On the asymptotic behavior of the Nadaraya-Watson estimator associated with the recursive SIR method

Bernard Bercu, Thi Mong Ngoc Nguyen, Jérôme Saracco

ON THE ASYMPTOTIC BEHAVIOR OF THE NADARAYA-WATSON ESTIMATOR ASSOCIATED WITH THE RECURSIVE SIR METHOD

BERNARD BERCU, THI MONG NGOC NGUYEN, AND JEROME SARACCO

Abstract. We investigate the asymptotic behavior of the Nadaraya-Watson estimator for the estimation of the regression function in a semiparametric regression model. On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown parameter of the model. On the other hand, we implement a recursive Nadaraya-Watson procedure for the estimation of the regression function which takes into account the previous estimation of the parameter of the semiparametric regression model. We establish the almost sure convergence as well as the asymptotic normality for our Nadaraya-Watson estimator. We also illustrate our semiparametric estimation procedure on simulated data.

1. INTRODUCTION

The goal of this paper is to investigate the asymptotic behavior of the Nadaraya-Watson estimator of the regression function f in the semiparametric regression model given, for all $n \geq 1$, by

$$Y_n = f(\theta' X_n) + \varepsilon_n \tag{1.1}$$

where (X_n) is a sequence of independent and identically distributed random vectors of $\mathbb{R}^p$ and the driven noise (ε_n) is a real martingale difference sequence independent of (X_n). We assume in all the sequel that the unknown p-dimensional parameter $\theta \neq 0$. On the one hand, we make use of the recursive version of the sliced inverse regression (SIR) method, originally proposed by Li [11] and Duan and Li [7], in order to estimate θ. On the other hand, we estimate the unknown regression function f via a recursive Nadaraya-Watson estimator which takes into account the previous estimation of the parameter θ. Our purpose is precisely to investigate the asymptotic behavior of the recursive Nadaraya-Watson estimator of f.

One can find a wide range of literature on nonparametric estimation of a regression function. We refer the reader to [6], [13], [19], [21] for some excellent books on density and regression function estimation. In the classical situation without any parameter θ, the almost sure convergence of the Nadaraya-Watson estimator [12], [22] was proved by Noda [15] and its asymptotic normality was established by Schuster [18]. Moreover, Choi, Hall and Rousson [4] propose three data-sharpening versions of the

2000 *Mathematics Subject Classification.* Primary: 62H12 Secondary: 62G05, 60F05, 62L12.

Key words and phrases. Semi-parametric regression, recursive estimation, Nadaraya-Watson estimator, Sliced inversion regression.

Nadaraya-Watson estimator in order to reduce the asymptotic variance in the central limit theorem. In our situation, we propose to make use of a recursive Nadaraya-Watson estimator [8] of f which takes into account the previous estimation of the parameter θ. It is given, for all $x \in \mathbb{R}^p$, by

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^{n} W_k(x) Y_k}{\sum_{k=1}^{n} W_k(x)} \tag{1.2}$$

with

$$W_n(x) = \frac{1}{h_n} K\Big(\frac{x - \widehat{\theta}'_{n-1} X_n}{h_n}\Big)$$

where the kernel K is a chosen probability density function and the bandwidth (h_n) is a sequence of positive real numbers decreasing to zero, such that nh_n tends to infinity. For the sake of simplicity, we propose to make use of $h_n = 1/n^{\alpha}$ with $\alpha \in]0,1[$. The main difficulty arising here is that we have to deal with the recursive SIR estimator $\widehat{\theta}_n$ of θ inside the kernel K.

The paper is organized as follows. Section 2 is devoted to the recursive SIR estimator $\widehat{\theta}_n$. Our main results on the asymptotic behavior of $\widehat{f}_n$ are given in Section 3. Under standard regularity assumptions on the kernel K, we establish the almost sure pointwise convergence of $\widehat{f}_n$ together with its asymptotic normality. Section 4 contains some numerical experiments on simulated data, illustrating the good performances of our semiparametric estimation procedure. All the technical proofs are postponed in Appendices A and B.

