_k^{-1}\right)\right] = O(h_k)$. We then deduce that there exist $C_1^*, C_2^* > 0$ such that, for all $x \in I$, for all k and n such that $n_0 \leq k \leq n$,

$$\begin{aligned}\mathbb{E}\left(\exp\left[s^{-1}V_{k,n}(x)\right]\right) &\leq 1 + C_1^* v_n^2 U_{n,k}^2(\varphi)\gamma_k^2 h_k^{-1} + C_2^* v_n^3 U_{k,n}^3(\varphi)\gamma_k^3 h_k^{-3} \\ &\leq \exp\left[C_1^* v_n^2 U_{k,n}^2(\varphi)\gamma_k^2 h_k^{-1} + C_2^* v_n^3 U_{k,n}^3(\varphi)\gamma_k^3 h_k^{-3}\right].\end{aligned}$$

Applying Lemma 3, we deduce from (3.61) that, for all $C > 0$,

$$\begin{aligned}&\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left(\tilde{T}_n^{(1)}(x) - \mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right)\right) \geq C\right] \\ &\leq n^{-\frac{C}{s}} \exp\left[C_1^* v_n^2 \sum_{k=n_0}^n U_{k,n}^2(\varphi)\gamma_k O\left(v_k^{-2}\right) + C_2^* v_n^3 \sum_{k=n_0}^n U_{k,n}^3(\varphi)\gamma_k O\left(v_k^{-3}\right)\right] \\ &= O\left(n^{-\frac{C}{s}}\right).\end{aligned}$$

We establish exactly in the same way that, for all $C > 0$,

$$\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left(\mathbb{E}\left(\tilde{T}_n^{(1)}(x)\right) - \tilde{T}_n^{(1)}(x)\right) \geq C\right] = O\left(n^{-\frac{C}{s}}\right),$$

which concludes the proof of Lemma 18.

3.4.4 Proof of Lemma 13

Set

$$\eta_k(x) = \left(W_{k+1}(x) - r(x) Z_{k+1}(x)\right). \tag{3.62}$$

In order to prove Lemma 13, we first establish a central limit theorem for

$$T_n(x) - \mathbb{E}(T_n(x)) = \frac{1}{\sum_{k=1}^n q_k} \sum_{k=n_0-1}^{n} q_k \left[\eta_k(x) - \mathbb{E}(\eta_k(x))\right].$$

In view of (3.37)-(3.39) and since $h_k/h_{k+1} = 1 + o(1)$, we have

$$\begin{aligned} Var(\eta_k(x)) &= Var(W_{k+1}(x)) + r^2(x) Var(Z_{k+1}(x)) - 2r(x) Cov(W_{k+1}(x), Z_{k+1}(x)), \\ &= \frac{1}{h_k}\left[Var[Y|X=x] f(x) \int_{\mathbb{R}} K^2(z)\,dz + o(1)\right]. \end{aligned}$$

Noting that $\left(q_n^2 h_n^{-1}\right) \in \mathcal{GS}(-2q+a)$ with $q < (1+a)/2$, and using (3.18), we get

$$\begin{aligned} v_n^2 &= \sum_{k=n_0-1}^{n} q_k^2 Var(\eta_k(x)) \\ &= \sum_{k=n_0-1}^{n} \frac{q_k^2}{h_k}\left[Var[Y|X=x] f(x) \int_{\mathbb{R}} K^2(z)\,dz + o(1)\right] \\ &= \frac{n q_n^2 h_n^{-1}}{1-2q+a}\left[Var[Y|X=x] f(x) \int_{\mathbb{R}} K^2(z)\,dz + o(1)\right]. \end{aligned} \tag{3.63}$$

Now, set $p \in]0,1]$ such that $q < (1 + a(1+p))/(2+p)$; it follows from (3.42) that

$$\begin{aligned} \sum_{k=n_0-1}^{n} q_k^{2+p} \mathbb{E}\left[|\eta_k(x)|^{2+p}\right] &= O\left(\sum_{k=n_0-1}^{n} \frac{q_k^{2+p}}{h_k^{2+p}} \mathbb{E}\left(|Y_k - r(x)|^{2+p} K^{2+p}\left(\frac{x-X_k}{h_k}\right)\right)\right) \\ &= O\left(\sum_{k=n_0-1}^{n} \frac{q_k^{2+p}}{h_k^{1+p}}\right). \end{aligned}$$

In view of (3.63) and using (3.18), we get

$$\begin{aligned} \frac{1}{v_n^{2+p}} \sum_{k=n_0-1}^{n} q_k^{2+p} \mathbb{E}\left[|\eta_k(x)|^{2+p}\right] &= O\left(\frac{n q_n^{2+p} h_n^{-(1+p)}}{\left(n q_n^2 h_n^{-1}\right)^{1+\frac{p}{2}}}\right) \\ &= O\left(\frac{1}{n^{\frac{p}{2}} h_n^{\frac{p}{2}}}\right) \\ &= o(1). \end{aligned}$$

The application of Lyapounov Theorem gives

$$\frac{\sum_{k=1}^n q_k}{\sqrt{n q_n^2 h_n^{-1}}} \left(T_n(x) - \mathbb{E}[T_n(x)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{1+a-2q} Var[Y|X=x] f(x) \int_{\mathbb{R}} K^2(z)\,dz\right),$$

and applying (3.18), we obtain

$$\sqrt{nh_n}\left(T_n(x)-\mathbb{E}\left[T_n(x)\right]\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\frac{(1-q)^2}{1+a-2q}Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz\right). \tag{3.64}$$

Now, note that

$$\mathbb{E}\left(T_n(x)\right) = \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^n q_k\left[\left(\mathbb{E}\left(W_{k+1}(x)\right)-a(x)\right)-r(x)\left(\mathbb{E}\left(Z_{k+1}(x)\right)-f(x)\right)\right].$$

Since $h_{n+1}/h_n = 1+o(1)$, it follows from (3.44) and (3.45) that

$$\begin{aligned}&\mathbb{E}\left(T_n(x)\right)\\ &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^n q_k h_{k+1}^2\left[\frac{1}{2}\left(\int_{\mathbb{R}}y\frac{\partial^2 g}{\partial x^2}(x,y)\,dy - r(x)\int_{\mathbb{R}}\frac{\partial^2 g}{\partial x^2}(x,y)\,dy\right)+o(1)\right]\int_{\mathbb{R}}z^2K(z)\,dz\\ &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^n q_k h_k^2\left[m^{(2)}(x)f(x)+o(1)\right].\end{aligned}$$

Applying (3.18), we obtain

$$\lim_{n\to\infty}\frac{1}{h_n^2}\mathbb{E}\left(T_n(x)\right) = \frac{1-q}{1-2a-q}m^{(2)}(x)f(x), \tag{3.65}$$

and Lemma 13 follows from the combination of (3.64) and (3.65).

3.4.5 Proof of Lemma 14

Set

$$S_n(x) = \sum_{k=n_0-1}^n q_k\left[\eta_k(x)-\mathbb{E}\left(\eta_k(x)\right)\right]$$

where η_k is defined in (3.62), and $H_n^{-2} = nh_n^{-1}q_n^2$. Let us first note that, since $\left(nh_n^{-1}q_n^2\right)\in\mathcal{GS}(1+a-2q)$ with $1+a-2q>0$, we have $\lim_{n\to\infty}H_n^{-2}=\infty$. Moreover, we have $\lim_{n\to\infty}H_n^2/H_{n-1}^2 = 1$, and, by (3.63),

$$\lim_{n\to\infty}H_n^2\sum_{k=n_0-1}^n q_k^2 Var\left[\eta_k(x)\right] = \left[1+a-2q\right]^{-1}Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz$$

and, by (3.42), $\mathbb{E}\left[\left|q_n\eta_n(x)\right|^3\right] = O\left(q_n^3h_n^{-2}\right)$. Since, for all $\epsilon>0$,

$$\begin{aligned}\frac{1}{n\sqrt{n}}\sum_{k=n_0-1}^n\mathbb{E}\left(\left|H_nq_k\eta_k(x)\right|^3\right) &= O\left(\frac{H_n^3}{n\sqrt{n}}\sum_{k=n_0-1}^n\frac{q_k^3}{h_k^2}\right)\\ &= O\left(n^{-3}h_n^{\frac{3}{2}}q_n^{-3}\left(n^{\epsilon}+nq_n^3h_n^{-2}\right)\right),\end{aligned}$$

we have

$$\frac{1}{n\sqrt{n}}\sum_{k=n_0-1}^n\mathbb{E}\left(\left|H_nq_k\eta_k(x)\right|^3\right) = o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right).$$

The application of Theorem 1 of Mokkadem and Pelletier (2006) ensures that, with probability one, the sequence

$$\left(\frac{H_n S_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right) = \left(\frac{\sum_{k=1}^n q_k}{nq_n}\frac{\sqrt{nh_n}\left(T_n(x) - \mathbb{E}\left(T_n(x)\right)\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\frac{1}{1+a-2q}Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz},\sqrt{\frac{1}{1+a-2q}Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz}\right].$$

Since $\lim_{n\to\infty}\ln\ln\left(H_n^{-2}\right)/\ln\ln n = 1$, and using (3.18), it follows that, with probability one, the sequence $\left(\sqrt{nh_n}\left(T_n(x) - \mathbb{E}\left(T_n(x)\right)\right)/\sqrt{2\ln\ln n}\right)$ is relatively compact, and its limit set is the interval

$$\left[-\sqrt{\frac{(1-q)^2}{1+a-2q}Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz},\sqrt{\frac{(1-q)^2}{1+a-2q}Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz}\right].$$

The application of (3.65) concludes the proof of Lemma 14.

3.4.6 Proof of Lemma 15

Let us write $T_n(x)$ as

$$T_n(x) = T_{n,1}(x) - r(x)\,T_{n,2}(x)$$

with

$$\begin{aligned} T_{n,1}(x) &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^{n}\frac{q_k}{h_{k+1}}Y_{k+1}K\left(\frac{x-X_{k+1}}{h_{k+1}}\right) \\ T_{n,2}(x) &= \frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^{n}\frac{q_k}{h_{k+1}}K\left(\frac{x-X_{k+1}}{h_{k+1}}\right). \end{aligned}$$

Lemma 15 is proved by showing that, for $i \in \{1,2\}$,

$$\sup_{x\in I}\left|T_{n,i}(x) - \mathbb{E}\left(T_{n,i}(x)\right)\right| = O\left(\sqrt{n^{-1}h_n^{-1}\ln n}\right) \quad a.s., \tag{3.66}$$

and that

$$\sup_{x\in I}\left|\mathbb{E}\left(T_{n,1}(x)\right) - r(x)\,\mathbb{E}\left(T_{n,2}(x)\right)\right| = O\left(h_n^2\right). \tag{3.67}$$

The proof of (3.67) relies on classical computations and is omitted. Moreover the proof of (3.66) for $i=2$ is similar to the one for $i=1$, and is omitted too. We now prove (3.66) for $i=1$. To this end, we first state the following lemma.

Lemma 19 *There exists $s>0$ such that, for all $C>0$,*

$$\sup_{x\in I}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\left|T_{n,1}(x) - \mathbb{E}\left(T_{n,1}(x)\right)\right| \geq C\right] = O\left(n^{-\frac{C}{s}}\right).$$

We first show how (3.66) for $i=1$ can be deduced from Lemma 19, and then prove Lemma 19. Set $(M_n) \in \mathcal{GS}(\tilde{m})$ with $\tilde{m}>0$, and note that, for all $C>0$, we have

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I}\left|T_{n,1}(x)-\mathbb{E}(T_{n,1}(x))\right|\geq C\right] \\
&\leq \mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I}\left|T_{n,1}(x)-\mathbb{E}(T_{n,1}(x))\right|\geq C \text{ and } \sup_{k\leq n}|Y_{k+1}|\leq M_n\right] \\
&\quad +\mathbb{P}\left[\sup_{k\leq n}|Y_{k+1}|\geq M_n\right].
\end{aligned}$$

Let $I_i^{(n)}$ be $N(n)$ intervals of length d_n such that $\cup_{i=1}^{N(n)} I_i^{(n)} = I$, and for all $i\in\{1,\ldots,N(n)\}$, set $x_i^{(n)} \in I_i^{(n)}$. We have

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I}\left|T_{n,1}(x)-\mathbb{E}(T_{n,1}(x))\right|\geq C\right] \\
&\leq \sum_{i=1}^{N(n)}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I_i^{(n)}}\left|T_{n,1}(x)-\mathbb{E}(T_{n,1}(x))\right|\geq C \text{ and } \sup_{k\leq n}|Y_{k+1}|\leq M_n\right] \\
&\quad +\mathbb{P}\left[\sup_{k\leq n}|Y_{k+1}|\geq M_n\right]. \qquad (3.68)
\end{aligned}$$

Since K is Lipschitz-continuous, there exist $k^*, c^*>0$, such that, for all $x,y\in\mathbb{R}$ satisfying $|x-y|\leq d_n$ and on $\left\{\sup_{k\leq n}|Y_{k+1}|\leq M_n\right\}$, we have :

$$\begin{aligned}
\left|T_{n,1}(x)-T_{n,1}(y)\right| &= \left|\frac{1}{\sum_{k=1}^n q_k}\sum_{k=n_0-1}^{n} q_k h_{k+1}^{-1}Y_{k+1}\left[K\left(\frac{x-X_{k+1}}{h_{k+1}}\right)-K\left(\frac{y-X_{k+1}}{h_{k+1}}\right)\right]\right| \\
&\leq k^* M_n d_n \frac{1}{\sum_{k=1}^n q_k}\sum_{k=1}^{n} q_k h_{k+1}^{-2} \\
&\leq c^* M_n h_n^{-2} d_n.
\end{aligned}$$

It follows that, for all $x\in I_i^{(n)}$, on $\left\{\sup_{k\leq n}|Y_{k+1}|\leq M_n\right\}$, we have

$$\begin{aligned}
&\left|T_{n,1}(x)-\mathbb{E}(T_{n,1}(x))\right| \\
&\leq \left|T_{n,1}(x)-T_{n,1}\left(x_i^{(n)}\right)\right|+\left|T_{n,1}\left(x_i^{(n)}\right)-\mathbb{E}\left(T_{n,1}\left(x_i^{(n)}\right)\right)\right|+\left|\mathbb{E}\left(T_{n,1}\left(x_i^{(n)}\right)\right)-\mathbb{E}(T_{n,1}(x))\right| \\
&\leq 2c^* M_n h_n^{-2} d_n+\left|T_{n,1}\left(x_i^{(n)}\right)-\mathbb{E}\left(T_{n,1}\left(x_i^{(n)}\right)\right)\right|.
\end{aligned}$$

In view of (3.68), we obtain, for all $C>0$,

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I}\left|T_{n,1}(x)-\mathbb{E}(T_{n,1}(x))\right|\geq C\right] \\
&\leq \sum_{i=1}^{N(n)}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\left|T_{n,1}\left(x_i^{(n)}\right)-\mathbb{E}\left(T_{n,1}\left(x_i^{(n)}\right)\right)\right|+2c^*M_n\frac{\sqrt{nh_n}}{\ln n}h_n^{-2}d_n\geq C\right] \\
&\quad +n\mathbb{P}\left[|Y|\geq M_n\right].
\end{aligned}$$

Now, set $(d_n) \in \mathcal{GS}\left(-\frac{1}{2}-\frac{3}{2}a-\tilde{m}\right)$ such that, for all n, $2c^*M_n\sqrt{nh_n}\left(\ln n\right)^{-1}h_n^{-2}d_n \leq C/2$; in view of Lemma 19 and Assumption $(A4)\,ii)$, there exists $s>0$ such that

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I}\left|T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right|\geq C\right]\\
&\leq\ N(n)\sup_{x\in I}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\left|T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right|\geq\frac{C}{2}\right]+n\exp\left(-t^*M_n\right)\mathbb{E}\left(\exp\left(t^*\left|Y\right|\right)\right)\\
&=\ O\left(d_n^{-1}n^{-\frac{C}{2s}}+n\exp\left(-t^*M_n\right)\right).
\end{aligned}$$

Since $\left(d_n^{-1}\right)\in\mathcal{GS}\left(\frac{1}{2}+\frac{3}{2}a+\tilde{m}\right)$ and since $(M_n)\in\mathcal{GS}(\tilde{m})$ with $\tilde{m}>0$, we can choose C large enough so that

$$\sum_{n\geq 0}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\sup_{x\in I}\left|T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right|\geq C\right]<\infty,$$

which gives (3.66) for $i=1$.
It remains to prove Lemma 19. For all $x\in I$ and all $s>0$, we have

$$\begin{aligned}
\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\left(T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right)\geq C\right] &= \mathbb{P}\left[\exp\left[s^{-1}\sqrt{nh_n}\left(T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right)\right]\geq n^{\frac{C}{s}}\right]\\
&\leq n^{-\frac{C}{s}}\mathbb{E}\left(\exp\left[s^{-1}\sqrt{nh_n}\left(T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right)\right]\right)\\
&\leq n^{-\frac{C}{s}}\prod_{k=n_0-1}^{n}\mathbb{E}\left(\exp\left(s^{-1}U_{k,n}(x)\right)\right)
\end{aligned}\tag{3.69}$$

with

$$U_{k,n}(x) = \frac{\sqrt{nh_n}}{\sum_{k=1}^n q_k}\frac{q_k}{h_{k+1}}\left[Y_{k+1}K\left(\frac{x-X_{k+1}}{h_{k+1}}\right)-\mathbb{E}\left(\left(Y_{k+1}K\left(\frac{x-X_{k+1}}{h_{k+1}}\right)\right)\right)\right].$$

For k and n such that $k\leq n$, set

$$\alpha_{k,n} = \frac{\sqrt{nh_n}}{\sum_{k=1}^n q_k}\frac{q_k}{h_{k+1}}.$$

We have, for all $x\in I$,

$$\begin{aligned}
&\mathbb{E}\left(\exp\left[s^{-1}U_{k,n}(x)\right]\right)\\
&\leq\ 1+\frac{1}{2}\mathbb{E}\left[s^{-2}U_{k,n}^2(x)\right]+\mathbb{E}\left[s^{-3}\left|U_{k,n}^3(x)\right|\right]\exp\left[\left|U_{k,n}(x)\right|\right]\\
&\leq\ 1+\frac{1}{2}s^{-2}\alpha_{k,n}^2 Var\left[Y_{k+1}K\left(\frac{x-X_{k+1}}{h_{k+1}}\right)\right]\\
&\quad +s^{-3}\alpha_{k,n}^3\|K\|_\infty^3\mathbb{E}\left[\left(\left|Y_{k+1}\right|^3+\left(\mathbb{E}\left(\left|Y_{k+1}\right|\right)\right)^3\right)\exp\left(s^{-1}\alpha_{k,n}\|K\|_\infty\left(\left|Y_{k+1}\right|+\mathbb{E}\left(\left|Y_{k+1}\right|\right)\right)\right)\right].
\end{aligned}$$

Now, note that

$$\begin{aligned}
\alpha_{k,n} &= \left(\frac{nq_n}{\sum_{k=1}^n q_k}\right)\frac{q_k\sqrt{kh_k^{-1}}}{q_n\sqrt{nh_n^{-1}}}\sqrt{k^{-1}h_kh_{k+1}^{-2}}\\
&= \left(\frac{nq_n}{\sum_{k=1}^n q_k}\right)\left(\prod_{j=k}^{n-1}\frac{q_j\sqrt{jh_j^{-1}}}{q_{j+1}\sqrt{(j+1)\,h_{j+1}^{-1}}}\right)\sqrt{k^{-1}h_kh_{k+1}^{-2}}.
\end{aligned}$$

Since $\left(q_j\sqrt{jh_j^{-1}}\right) \in \mathcal{GS}\left(-q+(1+a)/2\right)$ with $-q+(1+a)/2>0$, we have

$$\begin{aligned}\frac{q_j\sqrt{jh_j^{-1}}}{q_{j+1}\sqrt{(j+1)\,h_{j+1}^{-1}}} &= 1-\left(-q+\frac{1+a}{2}\right)\frac{1}{j}+o\left(\frac{1}{j}\right)\\ &\leq 1 \quad \text{for } j \text{ large enough.}\end{aligned}$$

It follows that $\sup_{k\leq n}\alpha_{k,n}<\infty$. Consequently, in view of Assumption $(A4)\,ii)$, there exist $s>0$ and $c^*>0$ such that, for all k and n such that $k\leq n$,

$$\mathbb{E}\left[\left(|Y_{k+1}|^3+\left(\mathbb{E}\left(|Y_{k+1}|\right)\right)^3\right)\exp\left(s^{-1}\alpha_{k,n}\|K\|_\infty\left(|Y_{k+1}|+\mathbb{E}\left(|Y_{k+1}|\right)\right)\right)\right]\leq c^*.$$

Recall that $\sup_{x\in I}Var\left[Y_kK\left((x-X_k)\,h_k^{-1}\right)\right]=O\left(h_k\right)$. We then deduce that there exist positive constants C_i^*, such that, for all $x\in I$, and for all k and n such that $k\leq n$,

$$\begin{aligned}\mathbb{E}\left(\exp\left[s^{-1}U_{k,n}(x)\right]\right) &\leq 1+C_1^*\frac{nh_n}{\left(\sum_{k=1}^n q_k\right)^2}\frac{q_k^2}{h_k}+C_2^*\frac{(nh_n)^{\frac{3}{2}}}{\left(\sum_{k=1}^n q_k\right)^3}\frac{q_k^3}{h_k^3}\\ &\leq \exp\left[C_3^*\frac{h_n}{nq_n^2}\frac{q_k^2}{h_k}+C_4^*\frac{h_n^{\frac{3}{2}}}{n^{\frac{3}{2}}q_n^3}\frac{q_k^3}{h_k^3}\right].\end{aligned}$$

Then, it follows from (3.69) that, for all $C>0$,

$$\begin{aligned}\sup_{x\in I}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\left(T_{n,1}(x)-\mathbb{E}\left(T_{n,1}(x)\right)\right)\geq C\right] &\leq n^{-\frac{C}{s}}\exp\left[C_3^*\frac{h_n}{nq_n^2}\sum_{k=1}^n\frac{q_k^2}{h_k}+C_4^*\frac{h_n^{\frac{3}{2}}}{n^{\frac{3}{2}}q_n^3}\sum_{k=1}^n\frac{q_k^3}{h_k^3}\right]\\ &= O\left(n^{-\frac{C}{s}}\right).\end{aligned}$$

We establish exactly in the same way that, for all $C>0$,

$$\sup_{x\in I}\mathbb{P}\left[\frac{\sqrt{nh_n}}{\ln n}\left(\mathbb{E}\left(T_{n,1}(x)\right)-T_{n,1}(x)\right)\geq C\right] = O\left(n^{-\frac{C}{s}}\right),$$

which concludes the proof of Lemma 19.

3.4.7 Proof of Lemma 16

In view of (3.20), we have

$$\Delta_n(x) = \Delta_n^{(1)}(x)+\Delta_n^{(2)}(x),$$

with

$$\Delta_n^{(1)}(x) = \sum_{k=n_0}^n U_{k,n}\left(f(x)\right)\gamma_k\left(\mathbb{E}\left[Z_k(x)\right]-Z_k(x)\right)\left(r_{k-1}(x)-r(x)\right), \tag{3.70}$$

$$\Delta_n^{(2)}(x) = \sum_{k=n_0}^n U_{k,n}\left(f(x)\right)\gamma_k\left(f(x)-\mathbb{E}\left[Z_k(x)\right]\right)\left(r_{k-1}(x)-r(x)\right). \tag{3.71}$$

Let us first note that, in view of (3.45) and by application of Lemma 3, we have

$$\begin{aligned}\left|\Delta_n^{(2)}(x)\right| &= O\left(\Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k h_k^2 w_k\right) \quad a.s.\\ &= O\left(\Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k O(m_k) w_k\right) \quad a.s.\\ &= O(m_n w_n) \quad a.s.\end{aligned}$$

Let us now bound $\Delta_n^{(1)}(x)$. To this end, we set

$$\begin{aligned}\varepsilon_k(x) &= \mathbb{E}(Z_k(x)) - Z_k(x),\\ G_k(x) &= r_k(x) - r(x),\\ S_n(x) &= \sum_{k=1}^{n}\Pi_k^{-1}(f(x))\gamma_k\varepsilon_k(x)G_{k-1}(x),\end{aligned}$$

and $\mathcal{F}_k = \sigma((X_1, Y_1), \ldots, (X_k, Y_k))$. In view of (3.38) and of Lemma 3, the increasing process of the martingale $(S_n(x))$ satisfies

$$\begin{aligned}< S >_n(x) &= \sum_{k=n_0}^{n}\mathbb{E}\left[\Pi_k^{-2}(f(x))\gamma_k^2\varepsilon_k^2(x)G_{k-1}^2(x)|\mathcal{F}_{k-1}\right]\\ &= \sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k^2 G_{k-1}^2(x)Var[Z_k(x)]\\ &= O\left(\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k^2 w_k^2\frac{1}{h_k}\right) \quad a.s.\\ &= O\left(\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k m_k^2 w_k^2\right) \quad a.s.\\ &= O\left(\Pi_n^{-2}(f(x))m_n^2 w_n^2\right) \quad a.s.\end{aligned}$$

• Let us first consider the case the sequence $(n\gamma_n)$ is bounded. We then have $\left(\Pi_n^{-1}(f(x)) \in \mathcal{GS}\left(\xi^{-1}f(x)\right)\right)$, and thus $\ln(< S >_n(x)) = O(\ln n)$ a.s. Theorem 1.3.15 in Duflo (1997) then ensures that, for any $\delta > 0$,

$$\begin{aligned}|S_n(x)| &= o\left(< S >_n^{\frac{1}{2}}(x)(\ln < S >_n(x))^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s.\\ &= o\left(\Pi_n^{-1}(f(x))m_n w_n(\ln n)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s.\end{aligned}$$

It follows from the application of Lemma 3, for any $\delta > 0$,

$$\begin{aligned}\left|\Delta_n^{(1)}(x)\right| &= o\left(m_n w_n(\ln n)^{\frac{1+\delta}{2}}\right) + O(\Pi_n(f(x))) \quad a.s.\\ &= o\left(m_n w_n(\ln n)^{\frac{1+\delta}{2}}\right) + o(m_n) \quad a.s.,\end{aligned}$$

which concludes the proof of Lemma 16 in this case.