2. ON THE RECURSIVE SIR METHOD

From the seminal work of Li [11] and Duan and Li [7] devoted to the SIR theory, we know that the eigenvector associated with the maximum eigenvalue of the matrix $\Sigma^{-1}\Gamma$ is collinear with θ where $\Sigma = \mathbb{V}(X_n)$ is positive definite, $\Gamma = \mathbb{V}(\mathbb{E}(X_n|T(Y_n))$ and T is a slicing of the range of Y_n into H non overlapping slices $s_1, \cdots, s_H$. One can observe that since the link function f is unknown in the semiparametric regression model (1.1), the parameter θ is not entirely identifiable. Only its direction can be identified without assuming additional constraints. Li [11] called effective dimension reduction (EDR), any direction collinear with θ. Moreover, the SIR theory mainly relies on the so-called linear condition (LC) which imposes that for all $b \in \mathbb{R}^p$, $\mathbb{E}[b'X_n|\theta'X_n]$ is linear in $\theta'X_n$. It means that one can find $\alpha, \beta \in \mathbb{R}$ such that

$$\mathbb{E}[b'X_n|\theta'X_n] = \alpha + \beta\theta'X_n. \tag{LC}$$

This condition is required to only hold for the true parameter θ. Since θ is unknown, it is not possible in practice to verify it a priori. Hence, we can assume that (LC) holds for all possible values of θ, which is equivalent to elliptical symmetry of the distribution of the identically distributed sequence (X_n). Finally, Hall and Li [10] mentioned that (LC) is not a severe restriction because (LC) holds to a good approximation in many problems as the dimension p of the regression vector X_n increases. Chen and Li [3] or Cook and Ni [5] also provide interesting discussions on the linear condition.

In order to obtain a recursive version of an EDR direction estimated with SIR approach, we need an analytic expression of the maximum eigenvector of $\Sigma^{-1}\Gamma$. It is easily tractable when the range of Y_n is divided into two non overlapping slices s_1 and s_2. Hereafter we shall assume that $H = 2$. In this special case, it is not hard to see that $\Gamma = p_1 z_1 + p_2 z_2$ where $p_h = P(Y_n \in s_h)$ and $z_h = \mathbb{E}[X_n | Y_n \in s_h] - \mathbb{E}[X_n]$ with $p_h \neq 0$ for $h = 1, 2$. Moreover, it is straightforward to show that the eigenvector associated to the maximum eigenvalue of $\Sigma^{-1}\Gamma$ can be written as

$$\widetilde{\theta} = \Sigma^{-1}(z_1 - z_2).$$

This vector $\widetilde{\theta}$ is therefore an EDR direction. For the sake of simplicity, we identify in all the sequel the EDR direction $\widetilde{\theta}$ with θ. Our purpose is now to propose an estimator of the EDR direction θ. First of all, let us recall the non recursive SIR estimator $\widetilde{\theta}_n$ of θ given by Nguyen and Saracco [14]. The estimator $\widetilde{\theta}_n$ can be easily obtained from the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ by substituting the theoritical moments by their sample couterparts. More precisely, $\widetilde{\theta}_n$ is given by

$$\widetilde{\theta}_n = \Sigma_n^{-1}(z_{1,n} - z_{2,n}) \tag{2.1}$$

where

$$\Sigma_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)(X_k - \overline{X}_n)', \qquad \overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{2.2}$$

and, for $h = 1, 2$, $z_{h,n} = m_{h,n} - \overline{X}_n$ where

$$m_{h,n} = \frac{1}{n_{h,n}}\sum_{k=1}^{n} X_k \mathrm{I}_{\{Y_k \in s_h\}}, \qquad n_{h,n} = \sum_{k=1}^{n} \mathrm{I}_{\{Y_k \in s_h\}}. \tag{2.3}$$