• Let us now consider the case $\lim_{n\to\infty}(n\gamma_n) = \infty$. In this case, for all $\delta > 0$, we have

$$\begin{aligned}\ln\left(\Pi_n^{-2}(f(x))\right) &= \sum_{k=n_0}^{n}\ln(1-\gamma_k f(x))^{-2}\\ &= \sum_{k=n_0}^{n}(2\gamma_k f(x) + o(\gamma_k))\\ &= O\left(\sum_{k=1}^{n}\gamma_k k^{\delta}\right).\end{aligned}$$

Since $\left(\gamma_n n^{\delta}\right) \in \mathcal{GS}(-(\alpha-\delta))$ with $(\alpha-\delta) < 1$, we have

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

It follows that $\ln\left(\Pi_n^{-2}(f(x))\right) = O\left(n^{1+\delta}\gamma_n\right)$. The sequence $(m_n w_n)$ being in $\mathcal{GS}(-m^*+w^*)$, we deduce that, for all $\delta > 0$, we have

$$\ln(<S>_n(x)) = O\left(n^{1+\delta}\gamma_n\right) \quad a.s.$$

Theorem 1.3.15 in Duflo (1997) then ensures that, for any $\delta > 0$,

$$\begin{aligned}|S_n(x)| &= o\left(<S>_n^{\frac{1}{2}}(x)\,(\ln <S>_n(x))^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s.\\ &= o\left(\Pi_n^{-1}(f(x))\,m_n w_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s.\end{aligned}$$

It follows from the application of Lemma 3 that, for any $\delta > 0$,

$$\begin{aligned}\left|\Delta_n^{(1)}(x)\right| &= o\left(m_n w_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) + O(\Pi_n(f(x))) \quad a.s.\\ &= o\left(m_n w_n\left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) \quad a.s.,\end{aligned}$$

which concludes the proof of Lemma 16.

3.4.8 Proof of Lemma 17

Let us first note that, in view of (3.71), and by application of Lemma 3, we have

$$\begin{aligned}\sup_{x\in I}\left|\Delta_n^{(2)}(x)\right| &= O\left(\sum_{k=n_0}^{n}\left(\sup_{x\in I}U_{k,n}(f(x))\right)\gamma_k h_k^2 w_k\right) \quad a.s.\\ &= O\left(\sum_{k=n_0}^{n}U_{k,n}(\varphi)\,\gamma_k m_k w_k\right) \quad a.s.\\ &= O(m_n w_n) \quad a.s.\end{aligned}$$

Now, set

$$A_n = \frac{3}{t^*}\ln n \tag{3.72}$$

(where t^* is defined in $(A4)\,ii)$), and write $\Delta_n^{(1)}$ (defined in (3.70)) as

$$\Delta_n^{(1)}(x) = \Pi_n(f(x))\, M_n^{(n)}(x) + \Pi_n(f(x))\, S_n(x)$$

with

$$\begin{aligned} S_n(x) &= \sum_{k=n_0}^{n} \Pi_k^{-1}(f(x))\,\gamma_k\,(\mathbb{E}[Z_k(x)] - Z_k(x))\,(r_{k-1}(x) - r(x))\,\mathbb{1}_{\sup_{l\leq k-1}|Y_l|>A_n}, \\ M_k^{(n)}(x) &= \sum_{j=n_0}^{k} \Pi_j^{-1}(f(x))\,\gamma_j\,(\mathbb{E}[Z_j(x)] - Z_j(x))\,(r_{j-1}(x) - r(x))\,\mathbb{1}_{\sup_{l\leq j-1}|Y_l|\leq A_n}. \end{aligned}$$

Let us first prove a uniform strong upper bound for S_n. For any $c > 0$, we have

$$\begin{aligned} \sum_{n\geq 0} \mathbb{P}\left[\sup_{x\in I} m_n^{-1} w_n^{-1} |S_n(x)| \geq c\right] &= O\left(\sum_{n\geq 0} \mathbb{P}\left(\sup_{l\leq n-1} |Y_l| > A_n\right)\right) \\ &= O\left(\sum_{n\geq 0} n\mathbb{P}(|Y| > A_n)\right) \\ &= O\left(\sum_{n\geq 0} n\exp(-t^* A_n)\right) \\ &< \infty. \end{aligned}$$

It follows that

$$\sup_{x\in I} |S_n(x)| = O(m_n w_n) \quad a.s.$$

To establish the strong uniform bound of $M_n^{(n)}$, we shall apply the following result given in Duflo (1997), page 209.

Lemma 20
Let $\left(M_k^{(n)}\right)_k$ be a martingale such that, for all $k \leq n$, $\left|M_k^{(n)} - M_{k-1}^{(n)}\right| \leq c_n$, and set $\Phi_c(\lambda) = c^{-2}\left(e^{\lambda c} - 1 - \lambda c\right)$. For all λ_n such that $\lambda_n c_n \leq 1$ and all $\alpha_n > 0$, we have

$$\mathbb{P}\left(\lambda_n\left(M_n^{(n)} - M_0^{(n)}\right) \geq \Phi_{c_n}(\lambda_n) < M^{(n)} >_n + \alpha_n\lambda_n\right) \leq e^{-\alpha_n\lambda_n}.$$

In view of (3.21), there exists $C^* > 0$ such that, on $\left\{\sup_{l\leq k} |Y_l| \leq A_n\right\}$,

$$|r_k(x) - r(x)| \leq C^* k\gamma_k h_k^{-1} A_n.$$

Consequently, there exists $C_1 > 0$ such that

$$\begin{aligned} \left|M_k^{(n)}(x) - M_{k-1}^{(n)}(x)\right| &\leq \Pi_k^{-1}(f(x))\,\gamma_k\,|Z_k(x) - \mathbb{E}(Z_k(x))|\,\left|(r_{k-1}(x) - r(x))\,\mathbb{1}_{\sup_{l\leq k-1}|Y_l|\leq A_n}\right| \\ &\leq \Pi_k^{-1}(f(x))\,\gamma_k\left(2h_k^{-1}\|K\|_\infty\right)\left(C^*(k-1)\,\gamma_{k-1}h_{k-1}^{-1}A_n\right) \\ &\leq C_1\Pi_k^{-1}(f(x))\,k\gamma_k^2 h_k^{-2} A_n. \end{aligned}$$

- In the case $\lim_{n\to\infty}(n\gamma_n)=\infty$, since $\left(n\gamma_n^2h_n^{-2}\right)\in\mathcal{GS}(1-2\alpha+2a)$ there exists $(u_k)\to 0$ such that

$$\begin{aligned}\frac{(k-1)\gamma_{k-1}^2h_{k-1}^{-2}}{k\gamma_k^2h_k^{-2}} &= 1-[1-2\alpha+2a]\frac{1}{k}+o\left(\frac{1}{k}\right)\\ &= 1+u_k\gamma_k.\end{aligned}$$

It follows that there exists $k_0\geq n_0$ such that, for all $k\geq k_0$ and for all $x\in I$,

$$\begin{aligned}\frac{\Pi_{k-1}^{-1}(f(x))(k-1)\gamma_{k-1}^2h_{k-1}^{-2}}{\Pi_k^{-1}(f(x))k\gamma_k^2h_k^{-2}} &= (1-\gamma_kf(x))(1+u_k\gamma_k)\\ &= 1-\gamma_kf(x)+u_k\gamma_k(1-\gamma_kf(x))\\ &\leq 1-\gamma_k\varphi+u_k\gamma_k(1+\gamma_k\|f\|_\infty)\\ &\leq 1.\end{aligned}$$

Consequently, there exists $C>0$ such that, for all $x\in I$ and all $k\leq n$,

$$\left|M_k^{(n)}(x)-M_{k-1}^{(n)}(x)\right| \leq C\Pi_n^{-1}(f(x))n\gamma_n^2h_n^{-2}A_n. \tag{3.73}$$

- In the case $\lim_{n\to\infty}(n\gamma_n)<\infty$ (in which case $\alpha=1$), we set $\epsilon\in\left]0,\min\left\{(1-3a)/2;\varphi\xi^{-1}-m^*\right\}\right[$ (where m^* is defined in (3.9)), and write

$$\left|M_k^{(n)}(x)-M_{k-1}^{(n)}(x)\right| \leq C_1\left[\Pi_k^{-1}(f(x))k^{-\epsilon}m_k\right]A_n\left[m_k^{-1}k^{1+\epsilon}\gamma_k^2h_k^{-2}\right].$$

Since $\left(m_n^{-1}n^{1+\epsilon}\gamma_n^2h_n^{-2}\right)\in\mathcal{GS}(m^*+1+\epsilon-2\alpha+2a)$ with

$$\begin{aligned}m^*+1+\epsilon-2\alpha+2a &\leq \frac{1-a}{2}+\epsilon-1+2a\\ &\leq \epsilon-\frac{1}{2}(1-3a)<0,\end{aligned}$$

the sequence $\left(m_n^{-1}n^{1+\epsilon}\gamma_n^2h_n^{-2}\right)$ is bounded. On the other hand, since $(n^{-\epsilon}m_n)\in\mathcal{GS}(-\epsilon-m^*)$, there exists $(u_k)\to 0$ and $k_0\geq n_0$ such that, for all $k\geq k_0$ and for all $x\in I$,

$$\begin{aligned}\frac{\Pi_{k-1}^{-1}(f(x))(k-1)^{-\epsilon}m_{k-1}}{\Pi_k^{-1}(f(x))k^{-\epsilon}m_k} &= (1-\gamma_kf(x))\left(1+(m^*+\epsilon)\frac{1}{k}+o\left(\frac{1}{k}\right)\right)\\ &= (1-\gamma_kf(x))(1+(m^*+\epsilon)\xi\gamma_k+u_k\gamma_k)\\ &\leq (1-\gamma_k\varphi)(1+(m^*+\epsilon)\xi\gamma_k+|u_k|\gamma_k)\\ &\leq 1-\frac{(\varphi-(m^*+\epsilon)\xi)\gamma_k}{2}\\ &\leq 1.\end{aligned}$$

It follows that there exists $C>0$ such that, for all $x\in I$ and all $k\leq n$,

$$\left|M_k^{(n)}(x)-M_{k-1}^{(n)}(x)\right| \leq C\Pi_n^{-1}(f(x))n^{-\epsilon}m_nA_n. \tag{3.74}$$

From now on, we set

$$c_n(x)=\begin{cases}C\Pi_n^{-1}(f(x))n\gamma_n^2h_n^{-2}A_n & \text{if } \lim_{n\to\infty}(n\gamma_n)=\infty,\\ C\Pi_n^{-1}(f(x))m_nn^{-\epsilon}A_n & \text{if } \lim_{n\to\infty}(n\gamma_n)<\infty,\end{cases}$$

so that in view of (3.73) and (3.74), for all $x \in I$ and all $k \leq n$, we have

$$\left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| \leq c_n(x).$$

Now, let (u_n) be a positive sequence such that, for all n,

$$\begin{cases} u_n \leq C^{-1} n^{-1} \gamma_n^{-2} h_n^2 A_n^{-1} & \text{if } \lim_{n\to\infty}(n\gamma_n) = \infty, \\ u_n \leq C^{-1} m_n^{-1} n^{\epsilon} A_n^{-1} & \text{if } \lim_{n\to\infty}(n\gamma_n) < \infty, \end{cases} \tag{3.75}$$

and set

$$\lambda_n(x) = u_n \Pi_n(f(x)).$$

Let us at first assume that the following lemma holds.

Lemma 21
There exist $C_2 > 0$ and $\rho > 0$ such that for all $x, y \in I$ such that $|x-y| \leq C_2 n^{-\rho}$, we have

$$\begin{aligned} \left| \lambda_n(x) M_n^{(n)}(x) - \lambda_n(y) M_n^{(n)}(y) \right| &\leq 1, \\ \left| \Phi_{c_n(x)}(\lambda_n(x)) < M^{(n)} >_n (x) - \Phi_{c_n(y)}(\lambda_n(y)) < M^{(n)} >_n (y) \right| &\leq 1. \end{aligned}$$

Set

$$\begin{aligned} d_n &= C_2 n^{-\rho}, \\ V_n(x) &= \lambda_n(x) M_n^{(n)}(x) - \Phi_{c_n(x)}(\lambda_n(x)) < M^{(n)} >_n (x), \\ \alpha_n(x) &= \frac{(\rho+2)\ln n}{\lambda_n(x)}. \end{aligned}$$

Let $I_i^{(n)}$ be $N(n)$ intervals of length d_n such that $\cup_{i=1}^{N(n)} = I$, and for all $i \in \{1, \ldots, N(n)\}$, set $x_i^{(n)} \in I_i^{(n)}$. Applying Lemma 21, we get, for n large enough,

$$\begin{aligned} \mathbb{P}\left[\sup_{x\in I} V_n(x) \geq 2(\rho+2)\ln n\right] &\leq \sum_{i=1}^{N(n)} \mathbb{P}\left[\sup_{x\in I_i^{(n)}} V_n(x) \geq 2(\rho+2)\ln n\right] \\ &\leq \sum_{i=1}^{N(n)} \mathbb{P}\left[V_n\left(x_i^{(n)}\right) + 2 \geq 2(\rho+2)\ln n\right] \\ &\leq N(n) \sup_{x\in I} \mathbb{P}\left[V_n(x) \geq (\rho+2)\ln n\right]. \end{aligned}$$

Now, the application of Lemma 20 ensures that, for all $x \in I$,

$$\begin{aligned} \mathbb{P}\left[V_n(x) \geq (\rho+2)\ln n\right] &\leq \mathbb{P}\left[\lambda_n(x) M_n^{(n)}(x) - \Phi_{c_n(x)}(\lambda_n(x)) < M^{(n)} >_n (x) \geq \alpha_n(x)\lambda_n(x)\right] \\ &\leq \exp\left[-\alpha_n(x)\lambda_n(x)\right] \\ &\leq n^{-(\rho+2)}. \end{aligned}$$

It follows that

$$\sum_{n\geq 1} \mathbb{P}\left[\sup_{x\in I} V_n(x) \geq 2(\rho+2)\ln n\right] = O\left(\sum_{n\geq 1} n^{-2}\right) < +\infty,$$

and, applying Borel-Cantelli Lemma, we obtain

$$\sup_{x\in I}\lambda_n(x)M_n^{(n)}(x) \quad\leq\quad \sup_{x\in I}\Phi_{c_n(x)}(\lambda_n(x)) < M^{(n)} >_n (x) + 2(\rho+2)\ln n \quad a.s.$$

Since $\Phi_c(\lambda)\leq\lambda^2$ as soon as $\lambda c\leq 1$, and since $\lambda_n(x) = u_n\Pi_n(f(x))$, it follows that

$$u_n\sup_{x\in I}\Pi_n(f(x))M_n^{(n)}(x) \quad\leq\quad u_n^2\sup_{x\in I}\Pi_n^2(f(x)) < M^{(n)} >_n (x) + 2(\rho+2)\ln n \quad a.s.$$

Establishing the same upper bound for the martingale $\left(-M_k^{(n)}\right)$, we obtain

$$\sup_{x\in I}\Pi_n(f(x))\left|M_n^{(n)}(x)\right| \quad\leq\quad u_n\sup_{x\in I}\Pi_n^2(f(x)) < M^{(n)} >_n (x) + 2\frac{(\rho+2)\ln n}{u_n} \quad a.s.$$

Now, since $\sup_{x\in I} Var(Z_k(x)) = O\left(h_k^{-1}\right)$, we have

$$\begin{aligned}
&\sup_{x\in I}\Pi_n^2(f(x)) < M^{(n)} >_n (x)\\
&= O\left(\sum_{k=n_0}^{n}\sup_{x\in I}U_{k,n}^2(f(x))\gamma_k^2\sup_{x\in I}|r_{k-1}(x)-r(x)|^2\sup_{x\in I}(Var[Z_k(x)])\right)\\
&= O\left(\sum_{k=n_0}^{n}U_{k,n}^2(\varphi)\gamma_k^2h_k^{-1}w_k^2\right) \quad a.s. \qquad (3.76)
\end{aligned}$$

• Let us first consider the case when the sequence $(n\gamma_n)$ is bounded. In this case, (3.76) and Lemma 3 imply that

$$\begin{aligned}
\sup_{x\in I}\Pi_n^2(f(x)) < M^{(n)} >_n (x) &= O\left(\sum_{k=n_0}^{n}U_{k,n}^2(\varphi)\gamma_k m_k^2 w_k^2\right) \quad a.s.\\
&= O\left(m_n^2w_n^2\right) \quad a.s.
\end{aligned}$$

In this case, we have thus proved that, for all positive sequence (u_n) satisfying (3.75), we have

$$\sup_{x\in I}\Pi_n(f(x))\left|M_n^{(n)}(x)\right| = O\left(u_nm_n^2w_n^2+\frac{\ln n}{u_n}\right) \quad a.s. \qquad (3.77)$$

Now, since the sequence $\left(\left[m_n^{-1}w_n^{-1}\sqrt{\ln n}\right]m_nn^{-\epsilon}A_n\right)$ belongs to $\mathcal{GS}(-(w^*+\epsilon))$ with $w^*+\epsilon>0$, there exists $u_0>0$ such that, for all n,

$$u_0m_n^{-1}w_n^{-1}\sqrt{\ln n}\leq C^{-1}m_n^{-1}n^{\epsilon}A_n^{-1}$$

(where C is defined in (3.75)). Applying (3.77) with $(u_n) = \left(u_0m_n^{-1}w_n^{-1}\sqrt{\ln n}\right)$, we obtain

$$\sup_{x\in I}\Pi_n(f(x))\left|M_n^{(n)}(x)\right| \quad=\quad O\left(m_nw_n\sqrt{\ln n}\right) \quad a.s.,$$

which concludes the proof of Lemma 17 in this case.
• Let us now consider the case $\lim_{n\to\infty}(n\gamma_n)=\infty$. In this case, (3.76) implies that

$$\sup_{x\in I}\Pi_n^2(f(x)) < M >_n^{(n)} (x) \quad=\quad O\left(\gamma_nh_n^{-1}w_n^2\right) \quad a.s.$$

In this case, we have thus proved that, for all positive sequence (u_n) satisfying (3.75), we have

$$\sup_{x\in I}\Pi_n\left(f\left(x\right)\right)\left|M_n^{(n)}\left(x\right)\right| = O\left(u_n\gamma_n h_n^{-1}w_n^2+\frac{\ln n}{u_n}\right) \quad a.s. \tag{3.78}$$

Now, in view of (3.30), of (3.72), and of the assumptions of Lemma 17, we have

$$\begin{aligned}\left[\sqrt{\gamma_n^{-1}h_n\ln n}\,w_n^{-1}\right]\left[n\gamma_n^2h_n^{-2}A_n\right] &= O\left(\sqrt{\gamma_n h_n^{-1}\ln n}\,w_n^{-1}B_n\right)\\ &= O\left(1\right).\end{aligned}$$

Thus, there exists $u_0>0$ such that, for all n,

$$u_0\sqrt{\gamma_n^{-1}h_n\ln n}\,w_n^{-1}\leq C^{-1}n^{-1}\gamma_n^{-2}h_n^2A_n^{-1}.$$

Applying (3.78) with $(u_n)=\left(u_0\sqrt{\gamma_n^{-1}h_n\ln n}\,w_n^{-1}\right)$, we obtain

$$\begin{aligned}\sup_{x\in I}\Pi_n\left(f\left(x\right)\right)\left|M_n^{(n)}\left(x\right)\right| &= O\left(\sqrt{\gamma_n h_n^{-1}\ln n}\,w_n\right) \quad a.s.\\ &= O\left(m_nw_n\sqrt{\ln n}\right) \quad a.s.,\end{aligned}$$

which concludes the proof of Lemma 17.

Proof of Lemma 21

Let $(\delta_n)\in\mathcal{GS}(-\delta^*)$, set $x,y\in I$ such that $|x-y|\leq\delta_n$, and let c_i^* denote generic constants. Let us first note that

$$\begin{aligned}&\left|Z_k\left(x\right)-Z_k\left(y\right)\right|\leq c_1^*\delta_nh_k^{-2},\\ &\left|Z_k\left(x\right)\right|\leq c_2^*h_k^{-1},\\ &\left|Var\left[Z_k\left(x\right)\right]-Var\left[Z_k\left(y\right)\right]\right|\\ &\quad=\left|\mathbb{E}\left\{\left(Z_k(x)-Z_k(y)-\left[\mathbb{E}(Z_k(x))-\mathbb{E}(Z_k(y))\right]\right)\left(Z_k(x)+Z_k(y)-\left[\mathbb{E}(Z_k(x))+\mathbb{E}(Z_k(y))\right]\right)\right\}\right|\\ &\quad\leq 8c_2^*h_k^{-1}\mathbb{E}\left[\left|Z_k\left(x\right)-Z_k\left(y\right)\right|\right]\\ &\quad\leq c_3^*h_k^{-3}\delta_n,\end{aligned}$$

and that, in view of (3.21), on $\left\{\sup_{l\leq k}|Y_l|\leq A_n\right\}$,

$$\left|r_k\left(x\right)-r\left(x\right)\right|\leq c_4^*k\gamma_kh_k^{-1}A_n.$$

Now, in view of (3.10), we have

$$r_k\left(x\right)-r\left(x\right)=\left[1-\gamma_kZ_k\left(x\right)\right]\left[r_{k-1}\left(x\right)-r\left(x\right)\right]+\gamma_k\left[W_k\left(x\right)-r\left(x\right)Z_k\left(x\right)\right],$$

so that

$$\begin{aligned}&\left[r_k\left(x\right)-r\left(x\right)\right]-\left[r_k\left(y\right)-r\left(y\right)\right]\\ &\quad= \left[1-\gamma_kZ_k\left(x\right)\right]\left[r_{k-1}\left(x\right)-r\left(x\right)\right]-\left[1-\gamma_kZ_k\left(y\right)\right]\left[r_{k-1}\left(y\right)-r\left(y\right)\right]\\ &\qquad+\gamma_k\left(\left[W_k\left(x\right)-W_k\left(y\right)\right]-\left[r\left(x\right)Z_k\left(x\right)-r\left(y\right)Z_k\left(y\right)\right]\right)\\ &\quad= \left[1-\gamma_kZ_k\left(x\right)\right]\left(\left[r_{k-1}\left(x\right)-r\left(x\right)\right]-\left[r_{k-1}\left(y\right)-r\left(y\right)\right]\right)\\ &\qquad-\gamma_k\left[Z_k\left(x\right)-Z_k\left(y\right)\right]\left[r_{k-1}\left(y\right)-r\left(y\right)\right]\\ &\qquad+\gamma_k\left(\left[W_k\left(x\right)-W_k\left(y\right)\right]-r\left(x\right)\left[Z_k\left(x\right)-Z_k\left(y\right)\right]-Z_k\left(y\right)\left[r\left(x\right)-r\left(y\right)\right]\right).\end{aligned}$$

For $k \geq n_0$, we have $|1 - \gamma_k Z_k(x)| \leq 1$. It follows that, for $k \geq n_0$ and on $\left\{\sup_{l \leq k} |Y_l| \leq A_n\right\}$,

$$\begin{aligned}
&|[r_k(x) - r(x)] - [r_k(y) - r(y)]| \\
&\leq |[r_{k-1}(x) - r(x)] - [r_{k-1}(y) - r(y)]| + \gamma_k |Z_k(x) - Z_k(y)| |r_{k-1}(y) - r(y)| \\
&\quad + \gamma_k (|Y_k| + |r(x)|) |Z_k(x) - Z_k(y)| + \gamma_k |Z_k(y)| |r(x) - r(y)| \\
&\leq |[r_{k-1}(x) - r(x)] - [r_{k-1}(y) - r(y)]| + \left(c_5^* k \gamma_k^2 h_k^{-3} \delta_n A_n\right) + \left(c_6^* \gamma_k A_n h_k^{-2} \delta_n\right) + \left(c_7^* \gamma_k h_k^{-1} \delta_n\right) \\
&\leq |[r_{k-1}(x) - r(x)] - [r_{k-1}(y) - r(y)]| + c_8^* k \gamma_k^2 h_k^{-3} \delta_n A_n \\
&\leq c_9^* \sum_{j=1}^{k} j \gamma_j^2 h_j^{-3} \delta_n A_n \\
&\leq c_{10}^* k^2 \gamma_k^2 h_k^{-3} \delta_n A_n.
\end{aligned}$$

Moreover, we note that, on $\left\{\sup_{l \leq k} |Y_l| \leq A_n\right\}$,

$$\begin{aligned}
&\left|[r_k(x) - r(x)]^2 - [r_k(y) - r(y)]^2\right| \\
&\leq |[r_k(x) - r(x)] - [r_k(y) - r(y)]| \, |[r_k(x) - r(x)] + [r_k(y) - r(y)]| \\
&\leq c_{11}^* k^3 \gamma_k^2 h_k^{-4} \delta_n A_n^2.
\end{aligned}$$

Using all the previous upper bounds, as well as (3.59) with $p = 1$, we get

$$\begin{aligned}
&\left|\lambda_n(x) M_n^{(n)}(x) - \lambda_n(y) M_n^{(n)}(y)\right| \\
&= u_n \left| \sum_{k=n_0}^{n} U_{k,n}(f(x)) \gamma_k (\mathbb{E}[Z_k(x)] - Z_k(x)) [r_{k-1}(x) - r(x)] \right. \\
&\quad \left. - \sum_{k=n_0}^{n} U_{k,n}(f(y)) \gamma_k (\mathbb{E}[Z_k(y)] - Z_k(y)) [r_{k-1}(y) - r(y)] \right| \mathbb{1}_{\sup_{l \leq k-1} |Y_l| \leq A_n} \\
&\leq u_n \sum_{k=n_0}^{n} U_{k,n}(f(x)) \gamma_k |\mathbb{E}[Z_k(x)] - Z_k(x)| \, |(r_{k-1}(x) - r(x)) - (r_{k-1}(y) - r(y))| \, \mathbb{1}_{\sup_{l \leq k-1} |Y_l| \leq A_n} \\
&\quad + u_n \sum_{k=n_0}^{n} U_{k,n}(f(x)) \gamma_k |r_{k-1}(y) - r(y)| (|Z_k(x) - Z_k(y)| + \mathbb{E}[|Z_k(x) - Z_k(y)|]) \, \mathbb{1}_{\sup_{l \leq k-1} |Y_l| \leq A_n} \\
&\quad + u_n \sum_{k=n_0}^{n} U_{k,n}(f(y)) \left| \frac{U_{k,n}(f(x))}{U_{k,n}(f(y))} - 1 \right| \gamma_k |r_{k-1}(y) - r(y)| (|Z_k(x)| + \mathbb{E}[|Z_k(x)|]) \, \mathbb{1}_{\sup_{l \leq k-1} |Y_l| \leq A_n} \\
&\leq c_{12}^* u_n \sum_{k=n_0}^{n} U_{k,n}(\varphi) \gamma_k \left(h_k^{-1}\right) \left(k^2 \gamma_k^2 h_k^{-3} \delta_n A_n\right) + c_{13}^* u_n \sum_{k=n_0}^{n} U_{k,n}(\varphi) \gamma_k (k \gamma_k h_k^{-1} A_n) \left(\delta_n h_k^{-2}\right) \\
&\quad + c_{14}^* u_n \sum_{k=n_0}^{n} U_{k,n}(\varphi) \left(\delta_n \gamma_n^{-1}\right) \gamma_k (k \gamma_k h_k^{-1} A_n) \left(h_k^{-1}\right) \\
&\leq c_{15}^* u_n n^2 \gamma_n^2 h_n^{-4} \delta_n A_n + c_{16}^* u_n n \gamma_n h_n^{-3} \delta_n A_n + c_{17}^* u_n n h_n^{-2} \delta_n A_n
\end{aligned}$$

In view of (3.72) and (3.75), it follows that there exist $s_1^* > 0$ and $\tilde{S}_n^{(1)} \in \mathcal{GS}(s_1^*)$ such that

$$\left|\lambda_n(x) M_n^{(n)}(x) - \lambda_n(y) M_n^{(n)}(y)\right| \leq \delta_n \tilde{S}_n^{(1)}. \tag{3.79}$$

Now, we have

$$\begin{aligned}
&\Phi_{c_n(x)}\left(\lambda_n(x)\right) < M^{(n)} >_n (x) - \Phi_{c_n(y)}\left(\lambda_n(y)\right) < M^{(n)} >_n (y) \\
&= \frac{\Phi_{c_n(x)}\left(\lambda_n(x)\right)}{\lambda_n^2(x)} \lambda_n^2(x) < M^{(n)} >_n (x) - \frac{\Phi_{c_n(y)}\left(\lambda_n(y)\right)}{\lambda_n^2(y)} \lambda_n^2(y) < M^{(n)} >_n (y) \\
&= \frac{\Phi_{c_n(x)}\left(\lambda_n(x)\right)}{\lambda_n^2(x)} \left[\lambda_n^2(x) < M^{(n)} >_n (x) - \lambda_n^2(y) < M^{(n)} >_n (y)\right] \\
&\quad + \left[\frac{\Phi_{c_n(x)}\left(\lambda_n(x)\right)}{\lambda_n^2(x)} - \frac{\Phi_{c_n(y)}\left(\lambda_n(y)\right)}{\lambda_n^2(y)}\right] \lambda_n^2(y) < M^{(n)} >_n (y).
\end{aligned}$$

Since $c_n(x)\,\lambda_n(x) = c_n(y)\,\lambda_n(y) = \tilde{B}_n$, we have

$$\frac{\Phi_{c_n(x)}\left(\lambda_n(x)\right)}{\lambda_n^2(x)} = \tilde{B}_n^{-2}\left(\exp\left(\tilde{B}_n\right) - 1 - \tilde{B}_n\right) = \frac{\Phi_{c_n(y)}\left(\lambda_n(y)\right)}{\lambda_n^2(y)}.$$

Using the fact that $\Phi_c(\lambda) \leq \lambda^2$ for $\lambda c \leq 1$, and applying (3.59) with $p = 2$, we deduce that

$$\begin{aligned}
&\left|\Phi_{c_n(x)}\left(\lambda_n(x)\right) < M^{(n)} >_n (x) - \Phi_{c_n(y)}\left(\lambda_n(y)\right) < M^{(n)} >_n (y)\right| \\
&\leq \left|\lambda_n^2(x) < M^{(n)} >_n (x) - \lambda_n^2(y) < M^{(n)} >_n (y)\right| \\
&\leq u_n^2 \left|\sum_{k=n_0}^{n} U_{k,n}^2\left(f(x)\right)\gamma_k^2 Var\left[Z_k(x)\right]\left[r_{k-1}(x) - r(x)\right]^2\right. \\
&\quad \left. - \sum_{k=n_0}^{n} U_{k,n}^2\left(f(y)\right)\gamma_k^2 Var\left[Z_k(y)\right]\left[r_{k-1}(y) - r(y)\right]^2\right| \mathbb{1}_{\sup_{l\leq k-1}|Y_l|\leq A_n} \\
&\leq u_n^2 \sum_{k=n_0}^{n} U_{k,n}^2\left(f(x)\right)\gamma_k^2 Var\left[Z_k(x)\right]\left|\left(r_{k-1}(x) - r(x)\right)^2 - \left(r_{k-1}(y) - r(y)\right)^2\right| \mathbb{1}_{\sup_{l\leq k-1}|Y_l|\leq A_n} \\
&\quad + u_n^2 \sum_{k=n_0}^{n} U_{k,n}^2\left(f(x)\right)\gamma_k^2\left(r_{k-1}(y) - r(y)\right)^2\left|Var\left[Z_k(x)\right] - Var\left[Z_k(y)\right]\right| \mathbb{1}_{\sup_{l\leq k-1}|Y_l|\leq A_n} \\
&\quad + u_n^2 \sum_{k=n_0}^{n} U_{k,n}^2\left(f(y)\right)\left|\frac{U_{k,n}^2\left(f(x)\right)}{U_{k,n}^2\left(f(y)\right)} - 1\right| \gamma_k^2\left(r_{k-1}(y) - r(y)\right)^2 Var\left[Z_k(y)\right] \mathbb{1}_{\sup_{l\leq k-1}|Y_l|\leq A_n} \\
&\leq c_{18}^* u_n^2 \sum_{k=n_0}^{n} U_{k,n}^2(\varphi)\gamma_k^2\left(h_k^{-1}\right)\left(k^3\gamma_k^2 h_k^{-4}\delta_n A_n^2\right) + c_{19}^* u_n^2 \sum_{k=n_0}^{n} U_{k,n}^2(\varphi)\gamma_k^2\left(k^2\gamma_k^2 h_k^{-2} A_n^2\right)\left(\delta_n h_k^{-3}\right) \\
&\quad + c_{20}^* u_n^2 \sum_{k=n_0}^{n} U_{k,n}^2(\varphi)\left(\delta_n\gamma_n^{-1}\right)\gamma_k^2\left(k^2\gamma_k^2 h_k^{-2} A_n^2\right) h_k^{-1}
\end{aligned}$$

In view of (3.72) and (3.75), it follows that there exist $s_2^* > 0$ and $\tilde{S}_n^{(2)} \in \mathcal{GS}(s_2^*)$ such that

$$\left|\Phi_{c_n(x)}\left(\lambda_n(x)\right) < M^{(n)} >_n (x) - \Phi_{c_n(y)}\left(\lambda_n(y)\right) < M^{(n)} >_n (y)\right| \leq \delta_n \tilde{S}_n^{(2)}. \qquad (3.80)$$

Lemma 21 follows from the combination of (3.79) and (3.80).

Bibliographie

[1] Blum, J.R. (1954), Multidimensional stochastic approximation methods, *Ann. Math. Statist.,* **25**, 737-744

[2] Bojanic, R. and Seneta, E. (1973), A unified theory of regularly varying sequences, *Math.* Z., **134**, 91-106.

[3] Chen, H. (1988), Lower rate of convergence for locating a maximum of a function, *Ann. Statist.,* **16**, 1330-1334

[4] Chen, H.F., Duncan, T.E., and Pasik-Duncan, B. (1999), A Kiefer-Wolfowitz algorithm with randomized differences, *IEEE Trans. Automat. Control,* **44**, 442-453

[5] Delyon, B. and Juditsky, A.B. (1992), Stochastic optimization with averaging of trajectories, *Stochastics Stochastic Rep.,* **39**, 107-118.

[6] Dippon, J. and Renz J. (1996), Weighted means of processes in stochastic approximation, *Math. Meth. Statist.,* **5**, 32-60

[7] Dippon, J. and Renz J. (1997), Weighted means in stochastic approximation of minima, *SIAM J. Control Optim.,* **35**, 1811-1827

[8] Dippon, J. (2003), Accelerated randomized stochastic optimization, *Ann. Statist.,* **31**, 1260-1281

[9] Duflo, M. (1996), *Algorithmes stochastiques,* Collection mathématiques et applications, Springer

[10] Duflo, M. (1997), *Random Iterative Models*, Collection Applications of mathematics, Springer.

[11] Fabian, V. (1967), Stochastic approximation of minima with improved asymptotic speed, *Ann. Math. Statist.,* **38**, 191-200

[12] Galambos, J. and Seneta, E. (1973), Regularly varying sequences, *Proc. Amer. Math. Soc.*, **41**, 110-116.

[13] Hall, P. and Heyde, C.C. (1980), *Martingale limit theory and its application,* Academic Press, Inc., New York-London

[14] Hall, P. (1992), Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Ann. Statist.* **20**, 675-694

[15] Kiefer, J. and Wolfowitz, J. (1952), Stochastic estimation of the maximum of a regression function, *Ann. Math. Statist.* **23**, 462-466

[16] Kushner H.J. and Clark, D.S. (1978) *Stochastic approximation methods for constrained and unconstrained systems,* Springer, New York

[17] Kushner H.J. and Yang, J. (1993) Stochastic approximation with averaging : of the iterates : Optimal asymptotic rate of convergence for general processes *SIAM J. Control Optim.,* **31**, 1045-1062.

[18] Le Breton, A. (1993), About the averaging approach schemes for stochastic approximation, *Math. Methods Statist.,* **2**, 295-315.

[19] Le Breton, A. and Novikov, A. (1995), Some results about averaging in stochastic approximation, *Metrika,* **42**, 153-171

[20] Ljung, L. Pflug, G. and Walk, H. (1992) *Stochastic approximation and optimization of random systems,* Birkhäuser

[21] Mokkadem, A. and Pelletier, M. (2004), A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *submitted*

[22] Mokkadem, A. and Pelletier, M. (2006), Compact law of the iterated logarithm for matrix-normalized sums of random vectors, *to appear in Theory Probab. Appl.*

[23] Nadaraya, E. A. (1964), On estimating regression. *Theory Probab. Appl.* **10**, 186-190.

[24] Nevels'on, M.B. and Has'minskii , R.Z. (1976), *Stochastic approximation and recursive estimation,* Amer. Math. Soc, Providence, RI

[25] Pelletier, M. (2000), Asymptotic almost sure efficiency of averaged stochastic algorithms, *SIAM J. Control Optim.,* **39**, 49-72.

[26] Polyak, B.T. (1990), New method of stochastic approximation type, *Automat. Remote Control,* **51**, 937-946.

[27] Polyak, B.T. and Juditsky, A.B. (1992), Acceleration of stochastic approximation by averaging, *SIAM J. Control Optim.,* **30**, 838-855.

[28] Polyak, B.T. and Tsybakov, A.B. (1990), Optimal orders of accuracy for search algorithms of stochastic optimization, *Problems Inform. Transmission,* **26**, 126-133.

[29] Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.*, **8**, 391-398.

[30] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119-126.

[31] Ruppert, D. (1982), Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise, *Ann. of Probab.,* **10**, 178-187.

[32] Ruppert, D. (1988), Efficient estimators from a slowly converging Robbins-Monro process, *Tech. Rep. No.781, School of Oper. Res. and Ind. Engrg., Cornell University, Ithaca, NY.* (See also § 2.8 of Ruppert, D. (1991), *Stochastic approximation,* in Handbook of Sequential Analysis, Ghosh B.K. and Sen P.K., eds., Marcel Dekker, New York, 503-529).

[33] Spall, J.C. (1988), A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting. In *Proc. Conference on Decision and Control,* 1544-1548. IEEE, New York.

[34] Spall, J.C. (1997), A one-measurement form of simultaneous perturbation stochastic approximation, *Automatica J. IFAC,* **33**, 109-112.

[35] Watson, G. S. (1964), Smooth regression analysis. *Sankhya Ser. A.* **26**, 359-372.

[36] Yin, G. (1991), On extensions of Polyak's averaging approach to stochastic approximation, *Stochastics Stochastic Rep.,* **33**, 245-264.

Chapitre 4

How to apply the method of two-time-scale stochastic approximation in the estimation of a regression function

4.1 Introduction

The purpose of this chapter is to introduce two-time-scale stochastic approximation algorithms in order to define a class of recursive estimators of a regression function, and to study the asymptotic properties of the estimators built up in that way.

The use of stochastic approximation algorithms in the framework of regression estimation has been introduced by Kiefer and Wolfowitz (1952). The famous Kiefer and Wolfowitz algorithm allows the approximation of the point at which a regression function reaches its maximum. This pioneer work has been widely discussed and extended in many directions (see, among many others, Blum (1954), Fabian (1967), Kushner and Clark (1978), Hall and Heyde (1980), Ruppert (1982), Chen (1988), Spall (1988), Polyak and Tsybakov (1990), Dippon and Renz (1997), Spall (1997), Chen, Duncan and Pasik-Duncan (1999), Dippon (2003), and Mokkadem and Pelletier (2004)).

The question of applying the method of stochastic approximation to estimate a regression function at a given point (instead of approximating its mode) has been introduced by Révész (1973). His approach is the following.

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent, identically distributed pairs of random variables, and f denote the probability density of X. In order to construct a stochastic approximation algorithm for the estimation of the regression function $r : x \mapsto \mathbb{E}(Y|X = x)$ at a point x such that $f(x) \neq 0$, Révész (1973) defines an algorithm, which approximates the zero of the function $h\ : y \mapsto r(x)f(x) - f(x)y$. Following the Robbins-Monro procedure, this algorithm is defined by setting an arbitrary $\hat{r}_0(x) \in \mathbb{R}$ and, for $n \geq 1$,

$$\hat{r}_n(x) = \hat{r}_{n-1}(x) + \gamma_n W_n(x) \tag{4.1}$$

where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero, and $W_n(x)$ is an "observation" of the function h at the point $\hat{r}_{n-1}(x)$. To define $W_n(x)$, Révész introduces a kernel K (that is, a function such that $\int_{\mathbb{R}} K(x)dx = 1$) and a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and sets

$$W_n(x) = h_n^{-1} Y_n K(h_n^{-1}[x - X_n]) - h_n^{-1} K(h_n^{-1}[x - X_n])\hat{r}_{n-1}(x).$$

Révész (1977) gives the pointwise weak convergence rate, as well as an upper bound of the strong uniform convergence rate, of $\hat{r}_n$ towards r. The main drawback of his approach is that it requires the knowledge of a lower bound of the density f. More precisely, Révész (1977) chooses the stepsize (γ_n) equal to (n^{-1}), the bandwidth (h_n) equal to (n^{-a}) with $a \in]1/2, 1[$ and assume that $f(x) > (1-a)/2$ to establish a central limit theorem for $\hat{r}_n(x) - r(x)$, and that $\inf_{x \in I} f(x) > (1-a)/2$ (where I is a bounded interval) to prove an almost sure upper bound of $\sup_{x \in I} |\hat{r}_n(x) - r(x)|$.

The two-time-scale stochastic approximation algorithms have been introduced recently by Borkar (1997), Konda and Borkar (1999), Baras and Borkar (2000), Bhatnagar, Fu, Marcus and Fard (2001), Bhatnagar, Fu, and Marcus (2001), Konda and Tsitsiklis (2003), their convergence rate established by Konda and Tsitsiklis (2004) and Mokkadem and Pelletier (2006a). They are algorithms of search of the common zero (θ^*, μ^*) of two unknown functions h_1 and h_2. They are used in frameworks where the approximation of only one of the two parameters θ^* or μ^* is subject of interest, the approximation of the other parameter being made only to allow the approximation of the interesting parameter. They are built up in the following way : (i) θ_0 and μ_0 are arbitrarily set ; (ii) for $n \geq 1$, θ_n and μ_n are defined by the recursive equations

$$\theta_n = \theta_{n-1} + \gamma_n W_n^{(1)} \tag{4.2}$$
$$\mu_n = \mu_{n-1} + \beta_n W_n^{(2)} \tag{4.3}$$

where $W_n^{(1)}$ and $W_n^{(2)}$ are observations of $h_1(\theta_{n-1}, \mu_{n-1})$ and $h_2(\theta_{n-1}, \mu_{n-1})$ respectively, and where the stepsizes (γ_n) and (β_n) are two sequences of positive real numbers that go to zero with different rates. In the case both algorithms (4.2) and (4.3) are "of Robbins-Monro type", the first one is known to converge weakly with the rate $\gamma_n^{-1/2}$, whereas the weak convergence rate of the second one is $\beta_n^{-1/2}$. Thus, in the case the parameter of interest is μ^*, the stepsizes are chosen such that $\lim_{n\to\infty} \beta_n \gamma_n^{-1} = 0$.

In order to construct a two-time-scale stochastic approximation algorithm for the estimation of the regression function r at a point x such that $f(x) \neq 0$, we define an algorithm of search of the common zero of the functions

$$h_1 \ : (y,z) \mapsto f(x) - y \ \text{ and } \ h_2 \ : (y,z) \mapsto \frac{r(x)f(x)}{y} - z.$$

We thus proceed in the following way : (i) we set $f_0(x) = f_0 > 0$ and $r_0(x) \in \mathbb{R}$; (ii) for all $n \geq 1$, we define

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

where $W_n^{(1)}(x)$ and $W_n^{(2)}(x)$ are "observations" of the functions h_1 and h_2 at the point $(f_{n-1}(x), r_{n-1}(x))$. Following Révész's approach, we set

$$W_n^{(1)}(x) = h_n^{-1} K\left(\frac{x - X_n}{h_n}\right) - f_{n-1}(x) \ \text{ and } \ W_n^{(2)}(x) = h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right) \frac{1}{f_{n-1}(x)} - r_{n-1}(x).$$

The two-time-scale stochastic approximation algorithm we set up can thus be rewritten as :

$$f_n(x) = (1-\gamma_n) f_{n-1}(x) + \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right) \tag{4.4}$$

$$r_n(x) = (1-\beta_n) r_{n-1}(x) + \beta_n h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right) \frac{1}{f_{n-1}(x)}. \tag{4.5}$$

Since we have set $f_0(x) = f_0 > 0$, choosing the stepsize in (4.4) such that $\gamma_n \leq 1$ for all n and using a nonnegative kernel K ensures that the algorithm (4.5) is well defined for all $x \in \mathbb{R}$. To complete the definition of the algorithm (4.4)-(4.5), let us specify that, since the parameter of interest is the function regression r, we choose the stepsizes (γ_n) and (β_n) such that $\lim_{n\to\infty} \beta_n \gamma_n^{-1} = 0$.

We establish the pointwise weak and strong convergence rate of the estimator r_n defined by the two-time-scale stochastic algorithm (4.4)-(4.5) at any point x such that $f(x) \neq 0$. The optimal weak convergence rate of r_n is $n^{2/5}$, whereas the one of the estimator $\hat{r}_n$ defined in Révész (1977) is smaller than $n^{1/4}$. Moreover, the only condition we require to prove the convergence rate of r_n is $f(x) \neq 0$, instead of $f(x) > (1-a)/2$ as it is the case in Révész (1977). We also establish an upper bound of the strong convergence rate of $\sup_{x\in I} |r_n(x) - r(x)|$ for any bounded interval I on which $\inf_{x\in I} f(x) > 0$ (instead of on I such that $\inf_{x\in I} f(x) > (1-a)/2$, as in Révész (1977)).

To conclude this introduction, let us underline that the stochastic approximation algorithms (4.4) and (4.5) are not "of Robbins-Monro type". As a matter of fact, the use of the kernel K and of the bandwidth (h_n) to construct the "observations" $W_n^{(1)}(x)$ and $W_n^{(2)}(x)$ is, in a certain sense, very similar to the approximation of the differential of the regression function made in the construction of the Kiefer and Wolfowitz algorithm, and it induces behaviours for the algorithms (4.4) and (4.5) very comparable with the behaviour of the Kiefer and Wolfowitz algorithm. As a consequence, the results on the convergence rate of the two-time scale stochastic approximation algorithms obtained in Konda and Tsitsiklis (2004) or in Mokkadem and Pelletier (2006a) do not apply in the present framework.

Our main results are stated in Section 4.2, whereas Section 4.4 is devoted to the proofs.

4.2 Assumptions and main results

Let us first define the class of positive sequences that will be used in the statement of our assumptions.

Definition 3 *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*

$$\lim_{n\to\infty} n\left[1-\frac{v_{n-1}}{v_n}\right]=\gamma. \tag{4.6}$$

Condition (4.6) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1973)). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\ln n)^b$, $n^{\gamma}(\ln\ln n)^b$ and so on.

Let $g(s,t)$ denote the density of the couple (X,Y) (in particular $f(x)=\int_{\mathbb{R}} g(x,t)\,dt$), and set $a(x)=r(x)f(x)$. We establish the pointwise weak and strong convergence rate of the estimator r_n defined by the two-time-scale stochastic approximation algorithm (4.4)-(4.5) under the following assumptions.

(A1) $K:\mathbb{R}\to\mathbb{R}$ is a nonnegative, continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)\,dz=1$, $\int_{\mathbb{R}} zK(z)\,dz=0$ and $\int_{\mathbb{R}} z^2K(z)\,dz<\infty$.

(A2) *iii)* $(h_n)\in\mathcal{GS}(-a)$ with $a\in\left]0,\frac{1}{3}\right[$.
i) $(\beta_n)\in\mathcal{GS}(-\beta)$ with $\beta\in]3a,1]$.
ii) $(\gamma_n)\in\mathcal{GS}(-\alpha)$ with $\alpha\in]\min\{3a,a+1-\beta\},\beta]$. Moreover, $\gamma_n<1$ for all n and $\lim_{n\to\infty}\beta_n^{-1}\gamma_n\left(\ln\left(\sum_{k=1}^n\gamma_k\right)\right)^{-1}=\infty$.
iv) $\lim_{n\to\infty}(n\beta_n)\in]\min\{2a,(1-a)/2\},\infty]$.

(A3) *i)* $g(s,t)$ is two times continuously differentiable with respect to s.
ii) For $q\in\{0,1,2\}$, $s\mapsto\int_{\mathbb{R}} t^q g(s,t)\,dt$ is a bounded function continuous at $s=x$.
For $q\in[2,3]$, $s\mapsto\int_{\mathbb{R}}|t|^q g(s,t)\,dt$ is a bounded function.
iii) For $q\in\{0,1\}$, $\int_{\mathbb{R}}|t|^q\left|\frac{\partial g}{\partial x}(x,t)\right|dt<\infty$, and $s\mapsto\int_{\mathbb{R}} t^q\frac{\partial^2 g}{\partial s^2}(s,t)\,dt$ is a bounded function continuous at $s=x$.