Next, we focus our attention on the recursive SIR estimator $\widehat{\theta}_n$ of θ proposed by Bercu, Nguyen and Saracco [1], [14]. We split the sample into two parts: the subsample of the first $(n-1)$ observations $(X_1, Y_1), \ldots, (X_{n-1}, Y_{n-1})$, and the new observation (X_n, Y_n). On the one hand, the inverse of the matrix Σ_n given by (2.4) may be recursively calculated via the Riccati equation [8],

$$\Sigma_n^{-1} = \frac{n}{n-1}\Sigma_{n-1}^{-1} - \frac{n}{(n-1)(n+\rho_n)}\Sigma_{n-1}^{-1}\Phi_n\Phi_n'\Sigma_{n-1}^{-1} \tag{2.4}$$

where $\rho_n = \Phi_n'\Sigma_{n-1}^{-1}\Phi_n$ and $\Phi_n = X_n - \overline{X}_{n-1}$. On the other hand, we can also obtain the recursive form of $z_{h,n}$. As a matter of fact, we have for $h = 1, 2$,

$$z_{h,n} = \begin{cases} z_{h^*,n-1} - \dfrac{1}{n}\Phi_n + \dfrac{1}{n_{h^*,n-1}+1}\Phi_{h^*,n} & \text{if } h = h^*, \\[2ex] z_{h,n-1} - \dfrac{1}{n}\Phi_n & \text{otherwise,} \end{cases} \tag{2.5}$$

where h^* denotes the slice containing the observation Y_n and $\Phi_{h^*,n} = X_n - m_{h^*,n-1}$. We deduce from (2.4) and (2.5) that the recursive SIR estimator $\widehat{\theta}_n$ is given by

$$\begin{aligned}\widehat{\theta}_n &= \left(\frac{n}{n-1}\right)\widehat{\theta}_{n-1} - \frac{n}{(n-1)(n+\rho_n)}\Sigma_{n-1}^{-1}\Phi_n\Phi_n'\widehat{\theta}_{n-1} \\ &- \frac{(-1)^{h^*}\,n}{(n_{h^*,n-1}+1)(n-1)}\left(\Sigma_{n-1}^{-1} - \frac{1}{n+\rho_n}\Sigma_{n-1}^{-1}\Phi_n\Phi_n'\Sigma_{n-1}^{-1}\right)\Phi_{h^*,n}.\end{aligned} \tag{2.6}$$

The SIR estimators $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ share the same asymptotic properties, previously established in [14], under the following classical hypothesis.

- (H_1) The random vectors (X_n) are square integrable, independent and identically distributed and $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independently drawn from (1.1).

Lemma 2.1. *Assume that (LC) and (H_1) hold. Then, $\widehat{\theta}_n$ converges a.s. to θ,*

$$||\widehat{\theta}_n - \theta||^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \qquad a.s. \tag{2.7}$$

In addition, we also have the asymptotic normality

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Delta) \tag{2.8}$$

where the limiting covariance matrix Δ may be explicitely calculated.

3. MAIN RESULTS

Our purpose is to investigate the asymptotic properties of the recursive Nadaraya-Watson estimator $\widehat{f}_n$ of the link function f given by (1.2). First of all, we assume that the kernel K is a positive symmetric function, bounded with compact support, twice differentiable with bounded derivatives, satisfying

$$\int_{\mathbb{R}} K(x)\,dx = 1 \qquad \text{and} \qquad \int_{\mathbb{R}} K^2(x)\,dx = \nu^2.$$

Moreover, it is necessary to add the following standard hypothesis.

- (H_2) The probability density function g associated with (X_n) is continuous, positive on all $\mathbb{R}^p$, twice differentiable with bounded derivatives.
- (H_3) The link function f is Lipschitz.

Our first result deals with the almost sure convergence of the estimator $\widehat{f}_n$.

Theorem 3.1. *Assume that (LC) and (H_1) to (H_3) hold. In addition, suppose that the sequence (X_n) has a finite moment of order $a > 2$. Then, for any $x \in \mathbb{R}$, we have*

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \qquad a.s. \tag{3.1}$$

More precisely, if the bandwidth (h_n) is given by $h_n = 1/n^\alpha$ with $0 < \alpha < 1/3$,

$$\widehat{f}_n(x) - f(x) = \mathcal{O}\left(n^{-\alpha}\right) + \mathcal{O}\left(n^{1/a}\sqrt{\frac{\log(\log n)}{n}}\right) \qquad a.s. \tag{3.2}$$

while, if $1/3 \leq \alpha < 1$,

$$\widehat{f}_n(x) - f(x) = \mathcal{O}\left(\sqrt{n^{\alpha-1}}\log n\right) + \mathcal{O}\Big(n^{1/a}\sqrt{\frac{\log(\log n)}{n}}\Big) \qquad a.s. \tag{3.3}$$