Assumption $(A2)$ *iv)* on the limit of $(n\beta_n)$ as n goes to infinity is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\beta_n]^{-1}\right)$ is finite. Throughout this chapter we will use the following notations :

$$\xi = \lim_{n\to\infty}(n\beta_n)^{-1}$$

and for, $f(x)\neq 0$,

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

Theorem 10 (Weak convergence rate)
Let Assumptions $(A1)-(A3)$ hold for $x\in\mathbb{R}$ such that $f(x)\neq 0$.

1. *If there exists $c\geq 0$ such that $\beta_n^{-1}h_n^5\to c$, then*

$$\sqrt{\beta_n^{-1}h_n}\left(r_n(x)-r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{1-2a\xi}m^{(2)}(x),\frac{Var[Y|X=x]}{(2-(1-a)\xi)f(x)}\int_{\mathbb{R}}K^2(z)\,dz\right). \tag{4.7}$$

2. If $\beta_n^{-1}h_n^5 \to \infty$, then

$$\frac{1}{h_n^2}\left(r_n\left(x\right)-r\left(x\right)\right) \stackrel{\mathbb{P}}{\to} \frac{1}{\left(1-2a\xi\right)}m^{(2)}\left(x\right),$$

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

Remark 2 *In the case $\lim_{n\to\infty}\left(n\beta_n\right)=\infty$, we have $\xi=0$. Let us consider the case $\lim_{n\to\infty}\left(n\beta_n\right)<\infty$ (in which case $\beta=1$). In Part 2 of Theorem 10, we have $a\leq 1/5$, so that (A2) iv) implies that $\lim_{n\to\infty}\left(n\beta_n\right)>2a$, which guarantees that $1-2a\xi>0$. In Part 1, we have $a\geq 1/5$, (A2) iv) ensures that $\lim_{n\to\infty}\left(n\beta_n\right)>\left(1-a\right)/2$, so that $2-\left(1-a\right)\xi>0$. Moreover, in the case $c\neq 0$, we have $a=1/5$, so that the condition $\lim_{n\to\infty}\left(n\beta_n\right)>\left(1-a\right)/2$ ensured by (A2) iv) also guarantees that $1-2a\xi>0$.*

Let us underline that the convergence rate of the regression estimator r_n does not depend on the choice of the stepsize (γ_n) used in the stochastic approximation algorithm (4.4), which defines the density estimator f_n. This phenomenon is similar to the one, which happens in the framework of two-time-scale stochastic approximation algorithms.

For a given choice of stepsize (β_n), the optimal weak convergence rate of r_n is obtained by choosing the bandwidth (h_n) such that $\lim_{n\to\infty}\beta_n^{-1}h_n^5=c>0$. Now, choosing (β_n) satisfying $\lim_{n\to\infty}n\beta_n=\infty$ (in which case $\xi=0$) leads, whatever the bandwidth (h_n) is, to a convergence rate of r_n smaller than $n^{2/5}$. The advisable choice of stepsize is thus $(\beta_n)=(\beta^*n^{-1})$, where β^* must be set larger than $(1-a)/2$ if $a\geq 1/5$, and larger than $2a$ if $a<1/5$. The existence of such conditions on the stepsize when this one converges at the rate n is usual in the framework of stochastic approximation algorithms ; in our context (and unlike what often happens for stochastic approximation algorithms), these conditions are not tricky at all, since they only depend on the known parameter a.

In the case $(\beta_n)=(\beta^*n^{-1})$, $\lim_{n\to\infty}nh_n^5=c$, and $\beta^*>(1-a)/2$, we have $\xi=[\beta^*]^{-1}$, and (4.7) can be rewritten as

$$\sqrt{nh_n}\left(r_n\left(x\right)-r\left(x\right)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(\sqrt{c}\frac{\beta^*}{\beta^*-2a}m^{(2)}\left(x\right),\frac{\beta^{*2}}{\left(2\beta^*-\left(1-a\right)\right)}\frac{Var\left[Y|X=x\right]}{f\left(x\right)}\int_{\mathbb{R}}K^2\left(z\right)dz\right).$$

In particular, the optimal pointwise weak convergence rate of r_n is obtained by choosing $(h_n)=\left(cn^{-1/5}\right)$, and equals $n^{2/5}$. Let us recall that, to establish the pointwise weak convergence rate of the estimator $\hat{r}_n$ constructed via the single-time-scale stochastic approximation algorithm (4.1), Révész (1977) sets $(\gamma_n)=(n^{-1})$, $(h_n)=(n^{-a})$ with $a>1/2$, and proves the central limit theorem

$$\sqrt{nh_n}\left(\hat{r}_n\left(x\right)-r\left(x\right)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(0,\frac{Var\left[Y|X=x\right]f\left(x\right)}{\left(2f\left(x\right)-\left(1-a\right)\right)}\int_{\mathbb{R}}K^2\left(z\right)dz\right),$$

so that the convergence rate of $\hat{r}_n$ is smaller than $n^{1/4}$. Moreover, to prove this central limit theorem, Révész (1977) requires the condition $f(x)>(1-a)/2$; this condition is a typical example of the tedious conditions, which are usual in the framework of stochastic approximation algorithms when the stepsize converges at the rate n.

Let us now consider the problem of constructing confidence intervals for r. Hall (1992) shows that, to minimize the coverage error of the confidence interval, avoiding bias estimation by a slight undersmoothing is more efficient than explicit bias correction. We thus choose a bandwidth (h_n) such that $\lim_{n\to\infty}nh_n^5=0$. On the other hand, it seems interesting to minimize the asymptotic variance of r_n. The function $\beta^*\mapsto\beta^{*2}\left[2\beta-(1-a)\right]^{-1}$ reaching its minimum at the point $\beta^*=1-a$, we can state the following corollary.

Corollary 5
Let Assumptions $(A1)-(A3)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$, and choose $(h_n) \in \mathcal{GS}(-a)$ such that $\lim_{n\to\infty} nh_n^5 = 0$. To minimize the asymptotic variance of r_n, the stepsize must be chosen as $(\beta_n) \in \mathcal{GS}(-1)$ with $\lim_{n\to\infty} n\beta_n = 1-a$, and we then have

$$\sqrt{nh_n}\left(r_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1-a)\frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz\right). \tag{4.8}$$

Let us mention that the asymptotic variance of the estimator r_n defined in Corollary 5 is smaller than the one of the well-known (nonrecursive) Nadaraya-Watson kernel estimator

$$\tilde{r}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(h_n^{-1}(x - X_i)\right)}{\sum_{i=1}^n K\left(h_n^{-1}(x - X_i)\right)}, \tag{4.9}$$

whose asymptotic variance equals

$$\frac{Var[Y|X=x]}{f(x)}\int_{\mathbb{R}} K^2(z)\,dz.$$

To deduce confidence intervals from (4.8), it remains to estimate f (which can be done by using a recursive estimator defined in Chapter 1) and $Var[Y|X=x]$.

The following theorem gives the pointwise strong convergence rate of r_n.

Theorem 11 (Strong pointwise convergence rate)
Let Assumptions $(A1)-(A3)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $\beta_n^{-1}h_n^5/\ln\left(\sum_{k=1}^n \beta_k\right) \to c$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{\beta_n^{-1}h_n}{2\ln\left(\sum_{k=1}^n \beta_k\right)}}\left(r_n(x) - r(x)\right)\right)$$

is relatively compact and its limit set is the interval

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

2. *If $\beta_n^{-1}h_n^5/\ln\left(\sum_{k=1}^n \beta_k\right) \to \infty$, then, with probability one,*

$$\lim_{n\to\infty} \frac{1}{h_n^2}\left(r_n(x) - r(x)\right) = \frac{1}{1-2a\xi}m^{(2)}(x).$$

To establish a uniform strong upper bound of the convergence rate of r_n, we need the following additional assumptions.

$(A4)$ $i)$ K is Lipschitz-continuous.
$ii)$ There exists $t^* > 0$ such that $\mathbb{E}\left(\exp\left(t^*|Y|\right)\right) < \infty$.
$iii)$ $\alpha > \frac{3}{2}a + \frac{\beta}{2}$.
$iv)$ For $q \in \{0,1\}$, $x \mapsto \int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt$ is bounded on the set $\{x, f(x) > 0\}$.

Theorem 12 (Strong uniform convergence rate)
Let I be a bounded interval of $\mathbb{R}$ on which $\inf_{x\in I} f(x) > 0$ and let Assumptions $(A1)-(A4)$ hold for all $x \in I$.

1. *If $\left(\beta_n^{-1} h_n^5/[\ln n]^2\right)$ is bounded, then, with probability one,*

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

2. *If $\lim_{n\to\infty} \beta_n^{-1} h_n^5/[\ln n]^2 = \infty$, then, with probability one,*

$$\sup_{x\in I} |r_n(x) - r(x)| = O\left(h_n^2\right).$$

4.3 Simulations

The object of this section is to provide a simulations study comparing the (non recursive) Nadaraya-Watson regression estimator, the averaged Révész estimator (studied in Chapter 2) and the estimator computed with the help of a two-time-scale stochastic approximation algorithm. We consider the regression model

$$Y = r(X) + d\epsilon$$

where $d > 0$ and where ϵ is $\mathcal{N}(0,1)$ distibuted. Whatever the estimator is, we choose the kernel $K(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$, and the bandwidth equal to $(h_n) = \left(n^{-1/5}(\ln n)^{-1}\right)$ (which corresponds to a slight undersmoothing). The confidence intervals of $r(x)$ we consider are the following.

- When the Nadaraya-Watson estimator $\tilde{r}_n$ defined in (4.9) is used, we set

$$\tilde{I}_n = \left[\tilde{r}_n(x) - 1.96\sqrt{\frac{\sum_{i=1}^n (Y_i - \tilde{r}_n(X_i))^2}{n^2 h_n \tilde{f}_n(x)} \int_{\mathbb{R}} K^2(z)\,dz}, \right. \left. \tilde{r}_n(x) + 1.96\sqrt{\frac{\sum_{i=1}^n (Y_i - \tilde{r}_n(X_i))^2}{n^2 h_n \tilde{f}_n(x)} \int_{\mathbb{R}} K^2(z)\,dz}\right],$$

  where $\tilde{f}_n(x)$ is the density Rosenblatt estimator. The asymptotic confidence level of $\tilde{I}_n$ is 95%.
- To define the averaged Révész estimator $\bar{r}_n$, we choose the weights (q_n) equal to (h_n). This choice guarantees that (h_n) and (q_n) both belong to $\mathcal{GS}(-a)$ (with $a = 1/5$ here), so that the asymptotic variance of $\bar{r}_n$ is minimal. We also let $(\gamma_n) = \left(n^{-0.9}\right)$ (this choice being allowed by the assumptions of Chapter 2). Finally, we set

$$\bar{I}_n = \left[\bar{r}_n(x) - 1.96\sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{r}_n(X_i))^2}{n^2 h_n \hat{f}_n(x)} \int_{\mathbb{R}} K^2(z)\,dz}, \right. \left. \bar{r}_n(x) + 1.96\sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{r}_n(X_i))^2}{n^2 h_n \hat{f}_n(x)} \int_{\mathbb{R}} K^2(z)\,dz}\right],$$

  where $\hat{f}_n(x)$ is the recursive density estimator with minimum asymptotic variance; recall $\hat{f}_n$ is defined as in (4.4) but with the "fast" stepsize $(\gamma_n) = \left(4/5\, n^{-1}\right)$ (see Chapter 1 for the optimality of $\hat{f}_n$). Let Φ denote the distribution function of the standard normal; the asymptotic level of $\bar{I}_n$ is $2\Phi\left(1.96/\sqrt{4/5}\right) - 1 = 97.14\%$.

– For the estimator r_n defined by the two-time-scale stochastic approximation algorithm (4.4)-(4.5), we set $(\gamma_n) = \left(n^{-0.601}\right)$, and $(\beta_n) = \left(4/5\, n^{-1}\right)$ (which is the choice which leads to the minimum asymptotic variance of r_n). We set

$$\begin{aligned} I_n &= \left[r_n(x) - 1.96\sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - r_n(X_i)\right)^2}{n^2 h_n \hat{f}_n(x)} \int_{\mathbb{R}} K^2(z)\,dz}, \right. \\ &\qquad \left. r_n(x) + 1.96\sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - r_n(X_i)\right)^2}{n^2 h_n \hat{f}_n(x)} \int_{\mathbb{R}} K^2(z)\,dz} \right], \end{aligned}$$

where the density estimator $\hat{f}_n(x)$ is defined in the same way as for $\bar{I}_n$.

We consider three sample sizes $n = 50$, $n = 100$ and $n = 200$, three points $x = -0.5$, $x = 0$ and $x = 0.5$ and three densities of X : standard normal, normal mixture and student with 6 degrees of freedom. In each case the number of simulations is $N = 5000$. In each table, the first line corresponds to the Nadaraya-Watson estimator $\tilde{r}_n$ and gives the empirical level $\#\left\{r(x) \in \tilde{I}_n\right\}/N$; the second line corresponds to the averaged Révész estimator $\bar{r}_n$ and gives the empirical level $\#\left\{r(x) \in \bar{I}_n\right\}/N$; the third line corresponds to the estimator r_n constructed with the two-time-scale stochastic approximation algorithm and gives the empirical level $\#\left\{r(x) \in I_n\right\}/N$. For convenience, we recall the theoretical levels in the last column CL.

The simulations results confirm the theoretical ones : the coverage error of the intervals bluit up with the recursive estimators is smaller than the coverage error of the intervals bluit up with the Nadaraya-Watson estimators. Moreover it seems that the recursive algorithm defined in (4.4)-(4.5), with $(\beta_n) = \left(4/5\, n^{-1}\right)$ and $(\gamma_n) = \left(n^{-0.601}\right)$ works slightly better than the averaged Révész estimator $\bar{r}_n$ with $(q_n) = (h_n)$ and $(\gamma_n) = \left(n^{-0.9}\right)$.

Model $\mathbf{r}(\mathbf{x}) = \cos(\mathbf{x})$

$\mathbf{X} \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{1})$

$x = -0.5$			$x = 0$			$x = 0.5$			CL
$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	
				$d = 1$					
96.5%	96.76%	96.5%	96.44%	96.62%	96.84%	96.7%	97.04%	96.92%	95%
99.82%	99.9%	99.92%	99.8%	99.68%	99.76%	99.94%	99.86%	99.88%	97.14%
96.52%	99.34%	99.96%	97.28%	99.4%	99.98%	96.58%	99.12%	99.96%	97.14%
				$d = 2$					
95.42%	95.32%	95.7%	94.94%	95.44%	95.08%	95.4%	95.44%	96.2%	95%
99.82%	99.86%	99.76%	99.66%	99.6%	99.44%	99.82%	99.9%	99.98%	97.14%
97.38%	99.4%	99.92%	97.98%	99.28%	99.96%	97.72%	99.22%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \sin(\mathbf{2x}) + \mathbf{2}\exp\left(-\mathbf{16x^2}\right)$

$\mathbf{X} \to \mathcal{N}(\mathbf{0}, \mathbf{1})$

	$x=-0.5$			$x=0$			$x=0.5$		CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
98.5%	98.84%	99.16%	92.84%	96.42%	97.64%	99.34%	99.32%	99.3%	95%
88.4%	93.04%	96.62%	82.54%	89.94%	95%	99.98%	99.96%	99.98%	97.14%
98.14%	99.68%	99.96%	97.98%	99.42%	99.98%	96.26%	99.1%	99.92%	97.14%
				$d=2$					
96.16%	96.46%	97.26%	93.34%	94.92%	95.62%	96.5%	96.88%	96.72%	95%
96.74%	96.86%	97.84%	92.68%	94.02%	95.18%	99.96%	99.92%	99.84%	97.14%
97.88%	99.6%	99.98%	98.24%	99.72%	99.98%	97.64%	99.46%	99.96%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{0.3}\exp\left(-\mathbf{4}(\mathbf{x}+\mathbf{1})^{\mathbf{2}}\right) + \mathbf{0.7}\exp\left(-\mathbf{16}(\mathbf{x}-\mathbf{1})^{\mathbf{2}}\right)$

$\mathbf{X} \to \mathcal{N}(\mathbf{0}, \mathbf{1})$

	$x=-0.5$			$x=0$			$x=0.5$		CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
95.04%	94.74%	95.08%	95.06%	95.28%	95.4%	95.44%	95.44%	95.84%	95%
99.8%	99.62%	99.46%	99.24%	99.34%	99.06%	99.34%	99.34%	99.12%	97.14%
97.56%	99.52%	99.94%	98.58%	99.8%	99.98%	98.22%	99.52%	99.98%	97.14%
				$d=2$					
95.26%	95.14%	95.34%	94.74%	94.88%	95.06%	94.48%	95.56%	95.62%	95%
99.86%	99.76%	99.72%	99.64%	99.52%	99.38%	99.62%	99.74%	99.6%	97.14%
97.68%	99.68%	99.98%	98.5%	99.68%	99.98%	97.48%	99.68%	99.94%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{0.8} + \sin(\mathbf{6x})$

$\mathbf{X} \to \mathcal{N}(\mathbf{0}, \mathbf{1})$

	$x=-0.5$			$x=0$			$x=0.5$		CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
97.36%	97.54%	97.66%	97.2%	97.46%	97.68%	97.16%	97.72%	98.14%	95%
99.84%	99.8%	99.8%	99.84%	99.9%	99.76%	99.96%	99.86%	99.86%	97.14%
97.82%	99.68%	99.98%	98.34%	99.66%	99.96%	97.78%	99.36%	99.96%	97.14%
				$d=2$					
95.4%	95.86%	96.28%	95.36%	95.64%	96.06%	96%	95.98%	96.1%	95%
99.86%	99.86%	99.8%	99.72%	99.78%	99.74%	99.84%	99.66%	99.86%	97.14%
97.94%	99.62%	99.98%	98.62%	99.74%	99.96%	97.62%	99.5%	99.94%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{x} + \mathbf{2}\exp\left(-\mathbf{16x^2}\right)$

$\mathbf{X} \to \mathcal{N}(\mathbf{0}, \mathbf{1})$

	$x=-0.5$			$x=0$			$x=0.5$		CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
99.28%	99.68%	99.4%	96.4%	98.08%	99.16%	99.58%	99.68%	99.68%	95%
98.34%	99.02%	99.18%	89.5%	94.64%	97.7%	99.98%	99.98%	99.98%	97.14%
99.44%	99.96%	99.98%	99.36%	99.9%	99.98%	98.94%	99.82%	99.98%	97.14%
				$d=2$					
99.92%	97.1%	97.16%	94.62%	95.92%	97.3%	97.14%	97.52%	97.64%	95%
96.36%	99.04%	98.78%	94.78%	95.38%	96.88%	99.94%	99.96%	99.92%	97.14%
99.98%	99.82%	99.98%	98.76%	99.82%	99.98%	98.18%	99.72%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{1} + \mathbf{0.4x}$

$\mathbf{X} \to \mathcal{N}(\mathbf{0}, \mathbf{1})$

	$x=-0.5$			$x=0$			$x=-0.5$		CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
96.32%	95.94%	96.1%	96.24%	96.2%	96%	96.1%	96.24%	96.62%	95%
99.84%	99.9%	99.6%	99.92%	99.82%	99.72%	99.86%	99.8%	99.76%	97.14%
98.3%	99.76%	99.98%	98.7%	99.76%	99.98%	97.54%	99.5%	99.98%	97.14%
				$d=2$					
95.46%	94.76%	95.16%	95.56%	95.38%	95.54%	94.98%	94.96%	95.62%	95%
99.82%	99.88%	99.62%	99.88%	99.78%	99.68%	99.88%	99.82%	99.68%	97.14%
98.3%	99.66%	99.98%	98.66%	99.72%	99.98%	97.28%	99.48%	99.6%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \cos(\mathbf{x})$

$\mathbf{X} \to \mathbf{1/2}\mathcal{N}(-\mathbf{1/2}, \mathbf{1}) + \mathbf{1/2}\mathcal{N}(\mathbf{1/2}, \mathbf{1})$

	$x=-0.5$			$x=0$			$x=0.5$		CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
96.96%	97.06%	97.12%	97.26%	96.8%	97.1%	97.46%	96.94%	96.94%	95%
99.96%	99.92%	99.88%	99.86%	99.8%	99.66%	99.96%	99.96%	99.8%	97.14%
97.66%	99.68%	99.94%	98.18%	99.44%	99.96%	97.48%	99.5%	99.96%	97.14%
				$d=2$					
95.6%	95.32%	95.56%	95.08%	95.36%	95.64%	96.38%	95.7%	95.34%	95%
99.82%	99.92%	99.74%	99.94%	99.78%	99.64%	99.96%	99.9%	99.64%	97.14%
98.10%	99.7%	99.92%	98.44%	99.7%	99.98%	98.04%	99.6%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \sin(\mathbf{2x}) + \mathbf{2}\exp\left(-\mathbf{16x^2}\right)$

$$\mathbf{X} \rightarrow \mathbf{1/2}\mathcal{N}(-\mathbf{1/2}, \mathbf{1}) + \mathbf{1/2}\mathcal{N}(\mathbf{1/2}, \mathbf{1})$$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
98.12%	98.88%	98.18%	93.92%	96.52%	97.54%	99.08%	99.38%	99.26%	95%
86.8%	92.24%	94.94%	84.46%	89.02%	94.76%	99.98%	99.98%	99.98%	97.14%
98.92%	99.7%	99.98%	98.22%	99.46%	99.98%	97.18%	99.68%	99.98%	97.14%
				$d=2$					
96.26%	96.76%	96.82%	93.96%	95.12%	95.06%	96.84%	96.9%	97.08%	95%
96.88%	97.36%	97.38%	93.6%	94.4%	96.06%	99.94%	99.9%	99.88%	97.14%
98.48%	99.66%	99.96%	98.46%	99.62%	99.98%	97.72%	99.74%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{0.3}\exp\left(-\mathbf{4}(\mathbf{x}+\mathbf{1})^{\mathbf{2}}\right) + \mathbf{0.7}\exp\left(-\mathbf{16}(\mathbf{x}-\mathbf{1})^{\mathbf{2}}\right)$

$$\mathbf{X} \rightarrow \mathbf{1/2}\mathcal{N}(-\mathbf{1/2}, \mathbf{1}) + \mathbf{1/2}\mathcal{N}(\mathbf{1/2}, \mathbf{1})$$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
94.9%	95.38%	95.3%	95.56%	94.56%	94.86%	95.24%	95.24%	95.48%	95%
99.74%	99.62%	99.58%	99.44%	99.22%	99.1%	99.34%	99.28%	99.06%	97.14%
98.14%	99.72%	99.98%	98.6%	99.66%	99.98%	98.32%	99.7%	99.98%	97.14%
				$d=2$					
94.54%	95.34%	94.92%	95.2%	94.4%	94.82%	95.24%	95.06%	95.14%	95%
99.82%	99.78%	99.74%	99.84%	99.74%	99.6%	99.8%	99.78%	99.58%	97.14%
98.02%	99.68%	99.98%	98.58%	99.66%	99.98%	98.28%	99.68%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{0.8} + \sin(\mathbf{6x})$

$$\mathbf{X} \rightarrow \mathbf{1/2}\mathcal{N}(-\mathbf{1/2}, \mathbf{1}) + \mathbf{1/2}\mathcal{N}(\mathbf{1/2}, \mathbf{1})$$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
96.86%	97.38%	98.06%	97.22%	97.18%	97.56%	96.98%	97.56%	97.04%	95%
99.82%	99.8%	99.96%	99.96%	99.92%	99.86%	99.98%	99.96%	99.9%	97.14%
98.3%	99.82%	99.98%	98.5%	99.64%	99.96%	98.42%	99.64%	99.98%	97.14%
				$d=2$					
95.52%	95.82%	96.08%	95.56%	95.3%	95.62%	95.68%	95.82%	96.18%	95%
99.84%	99.82%	99.86%	99.98%	99.88%	99.76%	99.92%	99.88%	99.84%	97.14%
98.22%	99.78%	99.98%	98.54%	99.66%	99.96%	98.3%	99.62%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{x} + \mathbf{2}\exp\left(-\mathbf{16x^2}\right)$

$$\mathbf{X} \rightarrow \mathbf{1/2}\mathcal{N}\left(-\mathbf{1/2}, \mathbf{1}\right) + \mathbf{1/2}\mathcal{N}\left(\mathbf{1/2}, \mathbf{1}\right)$$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
99.4%	99.82%	99.76%	97.52%	98.84%	99.42%	99.88%	99.72%	99.92%	95%
99%	99.48%	99.64%	92.6%	96.08%	97.98%	99.96%	99.98%	99.98%	97.14%
99.66%	99.98%	99.98%	99.46%	99.96%	99.98%	99.48%	99.96%	99.98%	97.14%
				$d=2$					
97.1%	97.62%	98.02%	95.78%	96.64%	96.96%	97.88%	97.6%	97.68%	95%
99.3%	99.5%	99.5%	95.92%	96.42%	96.08%	99.96%	99.9%	99.98%	97.14%
98.8%	99.84%	99.98%	98.98%	99.88%	99.98%	98.92%	99.82%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{1} + \mathbf{0.4x}$

$$\mathbf{X} \rightarrow \mathbf{1/2}\mathcal{N}\left(-\mathbf{1/2}, \mathbf{1}\right) + \mathbf{1/2}\mathcal{N}\left(\mathbf{1/2}, \mathbf{1}\right)$$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
96.32%	96.66%	96.84%	96.46%	96.74%	96.64%	96.6%	96.72%	97.2%	95%
99.92%	99.88%	99.8%	99.94%	99.98%	99.84%	99.88%	99.9%	99.86%	97.14%
98.8%	99.98%	99.98%	98.9%	99.9%	99.98%	98%	99.82%	99.98%	97.14%
				$d=2$					
95.18%	95.46%	96.1%	95.08%	95.52%	95.6%	95.58%	95.44%	95.74%	95%
99.94%	99.86%	99.78%	99.88%	99.96%	99.7%	99.9%	99.86%	99.8%	97.14%
98.52%	99.76%	99.98%	98.76%	99.86%	99.98%	97.84%	99.76%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \cos(\mathbf{x})$

$$\mathbf{X} \rightarrow \mathcal{T}\left(\mathbf{6}\right)$$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
96.98%	97.54%	97.64%	97.02%	97.28%	97.52%	97.6%	97.1%	96.98%	95%
99.9%	99.84%	99.62%	99.74%	99.86%	99.88%	99.98%	99.9%	99.86%	97.14%
99.06%	99.92%	99.96%	99.24%	99.96%	99.98%	99.16%	99.96%	99.98%	97.14%
				$d=2$					
95.6%	95.96%	95.94%	95.4%	95.84%	96.06%	96.26%	95.62%	95.24%	95%
99.88%	99.78%	99.82%	99.74%	99.72%	99.8%	99.98%	99.82%	99.68%	97.14%
99.24%	99.92%	99.94%	99.52%	99.96%	99.98%	99.2%	99.96%	99.98%	97.14%

Model $\mathbf{r}\left(\mathbf{x}\right)=\sin\left(\mathbf{2x}\right)+\mathbf{2}\exp\left(-\mathbf{16x^2}\right)$

$\mathbf{X}\rightarrow\mathcal{T}\left(\mathbf{6}\right)$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
98.28%	98.74%	99.04%	93.48%	96.12%	97.7%	99.22%	99.18%	98.86%	95%
87.42%	92.18%	95.46%	83.7%	89.84%	94.84%	99.94%	99.96%	99.98%	97.14%
99.42%	99.96%	99.98%	99.46%	99.98%	99.98%	98.96%	99.94%	99.98%	97.14%
				$d=2$					
96.32%	96.56%	96.86%	94.14%	95.14%	96.04%	96.94%	96.48%	96.56%	95%
96.78%	97.2%	97.78%	93.66%	94.26%	95.42%	99.9%	99.9%	99.82%	97.14%
99.36%	99.92%	99.96%	99.58%	99.96%	99.98%	99.2%	99.98%	99.98%	97.14%

Model $\mathbf{r}\left(\mathbf{x}\right)=\mathbf{0.3}\exp\left(-\mathbf{4}\left(\mathbf{x}+\mathbf{1}\right)^{\mathbf{2}}\right)+\mathbf{0.7}\exp\left(-\mathbf{16}\left(\mathbf{x}-\mathbf{1}\right)^{\mathbf{2}}\right)$

$\mathbf{X}\rightarrow\mathcal{T}\left(\mathbf{6}\right)$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
95.3%	94.88%	95.08%	95.5%	95.06%	95.02%	95.28%	95.48%	95.56%	95%
99.8%	99.68%	99.46%	99.16%	99.26%	99.18%	99.4%	99.24%	99.18%	97.14%
99.46%	99.94%	99.96%	99.58%	99.94%	99.96%	99.48%	99.9%	99.96%	97.14%
				$d=2$					
94.88%	94.5%	94.8%	95.28%	94.8%	94.64%	95.06%	95.3%	95.3%	95%
99.84%	99.82%	99.58%	99.64%	99.66%	99.58%	99.8%	99.7%	99.7%	97.14%
99.44%	99.94%	99.96%	99.56%	99.94%	99.96%	99.5%	99.9%	99.98%	97.14%

Model $\mathbf{r}\left(\mathbf{x}\right)=\mathbf{0.8}+\sin\left(\mathbf{6x}\right)$

$\mathbf{X}\rightarrow\mathcal{T}\left(\mathbf{6}\right)$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
97.22%	97.36%	97.96%	97.16%	97.32%	98.02%	97.06%	97.5%	97.92%	95%
99.86%	99.8%	99.82%	99.92%	99.88%	99.94%	99.88%	99.92%	99.9%	97.14%
99.44%	99.98%	99.98%	99.56%	99.96%	99.98%	99.36%	99.92%	99.96%	97.14%
				$d=2$					
95.8%	95.58%	96.42%	95.6%	95.76%	96.22%	95.9%	95.74%	95.84%	95%
99.96%	99.78%	99.68%	99.9%	99.76%	99.76%	99.9%	99.9%	99.74%	97.14%
99.44%	99.96%	99.98%	99.5%	99.96%	99.98%	99.34%	99.94%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{x} + \mathbf{2}\exp\left(-\mathbf{16x^2}\right)$

$\mathbf{X} \rightarrow \mathcal{T}(\mathbf{6})$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
99.6%	99.84%	99.9%	97.6%	99.08%	99.58%	99.84%	99.86%	99.94%	95%
99.44%	99.58%	99.8%	93.46%	96.88%	98.82%	99.96%	99.96%	99.98%	97.14%
99.92%	99.98%	99.98%	99.86%	99.98%	99.98%	99.78%	99.98%	99.98%	97.14%
				$d=2$					
97.78%	97.84%	98.34%	95.54%	96.68%	97.72%	98.24%	98.06%	98.34%	95%
99.46%	99.36%	99.52%	95.54%	96.94%	97.48%	99.96%	99.98%	99.98%	97.14%
99.62%	99.98%	99.98%	99.76%	99.96%	99.98%	99.64%	99.94%	99.98%	97.14%

Model $\mathbf{r}(\mathbf{x}) = \mathbf{1} + \mathbf{0.4x}$

$\mathbf{X} \rightarrow \mathcal{T}(\mathbf{6})$

$x=-0.5$			$x=0$			$x=0.5$			CL
$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	
				$d=1$					
96.62%	97.04%	97%	97.2%	97.08%	97.02%	96.36%	97.14%	97.22%	95%
99.84%	99.9%	99.92%	99.94%	99.88%	99.82%	99.86%	99.84%	99.86%	97.14%
99.48%	99.98%	99.98%	99.66%	99.98%	99.96%	99.44%	99.98%	99.98%	97.14%
				$d=2$					
95.04%	95.62%	95.54%	95.96%	95.58%	95.88%	94.94%	96.14%	95.86%	95%
99.82%	99.9%	99.82%	99.84%	99.78%	99.66%	99.86%	99.84%	99.76%	97.14%
99.22%	99.92%	99.96%	99.7%	99.96%	99.98%	99.4%	99.96%	99.98%	97.14%

4.4 Proof of the results

Throughout the proofs, we use the following notations :

$$
\begin{aligned}
s_n &= \sum_{k=1}^{n} \gamma_k, \\
Q_n &= \prod_{j=1}^{n} \left(1-\gamma_j\right), \\
u_n &= \sum_{k=1}^{n} \beta_k, \\
\Pi_n &= \prod_{j=1}^{n} \left(1-\beta_j\right), \\
Z_n(x) &= h_n^{-1} K\left(\frac{x-X_n}{h_n}\right), \\
W_n(x) &= h_n^{-1} Y_n K\left(\frac{x-X_n}{h_n}\right).
\end{aligned}
$$

We first note that (4.5) can be rewritten as :

$$\begin{aligned}
&r_n(x) - r(x) \\
&= (1-\beta_n)(r_{n-1}(x) - r(x)) + \beta_n \left(\frac{W_n(x)}{f_{n-1}(x)} - r(x) \right) \\
&= \sum_{k=1}^{n} \prod_{j=k+1}^{n} (1-\beta_j) \beta_k \left(\frac{W_k(x)}{f_{k-1}(x)} - r(x) \right) + \prod_{j=1}^{n} (1-\beta_j)(r_0(x) - r(x)) \\
&= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left(\frac{W_k(x)}{f_{k-1}(x)} - r(x) \right) + \Pi_n (r_0(x) - r(x)) \\
&= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left(\frac{W_k(x) - r(x) f_{k-1}(x)}{f_{k-1}(x)} \right) + \Pi_n (r_0(x) - r(x)) \\
&= \frac{\Pi_n}{f(x)} \sum_{k=1}^{n} \Pi_k^{-1} \beta_k (W_k(x) - r(x) f_{k-1}(x)) \left[1 + \frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)} \right] + \Pi_n (r_0(x) - r(x)) \\
&= \frac{\Pi_n}{f(x)} \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left[W_k(x) - r(x) f_{k-1}(x) \right] \\
&\quad + \frac{\Pi_n}{f(x)} \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left[(W_k(x) - a(x)) - r(x)(f_{k-1}(x) - f(x)) \right] \left[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)} \right] \\
&\quad + \Pi_n (r_0(x) - r(x)). \qquad (4.10)
\end{aligned}$$

Now, in view of (4.4), we have

$$f_{k-1}(x) = Z_k(x) - \frac{1}{\gamma_k} \left[f_k(x) - f_{k-1}(x) \right].$$

Thus, relation (4.10) can be rewritten as

$$\begin{aligned}
r_n(x) - r(x) &= \frac{1}{f(x)} T_n(x) + \frac{r(x)}{f(x)} R_n^{(0)}(x) + \frac{1}{f(x)} R_n^{(1)}(x) + \frac{1}{f(x)} R_n^{(2)}(x) \\
&\quad + \frac{r(x)}{f(x)} R_n^{(3)}(x) + R_n^{(4)}(x)
\end{aligned}$$

where

$$\begin{aligned}
T_n(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left[W_k(x) - r(x) Z_k(x) \right], \\
R_n^{(0)}(x) &= \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1} \beta_k}{\gamma_k} \left[f_k(x) - f_{k-1}(x) \right], \\
R_n^{(1)}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left[W_k(x) - \mathbb{E}(W_k(x)) \right] \left[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)} \right], \\
R_n^{(2)}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left[\mathbb{E}(W_k(x)) - a(x) \right] \left[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)} \right], \\
R_n^{(3)}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left[\frac{(f_{k-1}(x) - f(x))^2}{f_{k-1}(x)} \right], \\
R_n^{(4)}(x) &= \Pi_n (r_0(x) - r(x)).
\end{aligned}$$

In order to prove Theorems 10-12, we show that the asymptotic behaviour of $r_n - r$ is given by the one of T_n, the terms $R_n^{(i)}$, $0 \leq i \leq 4$, being negligeable in front of T_n. More precisely, we establish the following lemmas 22-26 stated below.

Lemma 22 (Weak convergence rate of T_n)
Let Assumptions $(A1)-(A3)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $\beta_n^{-1} h_n^5 \to c$, then*

$$\sqrt{\beta_n^{-1} h_n} T_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{c^{\frac{1}{2}}}{1-2a\xi} m^{(2)}(x) f(x), \frac{Var[Y|X=x] f(x)}{(2-(\beta-a)\xi)} \int_{\mathbb{R}} K^2(z)\, dz \right).$$

2. *If $\beta_n^{-1} h_n^5 \to \infty$, then*

$$\frac{1}{h_n^2} T_n(x) \xrightarrow{\mathbb{P}} \frac{1}{1-2a\xi} m^{(2)}(x) f(x).$$

Remark 3 *Note that $(\beta - a)\xi \neq 0$ if and only if $\lim_{n\to\infty}(n\beta_n) = \beta^* \in]0, \infty[$, in which case $\beta = 1$. The variance of the Gaussian law in Part 1 of Lemma 22 is thus equal to the one, which appears in Part 1 of Theorem 10. We keep the parameter β in the statement of Lemma 22 to help the reader throughout the proofs. The same remark can be made for the following lemma.*

Lemma 23 (Strong pointwise convergence rate of T_n)
Let Assumptions $(A1)-(A3)$ hold for $x \in \mathbb{R}$ such that $f(x) \neq 0$.

1. *If there exists $c \geq 0$ such that $\beta_n^{-1} h_n^5 / \ln(u_n) \to c$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{\beta_n^{-1} h_n}{2 \ln(u_n)}} T_n(x) \right)$$

is relatively compact and its limit set is the interval

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

2. *If $\beta_n^{-1} h_n^5 / \ln(u_n) \to \infty$, then, with probability one,*

$$\lim_{n\to\infty} \frac{1}{h_n^2} T_n(x) = \frac{1}{1-2a\xi} m^{(2)}(x) f(x).$$

Lemma 24 (Strong uniform convergence rate of T_n)
Let I be a bounded interval on which $\inf_{x\in I} f(x) > 0$ and let Assumptions $(A1)-(A4)$ hold for all $x \in I$.

1. *If $\left(\beta_n^{-1} h_n^5 / [\ln n]^2\right)$ is bounded, then, with probability one,*

$$\sup_{x\in I} |T_n(x)| = O\left(\sqrt{\beta_n h_n^{-1}} \ln n \right).$$

2. *If* $\lim_{n\to\infty}\beta_n^{-1}h_n^5/[\ln n]^2=\infty$*, then, with probability one,*

$$\sup_{x\in I}|T_n(x)|=O\left(h_n^2\right).$$

Lemma 25 (Strong pointwise convergence rate of $R_n^{(i)}$)
Let Assumptions $(A1)-(A3)$ *hold for* $x\in\mathbb{R}$*. For* $i\in\{0,1,2,3\}$*, we have*

$$R_n^{(i)}(x)=o\left(\max\left\{\sqrt{\beta_n h_n^{-1}},h_n^2\right\}\right)\quad a.s.$$

Lemma 26 (Uniform strong convergence rate of $R_n^{(i)}$)
Let Assumptions $(A1)-(A4)$ *hold for all* $x\in I$*. For* $i\in\{0,1,2,3\}$*, we have*

$$\sup_{x\in I}\left|R_n^{(i)}(x)\right|=o\left(\max\left\{\sqrt{\beta_n h_n^{-1}}\ln n,h_n^2\right\}\right)\quad a.s.$$

Theorem 10 (respectively Theorem 11) follows from the combination of Lemma 22 (respectively Lemma 23) and Lemma 25. Theorem 12 is a consequence of the combination of Lemmas 24 and 26. Before giving the proofs of Lemmas 22-26, let us state two lemmas, which are proved in the first chapter.

Lemma 27 (Technical Lemma)
Let $(v_n)\in\mathcal{GS}(v^*)$*,* $(\alpha_n)\in\mathcal{GS}(-\alpha^*)$ *with* $\alpha^*>0$*, and set* $m>0$*. If* $m-v^*\lim_{n\to\infty}(n\alpha_n)^{-1}>0$*, then*

$$\lim_{n\to\infty}v_n\left[\prod_{j=1}^n(1-\alpha_j)^m\right]\sum_{k=1}^n\left[\prod_{j=1}^k(1-\alpha_j)^{-m}\right]\frac{\alpha_k}{v_k}=\frac{1}{m-v^*\lim_{n\to\infty}(n\alpha_n)^{-1}}.$$

Moreover, for all positive sequence (u_n) *such that* $\lim_{n\to\infty}u_n=0$*, and all* C*,*

$$\lim_{n\to\infty}v_n\left[\prod_{j=1}^n(1-\alpha_j)^m\right]\left[\sum_{k=1}^n\left[\prod_{j=1}^k(1-\alpha_j)^{-m}\right]\frac{\alpha_k}{v_k}u_k+C\right]=0.$$

Lemma 28 (Convergence rate of the density estimator defined by (4.4))

1. *Under Assumptions* $(A1)$*-*$(A3)$*, we have*

$$|f_n(x)-f(x)|\;=\;O\left(\max\left\{\sqrt{\frac{\gamma_n\ln s_n}{h_n}},h_n^2\right\}\right)\quad a.s.$$

2. *Let Assumptions* $(A1)-(A4)$ *hold, and let* I *be a bounded interval of* $\mathbb{R}$*; we have*

$$\sup_{x\in I}|f_n(x)-f(x)|=O\left(\max\left\{\sqrt{\gamma_n h_n^{-1}}\ln n,h_n^2\right\}\right)\quad a.s.$$

Our proofs are now organized as follows. Lemmas 22, 23 and 24 are proved in Sections 4.4.1, 4.4.2 and 4.4.3 respectively. Section 4.4.4 is devoted to the proof of Lemmas 25 and 26.

4.4.1 Proof of Lemma 22

We establish that

- if $a \geq \beta/5$, then
$$\sqrt{\beta_n^{-1}h_n}\left(T_n(x) - \mathbb{E}\left(T_n(x)\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{Var\left[Y|X=x\right]f(x)}{\left(2-(\beta-a)\xi\right)}\int_{\mathbb{R}} K^2(z)\,dz\right), \qquad (4.11)$$
- if $a > \beta/5$, then $\sqrt{\beta_n^{-1}h_n}\mathbb{E}\left(T_n(x)\right) \to 0,$ (4.12)
- if $a \leq \beta/5$, then $h_n^{-2}\mathbb{E}\left(T_n(x)\right) \to \dfrac{1}{1-2a\xi}m^{(2)}(x)\,f(x),$ (4.13)
- if $a < \beta/5$, then $h_n^{-2}\left(T_n(x) - \mathbb{E}\left(T_n(x)\right)\right) \xrightarrow{\mathbb{P}} 0.$ (4.14)

As a matter of fact the combination of (4.11) and (4.12) (respectively of (4.11) and (4.13)) gives Part 1 of Lemma 22 in the case $a > \beta/5$ (respectively $a = \beta/5$), the one of (4.13) and (4.14) (respectively of (4.11) and (4.13)) gives Part 2 of Lemma 22 in the case $a < \beta/5$ (respectively $a = \beta/5$). We prove (4.11), (4.14), (4.13), and (4.12) successively.

Proof of (4.11) Set

$$\eta_k(x) = \Pi_k^{-1}\beta_k\left[W_k(x) - r(x)\,Z_k(x)\right], \qquad (4.15)$$

so that $T_n(x) - \mathbb{E}\left(T_n(x)\right) = \Pi_n\sum_{k=1}^n\left[\eta_k(x) - \mathbb{E}\left(\eta_k(x)\right)\right]$. We have

$$Var\left(\eta_k(x)\right) = \Pi_k^{-2}\beta_k^2\left[Var\left(W_k(x)\right) + r^2(x)\,Var\left(Z_k(x)\right) - 2r(x)\,Cov\left(W_k(x), Z_k(x)\right)\right].$$

In view of $(A3)$, classical computations give

$$\begin{aligned} Var\left(W_k(x)\right) &= \frac{1}{h_k}\left[\mathbb{E}\left[Y^2|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right], \qquad (4.16)\\ Var\left(Z_k(x)\right) &= \frac{1}{h_k}\left[f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right],\\ Cov\left(W_k(x), Z_k(x)\right) &= \frac{1}{h_k}\left[r(x)\,f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right]. \end{aligned}$$

It follows that

$$Var\left(\eta_k(x)\right) = \frac{\Pi_k^{-2}\beta_k^2}{h_k}\left[Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right], \qquad (4.17)$$

and, since $\lim_{n\to\infty}(n\beta_n) > (1-a)/2$, Lemma 27 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n Var\left(\eta_k(x)\right)\\ &= \sum_{k=1}^n \frac{\Pi_k^{-2}\beta_k^2}{h_k}\left[Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right]\\ &= \frac{1}{\Pi_n^2}\frac{\beta_n}{h_n}\frac{1}{2-(\beta-a)\xi}\left[Var\left[Y|X=x\right]f(x)\int_{\mathbb{R}}K^2(z)\,dz + o(1)\right]. \qquad (4.18) \end{aligned}$$

For all $p \in]0,1]$, and in view of $(A3)$, we have

$$\begin{aligned}
&\mathbb{E}\left(|Y_k - r(x)|^{2+p} K^{2+p}\left(\frac{x - X_k}{h_k}\right)\right) \\
&= h_k \int_{\mathbb{R}^2} |y - r(x)|^{2+p} K^{2+p}(s)\, g(x - h_k s, y)\, dy ds \\
&\leq 2^{1+p} h_k \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} |y|^{2+p} g(x - h_k s, y)\, dy + |r(x)|^{2+p} \int_{\mathbb{R}} g(x - h_k s, y)\, dy \right\} K^{2+p}(s)\, ds \\
&= O(h_k). \qquad (4.19)
\end{aligned}$$

Now, set $p \in]0,1]$ such that $\lim_{n\to\infty}(n\beta_n) > \frac{1+p}{2+p}(1-a)$. Applying Lemma 27, we get

$$\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}\left[|\eta_k(x)|^{2+p}\right] &= O\left(\sum_{k=1}^{n} \frac{\Pi_k^{-2-p}\beta_k^{2+p}}{h_k^{2+p}} \mathbb{E}\left[|Y_k - r(x)|^{2+p} K^{2+p}\left(\frac{x - X_k}{h_k}\right)\right]\right) \\
&= O\left(\sum_{k=1}^{n} \frac{\Pi_k^{-2-p}\beta_k^{2+p}}{h_k^{1+p}}\right) \\
&= O\left(\frac{1}{\Pi_n^{2+p}} \frac{\beta_n^{1+p}}{h_n^{1+p}}\right).
\end{aligned}$$

Using (4.18), we deduce that

$$\begin{aligned}
\frac{1}{v_n^{2+p}} \sum_{k=1}^{n} \mathbb{E}\left[|\eta_k(x)|^{2+p}\right] &= O\left(\left(\frac{\beta_n}{h_n}\right)^{\frac{p}{2}}\right) \\
&= o(1),
\end{aligned}$$

and (4.11) follows by application of Lyapounov Theorem.

Proof of (4.14) In view of (4.17), since $a < \beta/5$, and since $\lim_{n\to\infty}(n\beta_n) > 2a$, the application of Lemma 27 gives

$$\begin{aligned}
Var(T_n(x)) &= \Pi_n^2 \sum_{k=1}^{n} \frac{\Pi_k^{-2}\beta_k^2}{h_k} \left[Var[Y|X=x]\, f(x) \int_{\mathbb{R}} K^2(z)\, dz + o(1)\right] \\
&= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\beta_k o\left(h_k^4\right) \\
&= o\left(h_n^4\right),
\end{aligned}$$

and (4.14) follows.

Proof of (4.13) We have

$$\mathbb{E}(T_n(x)) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k \left[(\mathbb{E}(W_k(x)) - a(x)) - r(x)(\mathbb{E}(Z_k(x)) - f(x))\right].$$

In view of $(A3)$ we obtain

$$\begin{aligned}
\mathbb{E}(W_k(x)) - a(x) &= \frac{1}{2} h_k^2 \int_{\mathbb{R}} y \frac{\partial^2 g}{\partial x^2}(x,y)\, dy\, [1 + o(1)] \int_{\mathbb{R}} z^2 K(z)\, dz, \\
\mathbb{E}(Z_k(x)) - f(x) &= \frac{1}{2} h_k^2 \int_{\mathbb{R}} \frac{\partial^2 g}{\partial x^2}(x,y)\, dy\, [1 + o(1)] \int_{\mathbb{R}} z^2 K(z)\, dz.
\end{aligned}$$

Since $\lim_{n\to\infty}(n\beta_n) > 2a$, it follows from the application of Lemma 27 that

$$\begin{aligned}
\mathbb{E}(T_n(x)) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k h_k^2 \left[\frac{1}{2}\left(\int_{\mathbb{R}} y\frac{\partial^2 g}{\partial x^2}(x,y)\,dy - r(x)\int_{\mathbb{R}}\frac{\partial^2 g}{\partial x^2}(x,y)\,dy\right) + o(1)\right]\int_{\mathbb{R}} z^2 K(z)\,dz \\
&= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k h_k^2 \left[m^{(2)}(x)\,f(x) + o(1)\right] \\
&= \frac{1}{1-2a\xi} h_n^2 \left[m^{(2)}(x)\,f(x) + o(1)\right].
\end{aligned}$$

which gives (4.13).

Proof of (4.12) Since $a > \beta/5$ and $\lim_{n\to\infty}(n\beta_n) > (1-a)/2$, we have

$$\begin{aligned}
\mathbb{E}(T_n(x)) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o\left(\sqrt{\beta_k h_k^{-1}}\right) \\
&= o\left(\sqrt{\beta_n h_n^{-1}}\right).
\end{aligned}$$

which gives (4.12).

4.4.2 Proof of Lemma 23

Set

$$S_n(x) = \sum_{k=1}^{n}\left[\eta_k(x) - \mathbb{E}(\eta_k(x))\right]$$

where η_k is defined in (4.15).