Proof. The proof is given Appendix A. ☐

Remark 3.1. *In the particular case where* (X_n) *is a sequence of independent random vectors of* $\mathbb{R}^p$ *sharing the same* $\mathcal{N}(m,\Sigma)$ *distribution where the covariance matrix* Σ *is positive definite, we can replace* $n^{1/a}$ *by* $\log n$ *into* (3.2) *and* (3.3)*. Consequently, for any* $x \in \mathbb{R}$*, we obtain that if* $0 < \alpha < 1/3$,

$$\widehat{f}_n(x) - f(x) = \mathcal{O}\left(n^{-\alpha}\right) \qquad a.s.$$

while, if $1/3 \leq \alpha < 1$,

$$\widehat{f}_n(x) - f(x) = \mathcal{O}\left(\sqrt{n^{\alpha-1}}\log n\right) \qquad a.s.$$

The asymptotic normality of the estimator $\widehat{f}_n$ is as follows.

Theorem 3.2. *Assume that* (LC) *and* (H_1) *to* (H_3) *hold. In addition, suppose that the sequence* (X_n) *has a finite moment of order* $a = 6$ *and that the sequence* (ε_n) *has a finite conditional moment of order* $b > 2$*. Then, as soon as the bandwidth* (h_n) *satisfies* $h_n = 1/n^\alpha$ *with* $1/3 < \alpha < 1$*, we have for any* $x \in \mathbb{R}$*, the pointwise asymptotic normality*

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma^2\nu^2}{(1+\alpha)h(\theta,x)}\Big) \tag{3.4}$$

where $h(\theta,x)$ *stands for the probability density function associated with* $(\theta' X_n)$*.*

Proof. The proof is given Appendix B. ☐

4. NUMERICAL SIMULATIONS

The goal of this Section is to illustrate via some numerical experiments the theoretical results of Section 3. We will provide the numerical behavior of our recursive estimators combining the recursive Nadaraya-Watson estimator of the link function f together with the recursive SIR estimator of the parameter θ. First of all, we describe in Section 4.1 the simulated model used in the numerical study and we present the estimation procedure, in particular the choice of the bandwidth parameter α by a cross-validation criterion. Then, we illustrate in Sections 4.2 and 4.3 the almost sure convergence and the asymptotic normality of our recursive Nadaraya-Watson estimator of f.

4.1. **Simulated model and estimation procedures.** We consider the semiparametric regression model given, for all $n \geq 1$, by

$$Y_n = f(\theta' X_n) + \varepsilon_n \tag{M}$$

where the link function f is defined, for all $x \in \mathbb{R}$, by

$$f(x) = x\exp\Big(\frac{3x}{4}\Big).$$

The parameter θ belongs to $\mathbb{R}^p$ with $p = 10$ and is given by

$$\theta = \frac{1}{\sqrt{10}}\Big(1, 2, -2, -1, 0, \ldots, 0\Big).$$

Moreover, (X_n) is a sequence of independent random vectors of $\mathbb{R}^p$ sharing the same $\mathcal{N}(0, \mathrm{I}_p)$ distribution, while (ε_n) is a sequence of independent random variables with standard $\mathcal{N}(0, 1)$ distribution, independent of (X_n). In Figure 4.1, we present two scatterplots for a sample of size $n = 1000$ generated from model (M). On the left side, one can observe the data in the "true" reduction subspace, that is the scatterplot of $(\theta' X_1, Y_1), \ldots, (\theta' X_n, Y_n)$ based on the "true" EDR direction θ. On the right side, we plot the data obtained from the estimated EDR direction $\widehat{\theta}_n$ calculated via our recursive SIR procedure, that is the scatterplot of $(\widehat{\theta}_n' X_1, Y_1), \ldots, (\widehat{\theta}_n' X_n, Y_n)$. One can clearly notice that the EDR direction has been well estimated.