• Let us first consider the case $a \geq \beta/5$ (in which case $\lim_{n\to\infty}(n\beta_n) > (\beta-a)/2$). We set $H_n^2 = \Pi_n^2\beta_n^{-1}h_n$, and note that, since $\left(\beta_n^{-1}h_n\right) \in \mathcal{GS}(\beta-a)$, we have, by setting $\beta_0 = h_0 = 1$,

$$\begin{aligned}
\ln\left(H_n^{-2}\right) &= -2\ln(\Pi_n) + \ln\left(\prod_{k=1}^{n}\frac{\beta_{k-1}^{-1}h_{k-1}}{\beta_k^{-1}h_k}\right) \\
&= -2\sum_{k=1}^{n}\ln(1-\beta_k) + \sum_{k=1}^{n}\ln\left(1 - \frac{\beta-a}{k} + o\left(\frac{1}{k}\right)\right) \\
&= \sum_{k=1}^{n}(2\beta_k + o(\beta_k)) - \sum_{k=1}^{n}((\beta-a)\xi\beta_k + o(\beta_k)) \\
&= (2-\xi(\beta-a))\,u_n + o(u_n).
\end{aligned} \tag{4.20}$$

Since $2-\xi(\beta-a) > 0$, it follows in particular that $\lim_{n\to\infty} H_n^{-2} = \infty$. Moreover, we clearly have $\lim_{n\to\infty} H_n^2/H_{n-1}^2 = 1$, and by (4.18)

$$\lim_{n\to\infty} H_n^2 \sum_{k=1}^{n} Var[\eta_k(x)] = [2-(\beta-a)\xi]^{-1}\, Var[Y|X=x]\, f(x)\int_{\mathbb{R}} K^2(z)\,dz.$$

In view of (4.19), $\mathbb{E}\left[|\eta_n|^3\right] = O\left(\Pi_n^{-3}\beta_n^3 h_n^{-2}\right)$. Now, since $\lim_{n\to\infty}(n\beta_n) > (\beta - a)/2$, applying Lemma 27 and (4.20), we get

$$\begin{aligned}\frac{1}{n\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\left(|H_n\eta_k|^3\right) &= O\left(\frac{H_n^3}{n\sqrt{n}}\left(\sum_{k=1}^{n}\frac{\Pi_k^{-3}\beta_k^3}{h_k^2}\right)\right)\\ &= O\left(\frac{H_n^3}{n\sqrt{n}}\left(\sum_{k=1}^{n}\Pi_k^{-3}\beta_k o\left(\beta_k^{\frac{3}{2}}h_k^{-\frac{3}{2}}\right)\right)\right)\\ &= o\left(\frac{H_n^3}{n\sqrt{n}}\Pi_n^{-3}\beta_n^{\frac{3}{2}}h_n^{-\frac{3}{2}}\right)\\ &= o\left(\frac{1}{n\sqrt{n}}\right)\\ &= o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right).\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier (2006b) then ensures that, with probability one, the sequence

$$\left(\frac{H_nS_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right) = \left(\frac{\sqrt{\beta_n^{-1}h_n}\left(T_n(x) - \mathbb{E}\left(T_n(x)\right)\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\frac{Var\left[Y|X=x\right]f(x)}{(2-(\beta-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz},\sqrt{\frac{Var\left[Y|X=x\right]f(x)}{(2-(\beta-a)\xi)}\int_{\mathbb{R}}K^2(z)\,dz}\right]. \tag{4.21}$$

In view of (4.20), we have $\lim_{n\to\infty}\ln\ln\left(H_n^{-2}\right)/\ln u_n = 1$. It follows that, with probability one, the sequence $\left(\sqrt{\beta_n^{-1}h_n}\left(T_n(x) - \mathbb{E}\left(T_n(x)\right)\right)/\sqrt{2\ln u_n}\right)$ is relatively compact, and its limit set is the interval given in (4.21). The application of (4.12) (respectively of (4.13)) concludes the proof of Lemma 23 in the case $a > \beta/5$ (respectively $a = \beta/5$).

• Let us now consider the case $a < \beta/5$ (in which case $\lim_{n\to\infty}(n\beta_n) > 2a$). Set $H_n^{-2} = \Pi_n^{-2}h_n^4\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{-1}$ and note that, since $\left(h_n^{-4}\right) \in \mathcal{GS}(4a)$, we have

$$\begin{aligned}\ln\left(\Pi_n^{-2}h_n^4\right) &= -2\ln\left(\Pi_n\right) + \ln\left(\prod_{k=1}^{n}\frac{h_{k-1}^{-4}}{h_k^{-4}}\right)\\ &= -2\sum_{k=1}^{n}\ln\left(1-\beta_k\right) + \sum_{k=1}^{n}\ln\left(1 - \frac{4a}{k} + o\left(\frac{1}{k}\right)\right)\\ &= \sum_{k=1}^{n}\left(2\beta_k + o\left(\beta_k\right)\right) - \sum_{k=1}^{n}\left(4a\xi\beta_k + o\left(\beta_k\right)\right)\\ &= \left(2 - 4a\xi\right)u_n + o\left(u_n\right).\end{aligned} \tag{4.22}$$

Since $2 - 4a\xi > 0$, it follows in particular that $\lim_{n\to\infty}\Pi_n^{-2}h_n^4 = \infty$, and thus $\lim_{n\to\infty}H_n^{-2} = \infty$. Moreover, we clearly have $\lim_{n\to\infty}H_n^2/H_{n-1}^2 = 1$, $\mathbb{E}\left[|\eta_n|^3\right] < \infty$. Now, set $\epsilon \in]0, \beta - 5a[$ such that

$\lim_{n\to\infty}(n\beta_n) > 2a+\epsilon/2$; in view of (4.18), and applying Lemma 27, we get

$$\begin{aligned}
H_n^2\sum_{k=1}^n Var\left[\eta_k(x)\right] &= O\left(\Pi_n^2 h_n^{-4}\ln\ln\left(\Pi_n^{-2}h_n^4\right)\sum_{k=1}^n\frac{\Pi_k^{-2}\beta_k^2}{h_k}\right)\\
&= O\left(\Pi_n^2 h_n^{-4}\ln\ln\left(\Pi_n^{-2}h_n^4\right)\sum_{k=1}^n\Pi_k^{-2}\beta_k o\left(h_k^4k^{-\epsilon}\right)\right)\\
&= o\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)n^{-\epsilon}\right)\\
&= o(1).
\end{aligned}$$

Moreover, applying (4.22) and Lemma 27, we obtain

$$\begin{aligned}
\frac{1}{n\sqrt{n}}\sum_{k=1}^n\mathbb{E}\left(\left|H_n\eta_k\right|^3\right) &= O\left(\frac{\Pi_n^3h_n^{-6}}{n\sqrt{n}}\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{\frac{3}{2}}\left(\sum_{k=1}^n\frac{\Pi_k^{-3}\beta_k^3}{h_k^2}\right)\right)\\
&= O\left(\frac{\Pi_n^3h_n^{-6}}{n\sqrt{n}}\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{\frac{3}{2}}\left(\sum_{k=1}^n\Pi_k^{-3}\beta_k o\left(h_k^6\right)\right)\right)\\
&= o\left(\frac{\Pi_n^3h_n^{-6}}{n\sqrt{n}}\Pi_n^{-3}h_n^6\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{\frac{3}{2}}\right)\\
&= o\left(\frac{\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{\frac{3}{2}}}{n\sqrt{n}}\right)\\
&= o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right).
\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier (2006b) then ensures that, with probability one,

$$\lim_{n\to\infty}\frac{H_nS_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}} = \lim_{n\to\infty}h_n^{-2}\frac{\sqrt{\ln\ln\left(\Pi_n^{-2}h_n^4\right)}}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\left(T_n(x)-\mathbb{E}\left(T_n(x)\right)\right)=0.$$

Noting that (4.22) ensures that $\lim_{n\to\infty}\ln\ln\left(H_n^{-2}\right)/\ln\ln\left(\Pi_n^{-2}h_n^4\right)=1$, we get

$$\lim_{n\to\infty}h_n^{-2}\left[T_n(x)-\mathbb{E}\left(T_n(x)\right)\right]=0\quad a.s.,$$

and Lemma 23 in the case $a<\beta/5$ follows from (4.13).

4.4.3 Proof of Lemma 24

Let us write $T_n(x)$ as

$$T_n(x)=T_n^{(1)}(x)-r(x)T_n^{(2)}(x)$$

with

$$\begin{aligned}
T_n^{(1)}(x) &= \Pi_n\sum_{k=1}^n\Pi_k^{-1}\beta_kh_k^{-1}Y_kK\left(\frac{x-X_k}{h_k}\right)\\
T_n^{(2)}(x) &= \Pi_n\sum_{k=1}^n\Pi_k^{-1}\beta_kh_k^{-1}K\left(\frac{x-X_k}{h_k}\right).
\end{aligned}$$

Lemma 24 is proved by showing that, for $i \in \{1,2\}$,

- if $a \geq \beta/5$, then $\sup_{x\in I}\left|T_n^{(i)}(x) - \mathbb{E}\left(T_n^{(i)}(x)\right)\right| = O\left(\sqrt{\beta_n h_n^{-1}}\ln n\right) \quad a.s.,$ (4.23)
- if $a > \beta/5$, then $\sup_{x\in I}\left|\mathbb{E}\left(T_n(x)\right)\right| = o\left(\sqrt{\beta_n h_n^{-1}}\ln n\right),$ (4.24)
- if $a < \beta/5$, then $\sup_{x\in I}\left|T_n^{(i)}(x) - \mathbb{E}\left(T_n^{(i)}(x)\right)\right| = o\left(h_n^2\right) \quad a.s.,$ (4.25)
- if $a \leq \beta/5$, then $\sup_{x\in I}\left|\mathbb{E}\left(T_n(x)\right)\right| = O\left(h_n^2\right).$ (4.26)

As a matter of fact, lemma 24 follows from the combination of (4.23) and (4.24) in the case $a > \beta/5$, from the one of (4.23) and (4.26) in the case $a = \beta/5$, and from the one of (4.25) and (4.26) in the case $a < \beta/5$.

The proof of (4.24) and (4.26) is similar to the one of (4.12) and (4.13) and is omitted. Moreover the proof of (4.23) and (4.25) for $i = 2$ is similar to the one for $i = 1$, and is omitted too. To prove simultaneously (4.23) and (4.25) for $i = 1$, we introduce the sequence (v_n) defined as

$$(v_n) = \begin{cases} \left(\sqrt{\beta_n^{-1} h_n}\right) & \text{if} \quad a \geq \beta/5, \\ \left(h_n^{-2}\left[\ln n\right]^2\right) & \text{if} \quad a < \beta/5. \end{cases} \tag{4.27}$$

As a matter of fact, $(v_n) \in \mathcal{GS}(v^*)$ with $v^* = \min\left\{\frac{\beta - a}{2}, 2a\right\}$, and to prove that (4.23) and (4.25) for $i = 1$, we establish that

$$\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| = O\left(v_n^{-1}\ln n\right) \quad a.s. \tag{4.28}$$

Lemma 29 *There exists $s > 0$ such that, for all $C > 0$,*

$$\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C\right] = O\left(n^{-\frac{C}{s}}\right).$$

We first show how (4.28) can be deduced from Lemma 29, and then prove Lemma 29. Set $(M_n) \in \mathcal{GS}(m^*)$ with $m^* > 0$, and note that, for all $C > 0$, we have

$$\begin{aligned} &\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C\right] \\ &\leq \mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C \text{ and } \sup_{k\leq n}|Y_k| \leq M_n\right] \\ &\quad + \mathbb{P}\left[\sup_{k\leq n}|Y_k| \geq M_n\right]. \end{aligned} \tag{4.29}$$

Set $\rho_n = \rho_0 M_n^{-1} v_n^{-1} h_n^2$ where ρ_0 is chosen such that $\rho_n < 1$ for all n. One can choose $N(n)$ intervals $I_i^{(n)}$, $i \in \{1, \ldots, N(n)\}$, of length ρ_n and such that $\cup_{i=1}^{N(n)} I_i^{(n)} = I$. For all $i \in \{1, \ldots, N(n)\}$, set $x_i^{(n)} \in I_i^{(n)}$. We have

$$\begin{aligned} &\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C \text{ and } \sup_{k\leq n}|Y_k| \leq M_n\right] \\ &\leq \sum_{i=1}^{N(n)}\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I_i^{(n)}}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C \text{ and } \sup_{k\leq n}|Y_k| \leq M_n\right]. \end{aligned}$$

Since K is Lipschitz-continuous, and by application of Lemma 27, there exist $k^*, c^* > 0$, such that, for all $x, y \in \mathbb{R}$, satisfying $|x-y| \leq \rho_n$ and on $\left\{\sup_{k\leq n}|Y_k| \leq M_n\right\}$, we have :

$$\begin{aligned}\left|T_n^{(1)}(x) - T_n^{(1)}(y)\right| &= \left|\Pi_n \sum_{k=1}^n \Pi_k^{-1}\beta_k h_k^{-1} Y_k \left[K\left(\frac{x-X_k}{h_k}\right) - K\left(\frac{y-X_k}{h_k}\right)\right]\right| \\ &\leq k^* M_n \Pi_n \sum_{k=1}^n \Pi_k^{-1}\beta_k h_k^{-2}\rho_n \\ &\leq c^* M_n h_n^{-2}\rho_n.\end{aligned}$$

It follows that, for all $x \in I_i^{(n)}$, on $\left\{\sup_{k\leq n}|Y_k| \leq M_n\right\}$, we have

$$\begin{aligned}&\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \\ &\leq \left|T_n^{(1)}(x) - T_n^{(1)}\left(x_i^{(n)}\right)\right| + \left|T_n^{(1)}\left(x_i^{(n)}\right) - \mathbb{E}\left(T_n^{(1)}\left(x_i^{(n)}\right)\right)\right| + \left|\mathbb{E}\left(T_n^{(1)}\left(x_i^{(n)}\right)\right) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \\ &\leq 2c^* M_n h_n^{-2}\rho_n + \left|T_n^{(1)}\left(x_i^{(n)}\right) - \mathbb{E}\left(T_n^{(1)}\left(x_i^{(n)}\right)\right)\right|.\end{aligned}$$

In view of (4.29), we obtain, for all $C > 0$,

$$\begin{aligned}&\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C\right] \\ &\leq \sum_{i=1}^{N(n)} \mathbb{P}\left[\frac{v_n}{\ln n}\left|T_n^{(1)}\left(x_i^{(n)}\right) - \mathbb{E}\left(T_n^{(1)}\left(x_i^{(n)}\right)\right)\right| + 2c^* M_n \frac{v_n}{\ln n} h_n^{-2}\rho_n \geq C\right] \\ &\quad + n\mathbb{P}\left[|Y| \geq M_n\right].\end{aligned}$$

Now, for n large enough, we have $2c^* M_n v_n (\ln n)^{-1} h_n^{-2}\rho_n \leq C/2$. In view of Lemma 29 and Assumption $(A4)\, ii)$, there exists $s > 0$ such that, for n large enough,

$$\begin{aligned}&\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C\right] \\ &\leq N(n)\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq \frac{C}{2}\right] + n\exp\left(-t^* M_n\right)\mathbb{E}\left(\exp\left(t^*|Y|\right)\right) \\ &= O\left(\rho_n^{-1} n^{-\frac{C}{2s}} + n\exp\left(-t^* M_n\right)\right).\end{aligned}$$

Since $N(n) = O\left(\rho_n^{-1}\right)$ with $\left(\rho_n^{-1}\right) \in \mathcal{GS}(v^* + 2a + m^*)$, and since $(M_n) \in \mathcal{GS}(m^*)$ with $m^* > 0$, we can choose C large enough so that

$$\sum_{n\geq 2}\mathbb{P}\left[\frac{v_n}{\ln n}\sup_{x\in I}\left|T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right| \geq C\right] < \infty,$$

and (4.28) follows.
It remains to prove Lemma 29. For all $x \in I$ and all $s > 0$, we have

$$\begin{aligned}&\mathbb{P}\left[\frac{v_n}{\ln n}\left(T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right) \geq C\right] \\ &= \mathbb{P}\left[\exp\left[s^{-1} v_n\left(T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right)\right] \geq n^{\frac{C}{s}}\right] \\ &\leq n^{-\frac{C}{s}}\mathbb{E}\left(\exp\left[s^{-1} v_n\left(T_n^{(1)}(x) - \mathbb{E}\left(T_n^{(1)}(x)\right)\right)\right]\right) \\ &\leq n^{-\frac{C}{s}}\prod_{k=1}^n \mathbb{E}\left(\exp\left(s^{-1} U_{k,n}(x)\right)\right) \qquad (4.30)\end{aligned}$$

with

$$U_{k,n}(x) = v_n \Pi_n \Pi_k^{-1} \beta_k h_k^{-1} \left[Y_k K \left(\frac{x - X_k}{h_k} \right) - \mathbb{E} \left(Y_k K \left(\frac{x - X_k}{h_k} \right) \right) \right].$$

For k and n such that $k \leq n$, set

$$\alpha_{k,n} = v_n \Pi_n \Pi_k^{-1} \beta_k h_k^{-1}.$$

We have, for all $x \in I$,

$$\begin{aligned}
& \mathbb{E}\left(\exp\left[s^{-1}U_{k,n}(x)\right]\right) \\
\leq\ & 1 + \frac{1}{2}\mathbb{E}\left[s^{-2}U_{k,n}^2(x)\right] + \mathbb{E}\left[s^{-3}\left|U_{k,n}^3(x)\right|\right]\exp\left[s^{-1}\left|U_{k,n}(x)\right|\right] \\
\leq\ & 1 + \frac{1}{2}s^{-2}\alpha_{k,n}^2 Var\left[Y_k K\left(\frac{x - X_k}{h_k}\right)\right] \\
& + s^{-3}\alpha_{k,n}^3\|K\|_\infty^3 \mathbb{E}\left[\left(|Y_k|^3 + (\mathbb{E}(|Y_k|))^3\right)\exp\left(s^{-1}\alpha_{k,n}\|K\|_\infty\left(|Y_k| + \mathbb{E}(|Y_k|)\right)\right)\right].
\end{aligned}$$

Now, note that $\alpha_{k,n}$ can be rewritten as :

$$\alpha_{k,n} = \frac{v_n \Pi_n}{v_k \Pi_k} v_k \beta_k h_k^{-1}.$$

Since $(v_n) \in \mathcal{GS}(v^*)$ with $v^* = \min\left\{\frac{\beta - a}{2}, 2a\right\}$, we have $1 - v^*\xi > 0$, and thus

$$\begin{aligned}
\frac{\Pi_n}{\Pi_{n-1}} \frac{v_n}{v_{n-1}} &= (1 - \beta_n)\left(1 + \frac{v^*}{n} + o\left(\frac{1}{n}\right)\right) \\
&= 1 - (1 - v^*\xi)\beta_n + o(\beta_n) \\
&\leq 1 \qquad \text{for } n \text{ large enough.}
\end{aligned}$$

Writting

$$\frac{v_n \Pi_n}{v_k \Pi_k} = \prod_{i=k}^{n-1} \frac{v_{i+1}\Pi_{i+1}}{v_i \Pi_i},$$

we obtain

$$\sup_n \sup_{k \leq n} \frac{v_n \Pi_n}{v_k \Pi_k} < \infty.$$

Moreover, in view of (4.27), we have $\lim_{k\to\infty}\left(v_k \beta_k h_k^{-1}\right) = 0$. It follows that $\sup_{\substack{k,n \\ k \leq n}} \alpha_{k,n} < \infty$. Consequently, in view of Assumption $(A4)\, ii)$, there exist $s > 0$ and $c^* > 0$ such that, for all k and n, $(k \leq n)$,

$$\mathbb{E}\left[\left(|Y_k|^3 + (\mathbb{E}(|Y_k|))^3\right)\exp\left(s^{-1}\alpha_{k,n}\|K\|_\infty\left(|Y_k| + \mathbb{E}(|Y_k|)\right)\right)\right] \leq c^*.$$

From classical computation we have $\sup_{x\in I} Var\left[Y_k K\left((x - X_k)h_k^{-1}\right)\right] = O(h_k)$. Now let $C_1^*, C_2^* > 0$ be generic constants that may vary from line to line. Noting that, in view of (4.27), the sequence $\left(\beta_n h_n^{-1} v_n^2\right)$ is bounded, we obtain, for all $x \in I$, and for all k, n, such that $k \leq n$,

$$\begin{aligned}
& \mathbb{E}\left(\exp\left[s^{-1}U_{k,n}(x)\right]\right) \\
\leq\ & 1 + C_1^* v_n^2 \Pi_n^2 \Pi_k^{-2} \beta_k^2 h_k^{-1} + C_2^* v_n^3 \Pi_n^3 \Pi_k^{-3} \beta_k^3 h_k^{-3} \\
\leq\ & \exp\left[C_1^* v_n^2 \Pi_n^2 \Pi_k^{-2} \beta_k v_k^{-2}\left(\beta_k h_k^{-1} v_k^2\right) + C_2^* v_n^3 \Pi_n^3 \Pi_k^{-3} \beta_k v_k^{-3}\left(\left[\beta_k h_k^{-1} v_k^2\right]^{\frac{3}{2}} \beta_k^{\frac{1}{2}} h_k^{-\frac{3}{2}}\right)\right] \\
\leq\ & \exp\left[C_1^* v_n^2 \Pi_n^2 \Pi_k^{-2} \beta_k v_k^{-2} + C_2^* v_n^3 \Pi_n^3 \Pi_k^{-3} \beta_k v_k^{-3}\right].
\end{aligned}$$

Applying Lemma 27, we deduce from (4.30) that, for all $C > 0$,

$$\begin{aligned}
&\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left(T_n^{(1)}(x)-\mathbb{E}\left(T_n^{(1)}(x)\right)\right)\geq C\right]\\
&\leq\ n^{-\frac{C}{s}}\exp\left[C_1^* v_n^2\Pi_n^2\sum_{k=1}^n \Pi_k^{-2}\beta_k v_k^{-2}+C_2^* v_n^3\Pi_n^3\sum_{k=1}^n \Pi_k^{-3}\beta_k v_k^{-3}\right]\\
&=\ O\left(n^{-\frac{C}{s}}\right).
\end{aligned}$$

We establish exactly in the same way that, for all $C > 0$,

$$\sup_{x\in I}\mathbb{P}\left[\frac{v_n}{\ln n}\left(\mathbb{E}\left(T_n^{(1)}(x)\right)-T_n^{(1)}(x)\right)\geq C\right] = O\left(n^{-\frac{C}{s}}\right),$$

which concludes the proof of Lemma 29.

4.4.4 Proof of Lemmas 25 and 26

Let (m_n) and $(\tilde{m}_n)$ be the sequences defined as :

$$(m_n) = \begin{cases}\left(\sqrt{\beta_n h_n^{-1}}\right) & \text{if } \lim_{n\to\infty}\left(\beta_n h_n^{-5}\right)=\infty,\\ \left(h_n^2\right) & \text{otherwise}\end{cases} \tag{4.31}$$

$$(\tilde{m}_n) = \begin{cases}\left(\sqrt{\beta_n h_n^{-1}}\ln n\right) & \text{if } \lim_{n\to\infty}\left(\beta_n h_n^{-5}\ln n\right)=\infty,\\ \left(h_n^2\right) & \text{otherwise}\end{cases} \tag{4.32}$$

and note that (m_n) and $(\tilde{m}_n)$ are in $\mathcal{GS}(-m^*)$ with $m^* = \min\left\{\frac{\beta-a}{2}, 2a\right\}$. Lemmas 25 and 26 are proved by establishing that $\left|R_n^{(i)}(x)\right| = o(m_n)$ a.s. and that $\sup_{x\in I}\left|R_n^{(i)}(x)\right| = o(\tilde{m}_n)$ a.s. respectively. We prove Lemmas 25 and 26 for $i=4$, $i=0$, $i=2$, and $i=3$ in Sections 4.4.4.a, 4.4.4.b, 4.4.4.c, and 4.4.4.d respectively. Section 4.4.4.e is devoted to the proof of Lemma 25 for $i=1$, and Section 4.4.4.f to the one of Lemma 26 for $i=1$.

4.4.4.a Proof of Lemmas 25 and 26 for $i = 4$

We have $\sup_{x\in I}\left|R_n^{(4)}(x)\right| = o(\Pi_n)$ a.s. Since $\left(m_n^{-1}\right)\in\mathcal{GS}(m^*)$ with $1-m^*\xi > 0$, Lemma 27 ensures that

$$\Pi_n = o(m_n), \tag{4.33}$$

which proves Lemmas 25 and 26 for $i = 4$.

4.4.4.b Proof of Lemmas 25 and 26 for $i = 0$

We have

$$\begin{aligned}
R_n^{(0)}(x) &= \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k} \left[(f_k(x) - f(x)) - (f_{k-1}(x) - f(x))\right] \\
&= \Pi_n \sum_{k=1}^{n-1} \left(\frac{\Pi_k^{-1}\beta_k}{\gamma_k} - \frac{\Pi_{k+1}^{-1}\beta_{k+1}}{\gamma_{k+1}}\right)(f_k(x) - f(x)) \\
&\quad + \frac{\beta_n}{\gamma_n}(f_n(x) - f(x)) - \Pi_n \frac{\beta_1}{(1-\beta_1)\gamma_1}(f_0(x) - f(x)), \\
&= \Pi_n \sum_{k=1}^{n-1} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}\left[1 - \frac{\Pi_k}{\Pi_{k+1}}\frac{\beta_{k+1}}{\beta_k}\frac{\gamma_k}{\gamma_{k+1}}\right](f_k(x) - f(x)) \\
&\quad + \frac{\beta_n}{\gamma_n}(f_n(x) - f(x)) - \Pi_n \frac{\beta_1}{(1-\beta_1)\gamma_1}(f_0(x) - f(x)),
\end{aligned}$$

Since $\left(\beta_k^{-1}\gamma_k\right) \in \mathcal{GS}(\beta - \alpha)$, we have

$$\begin{aligned}
\left[1 - \frac{\Pi_k}{\Pi_{k+1}}\frac{\beta_{k+1}}{\beta_k}\frac{\gamma_k}{\gamma_{k+1}}\right] &= 1 - \frac{1}{(1-\beta_{k+1})}\left(\frac{\beta_k^{-1}\gamma_k}{\beta_{k+1}^{-1}\gamma_{k+1}}\right) \\
&= 1 - (1 + \beta_{k+1} + o(\beta_{k+1}))\left(1 - \frac{(\beta-\alpha)}{k} + o\left(\frac{1}{k}\right)\right) \\
&= O(\beta_k).
\end{aligned}$$

It follows that

$$\left|R_n^{(0)}(x)\right| = O\left(\Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k^2}{\gamma_k}|f_k(x) - f(x)| + \frac{\beta_n}{\gamma_n}|f_n(x) - f(x)| + \Pi_n\right). \qquad (4.34)$$

The application of the first part of Lemma 28 gives :

$$\begin{aligned}
&\left|R_n^{(0)}(x)\right| \\
&= O\left(\Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k^2}{\gamma_k}\left[\left(\frac{\gamma_k \ln(s_k)}{h_k}\right)^{\frac{1}{2}} + h_k^2\right] + \frac{\beta_n}{\gamma_n}\left[\left(\frac{\gamma_n \ln(s_n)}{h_n}\right)^{\frac{1}{2}} + h_n^2\right] + \Pi_n\right) \quad a.s. \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k\left(\beta_k\gamma_k^{-1}\ln(s_k)\right)^{\frac{1}{2}}\sqrt{\beta_k h_k^{-1}} + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k\left(\beta_k\gamma_k^{-1}\right)h_k^2\right. \\
&\quad \left. + \left(\beta_n\gamma_n^{-1}\ln(s_n)\right)^{\frac{1}{2}}\sqrt{\beta_n h_n^{-1}} + \left(\beta_n\gamma_n^{-1}\right)h_n^2 + \Pi_n\right) \quad a.s.
\end{aligned}$$

In view of Assumption $(A2)\, ii)$, we obtain

$$\begin{aligned}
&\left|R_n^{(0)}(x)\right| \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o\left(\sqrt{\beta_k h_k^{-1}}\right) + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o\left(h_k^2\right) + o\left(\sqrt{\beta_n h_n^{-1}}\right) + o\left(h_n^2\right) + \Pi_n\right) \quad a.s. \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o(m_k)\right) + o(m_n) + O(\Pi_n) \quad a.s.
\end{aligned}$$

In view of (4.33) and applying Lemma 27, we get

$$\left|R_n^{(0)}(x)\right| \;=\; o(m_n) \quad a.s.,$$

which proves Lemma 25 for $i = 0$. Now, in view of (4.34), the application of the second part of Lemma 28 implies that :

$$\begin{aligned}
&\sup_{x\in I}\left|R_n^{(0)}(x)\right| \\
&= O\left(\Pi_n\sum_{k=1}^{n}\frac{\Pi_k^{-1}\beta_k^2}{\gamma_k}\left[\left(\frac{\gamma_k}{h_k}\right)^{\frac{1}{2}}\ln k + h_k^2\right] + \frac{\beta_n}{\gamma_n}\left[\left(\frac{\gamma_n}{h_n}\right)^{\frac{1}{2}}\ln n + h_n^2\right] + \Pi_n\right) \quad a.s. \\
&= O\left(\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left(\beta_k\gamma_k^{-1}\right)^{\frac{1}{2}}\sqrt{\beta_k h_k^{-1}}\ln k + \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left(\beta_k\gamma_k^{-1}\right)h_k^2\right. \\
&\quad \left. + \left(\beta_n\gamma_n^{-1}\right)^{\frac{1}{2}}\sqrt{\beta_n h_n^{-1}}\ln n + \left(\beta_n\gamma_n^{-1}\right)h_n^2 + \Pi_n\right) \quad a.s.
\end{aligned}$$

In view of Assumption $(A2)\,ii)$, (4.33), and applying Lemma 27, we obtain

$$\begin{aligned}
\sup_{x\in I}\left|R_n^{(0)}(x)\right| &= O\left(\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k o\left(\sqrt{\beta_k h_k^{-1}}\ln k\right) + \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k o\left(h_k^2\right)\right. \\
&\quad \left. + o\left(\sqrt{\beta_n h_n^{-1}}\ln n\right) + o\left(h_n^2\right) + O\left(\Pi_n\right)\right) \quad a.s. \\
&= O\left(\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k o\left(\tilde{m}_k\right)\right) + o\left(\tilde{m}_n\right) \quad a.s. \\
&= o\left(\tilde{m}_n\right) \quad a.s.,
\end{aligned}$$

which proves Lemma 26 for $i = 0$.

4.4.4.c Proof of Lemmas 25 and 26 for $i = 2$

We have

$$R_n^{(2)}(x) \;=\; \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left[\mathbb{E}\left(W_k(x)\right) - a(x)\right]\frac{\left[f(x) - f_{k-1}(x)\right]}{f_{k-1}(x)}.$$

The application of first part of Lemma 28 and then of Lemma 27 gives

$$\begin{aligned}
\left|R_n^{(2)}(x)\right| &= O\left(\Pi_n\sum_{k=0}^{n-1}\Pi_k^{-1}\beta_k h_{k+1}^2\left|f(x) - f_k(x)\right|\right) \quad a.s. \\
&= O\left(\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k h_k^2 o(1)\right) \quad a.s. \\
&= O\left(\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k o\left(m_k\right)\right) \quad a.s. \\
&= o\left(m_n\right) \quad a.s.
\end{aligned}$$

which gives Lemma 25 for $i = 2$. Lemma 26 for $i = 2$ is obtained exactly in the same way (except that it is the second part of Lemma 28 which is applied).

4.4.4.d Proof of Lemmas 25 and 26 for $i = 3$

The application of first part of Lemma 28, and then of Lemma 27 ensures that

$$\begin{aligned}
R_n^{(3)}(x) &= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left(f_{k-1}(x) - f(x)\right)^2\right) \quad a.s. \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \left(\frac{\gamma_k \ln(s_k)}{h_k} + h_k^4\right)\right) \quad a.s. \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k o\left(h_k^2\right)\right) \quad a.s. \\
&= O\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k o(m_k)\right) \quad a.s. \\
&= o(m_n) \quad a.s.
\end{aligned}$$

which gives Lemma 25 for $i = 3$. Lemma 26 for $i = 3$ is obtained exactly in the same way (except that it is the second part of Lemma 28 which is applied).

4.4.4.e Proof of Lemma 25 for $i = 1$

We set

$$\begin{aligned}
\varepsilon_k(x) &= W_k(x) - \mathbb{E}(W_k(x)), \\
G_k(x) &= \left[\frac{f(x) - f_k(x)}{f_k(x)}\right], \\
M_n(x) &= \sum_{k=1}^{n} \Pi_k^{-1} \beta_k \varepsilon_k(x) G_{k-1}(x),
\end{aligned}$$

so that $R_n^{(1)}(x) = \Pi_n M_n(x)$. Set $\mathcal{F}_k = \sigma((X_1, Y_1), \ldots, (X_k, Y_k))$; in view of (4.16) and applying the first part of Lemma 28, we note that the increasing process of the martingale $(M_n(x))$ satisfies

$$\begin{aligned}
< M >_n (x) &= \sum_{k=1}^{n} \mathbb{E}\left[\Pi_k^{-2} \beta_k^2 \varepsilon_k^2(x) G_{k-1}^2(x) | \mathcal{F}_{k-1}\right] \\
&= \sum_{k=1}^{n} \Pi_k^{-2} \beta_k^2 G_{k-1}^2(x) Var\left[W_k(x)\right] \\
&= O\left(\sum_{k=1}^{n} \Pi_k^{-2} \frac{\beta_k^2}{h_k} \left(\frac{\gamma_k \ln(s_k)}{h_k} + h_k^4\right)\right) \quad a.s. \qquad (4.35)
\end{aligned}$$

• Let us first consider the case $\lim_{n\to\infty}(n\beta_n) = \infty$. The application of Lemma 27 gives

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

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

$$\begin{aligned}
\ln\left(\Pi_n^{-2}\right) &= \sum_{k=1}^{n} \ln\left(1-\beta_k\right)^{-2} \\
&= \sum_{k=1}^{n} \left(2\beta_k + o\left(\beta_k\right)\right) \\
&= O\left(\sum_{k=1}^{n} \beta_k k^{\epsilon}\right).
\end{aligned}$$

Since $(\beta_n n^{\epsilon}) \in \mathcal{GS}\left(-(\beta-\epsilon)\right)$ with $(\beta-\epsilon)<1$, we have

$$\lim_{n\to\infty} \frac{n\left(\beta_n n^{\epsilon}\right)}{\sum_{k=1}^{n}\beta_k k^{\epsilon}} = 1-(\beta-\epsilon)$$

and thus $\ln\left(\Pi_n^{-2}\right) = O\left(n^{1+\epsilon}\beta_n\right)$. The sequences $\left(\beta_n\gamma_n h_n^{-2}\ln\left(s_n\right)\right)$ and $\left(\beta_n h_n^3\right)$ being in $\mathcal{GS}$, it follows that, for all $\epsilon>0$, we have

$$\ln\left(< M >_n (x)\right) = O\left(n^{1+\epsilon}\beta_n\right) \quad a.s.$$

Theorem 1.3.15 in Duflo (1997) then ensures that, for any $\delta > 0$,

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

Set $\epsilon>0$ and $\delta>0$ such that $\left(\left[\gamma_n h_n^{-1}\right]^{\frac{1}{2}}\left[n^{1+\epsilon}\beta_n\right]^{\frac{1+\delta}{2}}\right) \in \mathcal{GS}\left(u^*\right)$ with $u^*<0$ (the existence of such real numbers ϵ and δ being ensured by the condition $\alpha > a+1-\beta$). We obtain

$$\begin{aligned}
&\Pi_n\left|M_n(x)\right| \\
&= o\left(\left(\frac{\beta_n}{h_n}\right)^{\frac{1}{2}}\left(\frac{\gamma_n}{h_n}\right)^{\frac{1}{2}}\left(n^{1+\epsilon}\beta_n\right)^{\frac{1+\delta}{2}}\left(\ln\left(s_n\right)\right)^{\frac{1}{2}} + h_n^2\left(\frac{\beta_n}{h_n}\right)^{\frac{1}{2}}\left(n^{1+\epsilon}\beta_n\right)^{\frac{1+\delta}{2}}\right) + O\left(\Pi_n\right) \quad a.s. \\
&= o\left(\left(\frac{\beta_n}{h_n}\right)^{\frac{1}{2}} + h_n^2\right) + o\left(m_n\right) \quad a.s. \\
&= o\left(m_n\right) \quad a.s.,
\end{aligned}$$

which gives Lemma 25 for $i=1$ in the case $\lim_{n\to\infty}\left(n\beta_n\right)=\infty$.

• Let us now consider the case the sequence $\left(n\beta_n\right)$ is bounded. In view of (4.35), and applying Lemma 27, we have, for all sequence $\left(\mathcal{L}_n\right)\in\mathcal{GS}(0)$,

$$\begin{aligned}
< M >_n (x) &= O\left(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k\left[\left(\beta_k h_k^{-1}\right)\left(\gamma_k h_k^{-1}\ln\left(s_k\right)\right) + \left(\beta_k h_k^{-1}\right)h_k^4\right]\right) \quad a.s. \\
&= O\left(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k\left[o\left(\beta_k h_k^{-1}\mathcal{L}_k\right) + o\left(h_k^4\mathcal{L}_k\right)\right]\right) \quad a.s. \\
&= o\left(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k\left[o\left(m_k^2\mathcal{L}_k\right)\right]\right) \quad a.s. \\
&= o\left(\Pi_n^{-2}m_n^2\mathcal{L}_n\right) \quad a.s.
\end{aligned}$$

Moreover, since $(n\beta_n)$ is bounded, we have $\Pi_n^{-1} \in \mathcal{GS}\left(\xi^{-1}\right)$, and thus $\ln\left(\Pi_n^{-2} m_n^2 \mathcal{L}_n\right) = O(\ln n)$. Applying Theorem 1.3.15 in Duflo (1997), we deduce that, for all $\delta > 0$,

$$\begin{aligned} |M_n(x)| &= o\left(< M >_n^{\frac{1}{2}}(x)\left(\ln < M >_n (x)\right)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s. \\ &= o\left(\Pi_n^{-1} m_n (\mathcal{L}_n)^{\frac{1}{2}} (\ln n)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s. \end{aligned}$$

Setting for instance $\delta = 1$ and $\mathcal{L}_n = (\ln n)^{-2}$, we get

$$\Pi_n |M_n(x)| = o(m_n) \quad a.s.,$$

which concludes the proof of Lemma 25 for $i = 1$.

4.4.4.f Proof of Lemma 26 for $i = 1$

Set $G_k^*(x) = [f(x) - f_k(x)]$ and $(A_n) \in \mathcal{GS}(a^*)$, where a^* is chosen such that

- if $a \geq \beta/5$, then $0 < a^* < \min\left\{\lim_{n\to\infty}(n\beta_n) - \frac{\beta - a}{2}, \alpha - \frac{\beta}{2} - \frac{3}{2}a, a\right\}$, (4.36)
- if $a < \beta/5$, then $0 < a^* < \min\left\{\lim_{n\to\infty}(n\beta_n) - 2a, \alpha - \frac{\beta}{2} - \frac{3}{2}a, a\right\}$. (4.37)

We write $R_n^{(1)}(x)$ as

$$R_n^{(1)}(x) = \frac{\Pi_n}{f(x)} M_n^{(n)}(x) + \frac{\Pi_n}{f(x)} S_n(x) + \frac{1}{f(x)} U_n^{(1)}(x) + \frac{1}{f(x)} U_n^{(2)}(x)$$

with

$$\begin{aligned} M_k^{(n)}(x) &= \sum_{j=1}^{k} \Pi_j^{-1} \beta_j G_{j-1}^*(x)\left[W_j(x)\mathbb{1}_{|Y_j|\leq A_n} - \mathbb{E}\left(W_j(x)\mathbb{1}_{|Y_j|\leq A_n}\right)\right], \\ S_n(x) &= \sum_{k=1}^{n} \Pi_k^{-1}\beta_k G_{k-1}^*(x)\left[W_k(x)\mathbb{1}_{|Y_k|>A_n} - \mathbb{E}\left(W_k(x)\mathbb{1}_{|Y_k|>A_n}\right)\right], \\ U_n^{(1)}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k\left[W_k(x) - \mathbb{E}(W_k(x))\right]\frac{\left(f(x) - f_{k-1}(x)\right)^2}{f_{k-1}(x)}\mathbb{1}_{\sup_{k\leq n}|Y_k|\leq A_n}, \\ U_n^{(2)}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k\left[W_k(x) - \mathbb{E}(W_k(x))\right]\frac{\left(f(x) - f_{k-1}(x)\right)^2}{f_{k-1}(x)}\mathbb{1}_{\sup_{k\leq n}|Y_k|> A_n}. \end{aligned}$$

We prove a uniform strong upper bound for $U_n^{(1)}$, $U_n^{(2)}$, $\Pi_n S_n$, and $\Pi_n M_n^{(n)}$ successively.

Upper bound of $U_n^{(1)}$
Noting that

$$\left|W_k(x) - \mathbb{E}(W_k(x))\right|\mathbb{1}_{\sup_{k\leq n}|Y_k|\leq A_n} \leq h_k^{-1} A_n \|K\|_\infty,$$

and applying the second part of Lemma 28, we get

$$\sup_{x\in I}\left|U_n^{(1)}(x)\right| = O\left(A_n \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k h_k^{-1}\left(\gamma_k h_k^{-1}(\ln k)^2 + h_k^4\right)\right) \quad a.s.$$

In view of (4.32), of Conditions (4.36) and (4.37), and applying Lemma 27, we obtain

$$\begin{aligned}\sup_{x\in I}\left|U_n^{(1)}(x)\right| &= O\left(A_n\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k o\left(A_k^{-1}\tilde{m}_k\right)\right) \quad a.s.\\ &= o\left(\tilde{m}_n\right) \quad a.s.\end{aligned}$$

Upper bound of $U_n^{(2)}$
In view of $(A4)\,ii)$, we have, for all $c>0$,

$$\begin{aligned}\sum_{n\geq 0}\mathbb{P}\left[\sup_{x\in I}\tilde{m}_n^{-2}\left|U_n^{(2)}(x)\right|\geq c\right] &= O\left(\sum_{n\geq 0}\mathbb{P}\left(\sup_{k\leq n}|Y_k|>A_n\right)\right)\\ &= O\left(\sum_{n\geq 0}n\mathbb{P}\left(|Y|>A_n\right)\right)\\ &= O\left(\sum_{n\geq 0}n\exp\left(-t^*A_n\right)\right)\\ &< \infty.\end{aligned}$$

It follows that

$$\sup_{x\in I}\left|U_n^{(2)}(x)\right| = O\left(\tilde{m}_n^2\right)=o\left(\tilde{m}_n\right) \quad a.s.$$

Upper bound of $\Pi_n S_n$
Let us first note that, in view of (4.4),

$$f_n(x)=Q_n\sum_{k=1}^{n}Q_k^{-1}\gamma_k h_k^{-1}K\left(\frac{x-X_k}{h_k}\right)+Q_n f_0, \tag{4.38}$$

(with $Q_n=\prod_{j=1}^{n}(1-\gamma_j)$). It follows from the application of Lemma 27 that there exists $c^*>0$ such that

$$\begin{aligned}\left|G_n^*(x)\right| &\leq |f_n(x)|+|f(x)|\\ &\leq Q_n\sum_{k=1}^{n}Q_k^{-1}\gamma_k h_k^{-1}\left|K\left(\frac{x-X_k}{h_k}\right)\right|+|f(x)|+Q_n f_0+\|f\|_\infty\\ &\leq c^*h_n^{-1}.\end{aligned} \tag{4.39}$$

Now, let c_i^* denote positive constants. In view of $(A4)\,ii)$, we have

$$\begin{aligned}\mathbb{E}\left[|W_k(x)|\,\mathbb{1}_{|Y_k|>A_n}\right] &\leq \left(\mathbb{E}\left[W_k^2(x)\right]\right)^{\frac{1}{2}}\left[\mathbb{P}\left(|Y_k|>A_n\right)\right]^{\frac{1}{2}}\\ &\leq c_1^*h_k^{-\frac{1}{2}}\left[\mathbb{P}\left(\exp\left(|t^*Y_k|\right)>\exp\left(t^*A_n\right)\right)\right]^{\frac{1}{2}}\\ &\leq c_2^*h_k^{-\frac{1}{2}}\exp\left(-\frac{t^*A_n}{2}\right)\left(\mathbb{E}\left[\exp\left(|t^*Y_k|\right)\right]\right)^{\frac{1}{2}}\\ &\leq c_3^*h_k^{-\frac{1}{2}}\exp\left(-\frac{t^*A_n}{2}\right).\end{aligned}$$

Applying (4.39) and Lemma 27, it follows that

$$\begin{aligned}
&\sup_{x\in I}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left|G_{k-1}^{*}(x)\right|\mathbb{E}\left(\left|W_k(x)\right|\mathbb{1}_{|Y_k|>A_n}\right)\\
&\leq\; c_5^{*}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k h_k^{-\frac{3}{2}}\exp\left(-\frac{t^{*}A_n}{2}\right)\\
&\leq\; c_6^{*}\beta_n^{-\frac{1}{2}}h_n^{-1}\exp\left(-\frac{t^{*}A_n}{2}\right).
\end{aligned}$$

Set $C>0$ and note that, since $(A_n)\in\mathcal{GS}(a^{*})$ with $a^{*}>0$, we have, for n large enough, $c_6^{*}\beta_n^{-\frac{1}{2}}h_n^{-1}\exp\left(-\frac{t^{*}A_n}{2}\right)\leq C/2$. For n large enough, we get

$$\begin{aligned}
&\mathbb{P}\left[\sup_{x\in I}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\left|S_n(x)\right|\geq C\right]\\
&\leq\;\mathbb{P}\left[\sup_{x\in I}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left|G_{k-1}^{*}(x)W_k(x)\right|\mathbb{1}_{|Y_k|>A_n}\right.\\
&\qquad\left.+\sup_{x\in I}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left|G_{k-1}^{*}(x)\right|\mathbb{E}\left(\left|W_k(x)\right|\mathbb{1}_{|Y_k|>A_n}\right)\geq C\right]\\
&\leq\;\mathbb{P}\left[\sup_{x\in I}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left|G_{k-1}^{*}(x)W_k(x)\right|\mathbb{1}_{|Y_k|>A_n}\geq\frac{C}{2}\right]\\
&\leq\;\mathbb{P}\left[\sup_{k\leq n}\left|Y_k\right|\geq A_n\right]\\
&\leq\; n\mathbb{P}\left[|Y|\geq A_n\right]\\
&\leq\; n\exp\left(-t^{*}A_n\right)\mathbb{E}\left(\exp\left(t^{*}|Y|\right)\right).
\end{aligned}$$

Since $(A_n)\in\mathcal{GS}(a^{*})$ with $a^{*}>0$, it follows that

$$\sum_{n\geq 0}\mathbb{P}\left[\sup_{x\in I}\sqrt{\beta_n^{-1}h_n}\,\Pi_n\left|S_n(x)\right|\geq C\right]\;<\;\infty,$$

and thus

$$\sup_{x\in I}\Pi_n\left|S_n(x)\right|\;=\;O\left(\sqrt{\beta_n h_n^{-1}}\right)=o\left(\tilde{m}_n\right)\quad a.s.$$

Upper bound of $\Pi_n M_n^{(n)}$

To establish the strong uniform bound of $M_n^{(n)}$, we shall apply the following result given in Duflo (1997), page 209.

Lemma 30 *Let $\left(M_k^{(n)}\right)_k$ be a martingale such that, for all $k\leq n$, $\left|M_k^{(n)}-M_{k-1}^{(n)}\right|\leq c_n$, and set $\Phi_c(\lambda)=c^{-2}\left(e^{\lambda c}-1-\lambda c\right)$. For all λ_n such that $\lambda_n c_n\leq 1$ and all $\alpha_n>0$, we have*

$$\mathbb{P}\left(\lambda_n\left(M_n^{(n)}-M_0^{(n)}\right)\geq\Phi_{c_n}(\lambda_n)<M^{(n)}>_n+\alpha_n\lambda_n\right)\leq e^{-\alpha_n\lambda_n}.$$

Set ϵ such that

$$\begin{cases} \epsilon \in \left]0, \frac{\beta}{2} - \frac{3}{2}a - a^*\right[& \text{if} \quad a \geq \frac{\beta}{5} \\ \epsilon \in]0, \beta - 4a - a^*[& \text{if} \quad a < \frac{\beta}{5} \end{cases}$$

(The existence of such an ϵ is ensured by the condition $a^* < \alpha - \frac{1}{2}\beta - \frac{3}{2}a$ in (4.36), and in the case (4.37) by the condition $a^* < \alpha - \frac{1}{2}\beta - \frac{3}{2}a < \alpha - 4a$). We first prove that, for all $k \leq n$, we have

$$\left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| \leq c_n$$

with

$$c_n = \begin{cases} C_1 A_n \Pi_n^{-1} \beta_n h_n^{-2} & \text{if} \quad \lim_{n\to\infty}(n\beta_n) > \beta - 2a, \\ C_1 A_n \Pi_n^{-1} \beta_n h_n^{-2} n^\epsilon & \text{if} \quad \lim_{n\to\infty}(n\beta_n) = \beta - 2a, \\ C_1 A_n & \text{if} \quad \lim_{n\to\infty}(n\beta_n) < \beta - 2a \end{cases} \tag{4.40}$$

where $C_1 > 0$ and A_n is defined in (4.36) and (4.37). To this end, we note that, in view of (4.39), there exists $c_1^* > 0$ such that,

$$\begin{aligned} \left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| &\leq \Pi_k^{-1} \beta_k \left| W_k(x) \mathbb{1}_{|Y_k| \leq A_n} - \mathbb{E}\left(W_k(x) \mathbb{1}_{|Y_k| \leq A_n}\right) \right| \left| G_{k-1}^*(x) \right| \\ &\leq \Pi_k^{-1} \beta_k \left(2A_n h_k^{-1} \|K\|_\infty\right) \left(c^* h_{k-1}^{-1}\right) \\ &\leq c_1^* A_n \left(\Pi_k^{-1} \beta_k h_k^{-2}\right). \end{aligned}$$

Now, since $\left(\beta_n h_n^{-2}\right) \in \mathcal{GS}(2a - \beta)$, we have

$$\begin{aligned} \frac{\Pi_{k-1}^{-1} \beta_{k-1} h_{k-1}^{-2}}{\Pi_k^{-1} \beta_k h_k^{-2}} &= (1 - \beta_k) \left(1 - [2a - \beta] \frac{1}{k} + o\left(\frac{1}{k}\right)\right) \\ &= (1 - \beta_k)(1 - [2a - \beta] \xi \beta_k + o(\beta_k)) \\ &= 1 - (1 - [\beta - 2a] \xi) \beta_k + o(\beta_k). \end{aligned}$$

• In the case $\lim_{n\to\infty}(n\beta_n) > \beta - 2a$, it follows that there exists k_0 such that the sequence $\left(\Pi_k^{-1} \beta_k h_k^{-2}\right)_{k \geq k_0}$ is increasing. Consequently, there exists C_1 such that, in this case,

$$\left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| \leq C_1 A_n \left(\Pi_n^{-1} \beta_n h_n^{-2}\right) \quad \forall k \leq n.$$

• In the case $\lim_{n\to\infty}(n\beta_n) = \beta - 2a$, we first note that, for all $k \leq n$,

$$\left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| \leq c_1^* A_n \left(\Pi_n^{-1} \beta_n h_n^{-2} n^\epsilon\right),$$

and prove similarly that there exists C_1 such that

$$\left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| \leq C_1 A_n \left(\Pi_n^{-1} \beta_n h_n^{-2} n^\epsilon\right) \quad \forall k \leq n.$$

• Finally, in the case $\lim_{n\to\infty}(n\beta_n) < \beta - 2a$, we note that the sequence $\left(\Pi_k^{-1} \beta_k h_k^{-2}\right)$ is bounded, so that there exists C_1 such that

$$\left| M_k^{(n)}(x) - M_{k-1}^{(n)}(x) \right| \leq C_1 A_n \quad \forall k \leq n.$$

which concludes the proof of (4.40).
Now, let (w_n) be a positive sequence such that $w_n \leq \Pi_n^{-1} c_n^{-1}$ for all n (where c_n is defined in (4.40)), and set $\lambda_n = \Pi_n w_n$ and q such that

$$\begin{cases} q > 2\beta - 4a & \text{if} \quad \lim_{n\to\infty}(n\beta_n) \geq \beta - 2a, \\ q > 2\xi^{-1} & \text{if} \quad \lim_{n\to\infty}(n\beta_n) < \beta - 2a. \end{cases}$$

Let us at first assume that the following lemma holds.

Lemma 31 *There exist $C_2 > 0$ and $\rho > 0$ such that, if $|x-y| \leq C_2 n^{-\rho}$, then*

$$\begin{aligned} \left| M_n^{(n)}(x) - M_n^{(n)}(y) \right| &\leq \Pi_n^{-1} A_n n^{-q}, \\ \left| < M^{(n)} >_n (x) - < M^{(n)} >_n (y) \right| &\leq \Pi_n^{-2} A_n^2 n^{-q}. \end{aligned}$$

Set $\rho_n = C_2 n^{-\rho}$ and

$$V_n(x) = \lambda_n M_n^{(n)}(x) - \Phi_{c_n}(\lambda_n) < M^{(n)} >_n (x).$$

Let $I_i^{(n)}$ be $N(n)$ intervals of length ρ_n such that $\cup_{i=1}^{N(n)} I_i^{(n)} = I$, and for all $i \in \{1, \ldots, N(n)\}$, set $x_i^{(n)} \in I_i^{(n)}$. We have

$$\mathbb{P}\left(\sup_{x \in I} V_n(x) \geq \alpha_n \lambda_n\right) \leq \sum_{i=1}^{N(n)} \mathbb{P}\left(\sup_{x \in I_n^{(i)}} V_n(x) \geq \alpha_n \lambda_n\right).$$

Applying Lemma 31, and using the facts that $\Phi_c(\lambda) \leq \lambda^2$ as soon as $\lambda c \leq 1$, that $\lambda_n = \Pi_n w_n$, and that $w_n \leq \Pi_n^{-1} c_n^{-1}$ successively, we note that, for all $x \in I_i^{(n)}$,

$$\begin{aligned} V_n(x) &= V_n\left(x_i^{(n)}\right) + \lambda_n \left[M_n^{(n)}(x) - M_n^{(n)}\left(x_i^{(n)}\right)\right] \\ &\quad - \Phi_{c_n}(\lambda_n)\left[< M^{(n)} >_n (x) - < M^{(n)} >_n \left(x_i^{(n)}\right)\right] \\ &\leq V_n\left(x_i^{(n)}\right) + \lambda_n \Pi_n^{-1} n^{-q} A_n + \Phi_{c_n}(\lambda_n) \Pi_n^{-2} n^{-q} A_n^2 \\ &\leq V_n\left(x_i^{(n)}\right) + w_n n^{-q} A_n + w_n^2 n^{-q} A_n^2 \\ &\leq V_n\left(x_i^{(n)}\right) + \Pi_n^{-1} c_n^{-1} n^{-q} A_n + \Pi_n^{-2} c_n^{-2} n^{-q} A_n^2. \end{aligned}$$

In the case $\lim_{n\to\infty}(n\beta_n) \geq \beta - 2a$, in view of (4.40) and since $q > 2\beta - 4a$, we get, for all $x \in I_i^{(n)}$,

$$\begin{aligned} V_n(x) &\leq V_n\left(x_i^{(n)}\right) + C_1^{-1}\beta_n^{-1} h_n^2 n^{-q} + C_1^{-2} \beta_n^{-2} h_n^4 n^{-q} \\ &\leq V_n\left(x_i^{(n)}\right) + 1 \quad \text{for } n \text{ large enough.} \end{aligned}$$

In the case $\lim_{n\to\infty}(n\beta_n) < \beta - 2a$, we have $\left(\Pi_n^{-1}\right) \in \mathcal{GS}\left(\xi^{-1}\right)$; in view of (4.40) and since $q > 2\xi^{-1}$, we obtain, for all $x \in I_i^{(n)}$,

$$\begin{aligned} V_n(x) &\leq V_n\left(x_i^{(n)}\right) + \Pi_n^{-1} n^{-q} + \Pi_n^{-2} n^{-q} \\ &\leq V_n\left(x_i^{(n)}\right) + 1 \quad \text{for } n \text{ large enough.} \end{aligned}$$

Set α_n such that $\alpha_n \lambda_n = 2(\rho+2)\ln n$. Applying Lemma 30, we deduce that, for n large enough,

$$\begin{aligned} \mathbb{P}\left[\sup_{x\in I} V_n(x) \geq \frac{\alpha_n \lambda_n}{2}\right] &\leq \sum_{i=1}^{N(n)} \mathbb{P}\left[V_n\left(x_i^{(n)}\right) \geq \frac{\alpha_n\lambda_n}{2}\right] \\ &\leq N(n) \sup_{x \in I} \mathbb{P}\left[V_n(x) \geq \frac{\alpha_n \lambda_n}{2}\right] \\ &= O\left(n^{\rho} \exp\left(-\frac{\alpha_n\lambda_n}{2}\right)\right) \\ &= O\left(n^{-2}\right). \end{aligned}$$

It follows that

$$\sum_{n\geq 0}\mathbb{P}\left[\sup_{x\in I}V_n(x)\geq 2(\rho+2)\ln n\right]<\infty,$$

and Borel-Cantelli Lemma then ensures that

$$\sup_{x\in I}\lambda_n M_n^{(n)}(x) \quad \leq \quad \lambda_n^2\sup_{x\in I}<M^{(n)}>_n(x)+2(\rho+2)\ln n \quad a.s.$$

Establishing the same upper bound for the martingale $\left(-M_k^{(n)}(x)\right)$, we deduce that, since $\lambda_n = \Pi_n w_n$,

$$\sup_{x\in I}\Pi_n\left|M_n^{(n)}(x)\right| \quad \leq \quad w_n\sup_{x\in I}\Pi_n^2<M^{(n)}>_n(x)+2\frac{(\rho+2)\ln n}{w_n} \quad a.s.$$

Now, applying the second part of Lemm 28, and then Lemma 27, for any sequence $(\mathcal{L}_n)\in\mathcal{GS}(0)$, we have

$$\begin{aligned}
\sup_{x\in I}<M^{(n)}>_n(x) &= O\left(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k^2\sup_{x\in I}\left|G_{k-1}^{*2}(x)\right|\sup_{x\in I}\left|Var\left[W_k(x)\mathbb{1}_{|Y_k|\leq A_n}\right]\right|\right)\\
&= O\left(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k^2\left(\frac{\gamma_k(\ln k)^2}{h_k^2}+h_k^3\right)\right) \quad a.s.\\
&= O\left(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k m_k^2\mathcal{L}_k\right) \quad a.s.\\
&= O\left(\Pi_n^{-2}m_n^2\mathcal{L}_n\right) \quad a.s.
\end{aligned}$$

(where (m_n) is defined in (4.31)).
We have thus proved that, for all positive sequence (w_n) such that $w_n\leq\Pi_n^{-1}c_n^{-1}$ (with c_n defined in (4.40)), and for all sequence $(\mathcal{L}_n)\in\mathcal{GS}(0)$,

$$\sup_{x\in I}\Pi_n\left|M_n^{(n)}(x)\right| = O\left(w_n m_n^2\mathcal{L}_n+\frac{\ln n}{w_n}\right) \quad a.s. \tag{4.41}$$

• Let us first consider the case $a\geq\beta/5$. If $\lim_{n\to\infty}(n\beta_n)\geq\beta-2a$, then, since $\epsilon<\beta/2-3a/2-a^*$, we have, in view of (4.40),

$$\begin{aligned}
\Pi_n c_n\sqrt{\beta_n^{-1}h_n}\ln n &\leq C_1A_n\beta_n^{\frac{1}{2}}h_n^{-\frac{3}{2}}n^{\epsilon}\ln n\in\mathcal{GS}\left(a^*-\frac{\beta}{2}+\frac{3}{2}a+\epsilon\right)\\
&\to 0 \quad \text{as} \quad n\to\infty,
\end{aligned}$$

and, if $\lim_{n\to\infty}(n\beta_n)<\beta-2a$, then, $\Pi_n^{-1}\in\mathcal{GS}\left(\xi^{-1}\right)$, and, since $a^*<\xi^{-1}-\beta/2+a/2$ (see (4.36)), we get

$$\begin{aligned}
\Pi_n c_n\sqrt{\beta_n^{-1}h_n}\ln n &\leq C_1\Pi_nA_n\sqrt{\beta_n^{-1}h_n}\ln n\in\mathcal{GS}\left(-\xi^{-1}+a^*+\frac{\beta}{2}-\frac{a}{2}\right)\\
&\to 0 \quad \text{as} \quad n\to\infty.
\end{aligned}$$

It follows that there exists $w_0>0$ such that, for all n, $w_0\sqrt{\beta_n^{-1}h_n}\ln n\leq\Pi_n^{-1}c_n^{-1}$. Setting $(w_n) = \left(w_0\sqrt{\beta_n^{-1}h_n}\ln n\right)$ and $(\mathcal{L}_n)=(\ln n)^{-1}$ in (4.41), we obtain

$$\begin{aligned}
\sup_{x\in I}\Pi_n\left|M_n^{(n)}(x)\right| &= O\left(\sqrt{\beta_n h_n^{-1}}\right) \quad a.s,\\
&= o(\tilde{m}_n) \quad a.s.
\end{aligned}$$

• Let us now consider the case $a < \beta/5$. If $\lim_{n\to\infty}(n\beta_n) \geq \beta - 2a$, then, since $\epsilon < \beta - 4a - a^*$, we have, in view of (4.40),

$$\begin{aligned}\Pi_n c_n h_n^{-2}(\ln n)^2 &\leq C_1 A_n \beta_n h_n^{-4} n^{\epsilon}(\ln n)^2 \in \mathcal{GS}(a^* - \beta + 4a + \epsilon)\\ &\to 0 \quad \text{as} \quad n\to\infty,\end{aligned}$$

and, if $\lim_{n\to\infty}(n\beta_n) < \beta - 2a$, then, since $a^* < \xi^{-1} - 2a$ (see (4.37)), we get

$$\begin{aligned}\Pi_n c_n h_n^{-2}(\ln n)^2 &\leq C_1 \Pi_n A_n h_n^{-2}(\ln n)^2 \in \mathcal{GS}\left(-\xi^{-1} + a^* + 2a\right)\\ &\to 0 \quad \text{as} \quad n\to\infty.\end{aligned}$$

Consequently, there exists $w_0 > 0$ such that, for all n, $w_0 h_n^{-2}(\ln n)^2 \leq \Pi_n^{-1} c_n^{-1}$. Applying (4.41) with $(w_n) = \left(w_0 h_n^{-2}(\ln n)^2\right)$ and $(\mathcal{L}_n) = (\ln n)^{-3}$, we obtain

$$\begin{aligned}\sup_{x\in I}\Pi_n\left|M_n^{(n)}(x)\right| &= O\left(h_n^2(\ln n)^{-1}\right) \quad a.s,\\ &= o(\tilde{m}_n) \quad a.s.\end{aligned}$$

which concludes the proof of Lemma 26 for $i = 1$.

Proof of Lemma 31

Set $x, y \in I$ such that $|x - y| < \rho_n$, and let c_i^* denote positive constants. In view of (4.38), we have

$$\begin{aligned}\left|G_{k-1}^*(x) - G_{k-1}^*(y)\right| &\leq |f_{k-1}(x) - f_{k-1}(y)| + |f(x) - f(y)|\\ &\leq Q_{k-1}\sum_{j=1}^{k-1} Q_{j-1}^{-1}\gamma_j h_j^{-1}\left|K\left(\frac{x - X_j}{h_j}\right) - K\left(\frac{y - X_j}{h_j}\right)\right| + |f(x) - f(y)|\\ &\leq c_1^*\rho_n Q_{k-1}\sum_{j=1}^{k-1} Q_{j-1}^{-1}\gamma_j h_j^{-2} + c_2^*\rho_n\\ &\leq c_3^*\rho_n h_k^{-2}.\end{aligned}$$

Moreover, we get

$$\begin{aligned}|W_k(x) - W_k(y)|\,\mathbb{1}_{|Y_k|\leq A_n} &\leq A_n h_k^{-1}\left|K\left(\frac{x - X_k}{h_k}\right) - K\left(\frac{y - X_k}{h_k}\right)\right|\\ &\leq c_4^* A_n\rho_n h_k^{-2},\end{aligned}$$

and

$$|W_k(x)|\,\mathbb{1}_{|Y_k|\leq A_n} \leq A_n h_k^{-1}\|K\|_\infty.$$

Using (4.39), it follows that

$$\begin{aligned}&\left|M_n^{(n)}(x) - M_n^{(n)}(y)\right|\\ &\leq \sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left|G_{k-1}^*(x) - G_{k-1}^*(y)\right|\left(|W_k(x)|\,\mathbb{1}_{|Y_k|\leq A_n} + \mathbb{E}\left[|W_k(x)|\,\mathbb{1}_{|Y_k|\leq A_n}\right]\right)\\ &\quad + \sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left|G_{k-1}^*(y)\right|\left(|W_k(x) - W_k(y)|\,\mathbb{1}_{|Y_k|\leq A_n} + \mathbb{E}\left[|W_k(x) - W_k(y)|\,\mathbb{1}_{|Y_k|\leq A_n}\right]\right)\\ &\leq c_5^*\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\left(\rho_n h_k^{-2}\right)\left(2A_n h_k^{-1}\|K\|_\infty\right) + c_6^* A_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k h_k^{-3}\rho_n\\ &\leq c_7^* A_n\rho_n\Pi_n^{-1}h_n^{-3}. \qquad (4.42)\end{aligned}$$

Similary, we get

$$\begin{aligned}
&\left|< M^{(n)} >_n (x) - < M^{(n)} >_n (y)\right| \\
\leq\ & \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 \left|G_{k-1}^{*2}(x) - G_{k-1}^{*2}(y)\right| Var\left[W_k(x)\mathbb{1}_{|Y_k|\leq A_n}\right] \\
&+ \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 \left|G_{k-1}^{*2}(y)\right| \left|Var\left[W_k(x)\mathbb{1}_{|Y_k|\leq A_n}\right] - Var\left[W_k(y)\mathbb{1}_{|Y_k|\leq A_n}\right]\right| \\
\leq\ & \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 \left|G_{k-1}^{*}(x) - G_{k-1}^{*}(y)\right| \left|G_{k-1}^{*}(x) + G_{k-1}^{*}(y)\right| Var\left[W_k(x)\mathbb{1}_{|Y_k|\leq A_n}\right] \\
&+ \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 \left|G_{k-1}^{*2}(y)\right| \mathbb{E}\left(\mathbb{1}_{|Y_k|\leq A_n} |W_k(x) - W_k(y)| \, |W_k(x) + W_k(y)|\right) \\
&+ \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 \left|G_{k-1}^{*2}(y)\right| \mathbb{E}\left[|W_k(x) - W_k(y)| \mathbb{1}_{|Y_k|\leq A_n}\right] \mathbb{E}\left[|W_k(x) + W_k(y)| \mathbb{1}_{|Y_k|\leq A_n}\right] \\
\leq\ & c_8^*\rho_n \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 h_k^{-4} + c_9^* A_n^2 \rho_n \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 h_k^{-5} \\
\leq\ & c_{10}^* A_n^2 \rho_n \Pi_n^{-2}\beta_n h_n^{-5}. \qquad (4.43)
\end{aligned}$$

For all $q > 0$, there exists $\rho > 0$ such that $\lim_{n\to\infty} h_n^{-3} n^{-\rho} = 0$ and $\lim_{n\to\infty} \beta_n h_n^{-5} n^{-\rho} = 0$. Lemma 31 Thus follows from (4.42) and (4.43).

Bibliographie

[1] Baras, J.S. and Borkar, V.S. (2000), A learning algorithm for Markov decision processes with adaptive state aggregation. In *Proc. 39th IEEE Conference on Decision and Control.* IEEE, New York

[2] Bhatnagar, S., Fu, M.C., Marcus S.I. and Bathnagar, S. (2001), Two timescale algorithms for simulation optimization of hidden Markov models,*IEEE Transactions* **3**, 245-258

[3] Bhatnagar, S., Fu, M.C., Marcus S.I. and Fard, P.J. (2001), Optimal structured feedback policies for ABR flow control using two timescale SPSA, *IEEE/ACM Transactions on Networkings* **9**, 479-491

[4] Blum, J.R. (1954), Multidimensional stochastic approximation methods, *Ann. Math. Statist.,* **25**, 737-744

[5] Bojanic, R. and Seneta, E. (1973), A unified theory of regularly varying sequences, *Math.* Z., **134**, 91-106.

[6] Borkar, V.S. (1997), Stochastic approximation with two time scales, *Systems Control Lett.* **29**, 291-294

[7] Chen, H. (1988), Lower rate of convergence for locating a maximum of a function, *Ann. Statist.,* **16**, 1330-1334

[8] Chen, H.F., Duncan, T.E., and Pasik-Duncan, B. (1999), A Kiefer-Wolfowitz algorithm with randomized differences, *IEEE Trans. Automat. Control,* **44**, 442-453

[9] Dippon, J. and Renz J. (1997), Weighted means in stochastic approximation of minima, *SIAM J. Control Optim.,* **35**, 1811-1827

[10] Dippon, J. (2003), Accelerated randomized stochastic optimization, *Ann. Statist.,* **31**, 1260-1281

[11] Duflo, M. (1997), *Random Iterative Models*, Collection Applications of mathematics, Springer.

[12] Fabian, V. (1967), Stochastic approximation of minima with improved asymptotic speed, *Ann. Math. Statist.,* **38**, 191-200

[13] Galambos, J. and Seneta, E. (1973), Regularly varying sequences, *Proc. Amer. Math. Soc.*, **41**, 110-116.

[14] Hall, P. and Heyde, C.C. (1980), *Martingale limit theory and its application,* Academic Press, Inc., New York-London

[15] Hall, P. (1992), Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Ann. Statist.* **20**, 675-694

[16] Kiefer, J. and Wolfowitz, J. (1952), Stochastic approximation of the maximum of a regression functions, *Ann. Math. Statist.,* **23**, 462-466

[17] Konda, V.R. and Borkar, V.S. (1999), Actor-critic like learning algorithms for Markov decision processes, *SIAM J. Control Optim.* **38**, 94-123

[18] Konda, V.R. and Tsitsiklis, J.N. (2003), On actor-critic algorithms, *SIAM J. Control Optim.* **42**, 1143-1166

[19] Konda, V.R. and Tsitsiklis, J.N. (2004), Convergence rate of linear two-time-scale stochastic approximation, *Ann. of Appl. Probab.* **14**, 796-819

[20] Kushner H.J. and Clark, D.S. (1978) *Stochastic approximation methods for constrained and unconstrained systems,* Springer, New York

[21] Mokkadem, A. and Pelletier, M. (2004), A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *submitted*

[22] Mokkadem, A. and Pelletier, M. (2006a), Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms, *Ann. of Appl. Probab.* **16**, 1671-1702

[23] Mokkadem, A. and Pelletier, M. (2006b), Compact law of the iterated logarithm for matrix-normalized sums of random vectors, *to appear in Theory Probab. Appl.*

[24] Polyak, B.T. and Tsybakov, A.B. (1990), Optimal orders of accuracy for search algorithms of stochastic optimization, *Problems Inform. Transmission,* **26**, 126-133.

[25] Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.*, **8**, 391-398.

[26] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119-126.

[27] Ruppert, D. (1982), Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise, *Ann. of Probab.,* **10**, 178-187.

[28] Spall, J.C. (1988), A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting. In *Proc. Conference on Decision and Control*, 1544-1548. IEEE, New York.

[29] Spall, J.C. (1997), A one-measurement form of simultaneous perturbation stochastic approximation, *Automatica J. IFAC*, **33**, 109-112.

NOM : Yousri Slaoui

Titre : Application des méthodes d'approximation stochastique à l'estimation de la densité et de la régression

RÉSUMÉ L'objectif de cette thèse est d'appliquer les méthodes d'approximation stochastique à l'estimation de la densité et de la régression. Dans le premier chapitre, nous construisons un algorithme stochastique à pas simple qui définit toute une classe d'estimateurs récursifs à noyau d'une densité de probabilité. Nous étudions les différentes propriétés de cet algorithme. En particulier, nous identifions deux classes d'estimateurs ; la première correspond à un choix de pas qui permet d'obtenir un risque minimal, la seconde une variance minimale. Dans le deuxième chapitre, nous nous intéressons à l'estimateur proposé par Révész (1973, 1977) pour estimer une fonction de régression $r : x \mapsto \mathbb{E}[Y|X = x]$. Son estimateur r_n, construit à l'aide d'un algorithme stochastique à pas simple, a un gros inconvénient : les hypothèses sur la densité marginale de X nécessaires pour établir la vitesse de convergence de r_n sont beaucoup plus fortes que celles habituellement requises pour étudier le comportement asymptotique d'un estimateur d'une fonction de régression. Nous montrons comment l'application du principe de moyennisation des algorithmes stochastiques permet, tout d'abord en généralisant la définition de l'estimateur de Révész, puis en moyennisant cet estimateur généralisé, de construire un estimateur récursif $\bar{r}_n$ qui possède de bonnes propriétés asymptotiques. Dans le troisième chapitre, nous appliquons à nouveau les méthodes d'approximation stochastique à l'estimation d'une fonction de régression. Mais cette fois, plutôt que d'utiliser des algorithmes stochastiques à pas simple, nous montrons comment les algorithmes stochastiques à pas doubles permettent de construire toute une classe d'estimateurs récursifs d'une fonction de régression, et nous étudions les propriétés asymptotiques de ces estimateurs. Cette approche est beaucoup plus simple que celle du deuxième chapitre : les estimateurs construits à l'aide des algorithmes à pas doubles n'ont pas besoin d'être moyennisés pour avoir les bonnes propriétés asymptotiques.

TITLE : Application of stochastic approximation methods to estimate a density and a regression function

ABSTRACT The objective of this thesis is to apply the stochastic approximation methods to the estimation of a density and of a regression function. In the first chapter, we build up a stochastic algorithm with single stepsize, which defines a whole class of recursive kernel estimators of a probability density. We study the properties of this algorithm. In particular, we identify two classes of estimators ; the first one corresponds to a choice of stepsize which allows to get a minimum mean squared error, the second one a minimum variance. In the second chapter, we consider the estimator proposed by Révész (1973, 1977) to estimate a regression function $r : x \mapsto \mathbb{E}[Y|X = x]$. His estimator r_n, built up by using a single-time-scale stochastic algorithm, has a big disadvantage : the assumptions on the marginal density of X necessary to establish the convergence rate of r_n are much stronger than those usually required to study the asymptotic behavior of an estimator of a regression function. We show how the application of the averaging principle of stochastic algorithms allows, by first generalizing the definition of the estimator of Révész and then by averaging this generalized estimator, to build up a recursive estimator $\bar{r}_n$ which has good asymptotic properties. In the third chapter, we still apply stochastic approximation methods to estimate a regression function. But this time, rather than to use single-time-scale stochastic algorithm, we show how the two-time-scale stochastic algorithms allow to build up a whole class of recursive estimators of a regression function, and we study the asymptotic properties of these estimators. This approach is much easier than the one of the second chapter : the estimators built up using the two-time-scale algorithms do not need to be averaged to have good asymptotic properties.

DISCIPLINEE : Mathématiques appliquées, orientation Probabilités-Statistiques

MOTS CLÉS : Algorithme d'approximation stochastique ; Estimation ; Estimation de densité ; La régression nonparamétrique.

CLASSIFICATION AMS : 62L20, 62G05, 62G07, 62G08.