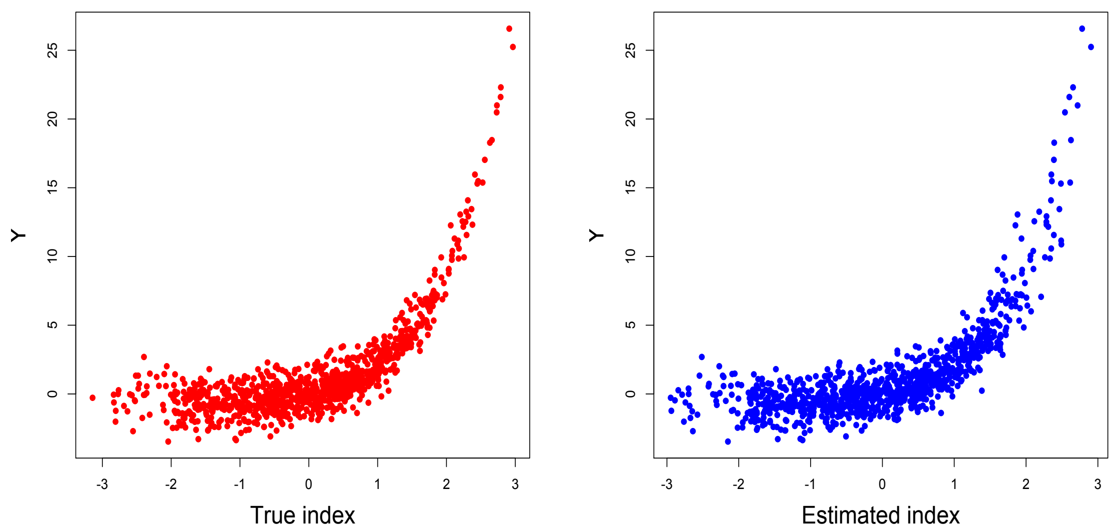

Figure 4.1.
Scatterplots of $(\theta' X_1, Y_1), \ldots, (\theta' X_n, Y_n)$ and $(\widehat{\theta}_n' X_1, Y_1), \ldots, (\widehat{\theta}_n' X_n, Y_n)$.

For the recursive Nadaraya-Watson estimator $\widehat{f}_n$ of f, we have chosen the well-known Epanechnikov kernel

$$K(x) = \frac{3}{4}(1 - x^2)\mathrm{I}_{\{|x| \leq 1\}}$$

and the bandwidth $h_n = 1/n^\alpha$ with $0 < \alpha < 1$. We now need to evaluate an optimal value for the smoothing parameter α. The problem of deciding how much to smooth is of great importance in nonparametric regression. We propose to make use of the optimal data-driven bandwidth α which minimizes the cross-validation criterion

$$CV(\alpha) = \sum_{k=p+1}^{n} (Y_k - \widehat{Y}_{k,\alpha})^2 \qquad \text{where} \qquad \widehat{Y}_{k,\alpha} = \widehat{f}_{k-1}(\widehat{\theta}_{k-1}' X_k).$$

We can observe by simulations that the $CV(\alpha)$ functions are all convex and the corresponding optimal data-driven bandwidth α lies into the interval $[0.33, 0.38]$. Consequently, in all Section 4, we have chosen the optimal value $\alpha = 0.35$.

4.2. **Almost sure convergence.** The good numerical performances of the recursive SIR estimator $\widehat{\theta}_n$ were perviously illustrated in [1], [14]. In order to keep this section brief, we only focus our attention on the almost sure convergence of $\widehat{f}_n$. We generate $N = 1000$ samples of different sizes $n = 200$, 500, 1000, 2000 from model (M) with $p = 10$. For each sample, we calculate the estimation $\widehat{f}_n(\widehat{\theta}'_n x)$ of $f(\theta' x)$ for 10 different values of $x \in \mathbb{R}^p$. The boxplots of the $\widehat{f}_n(\widehat{\theta}'_n x)$'s are given in Figure 4.2. The circle point in each boxplot represents the true value $f(\theta' x)$ to easily judge the quality of the estimations. One can observe that the dispersion of the $\widehat{f}_n(\widehat{\theta}'_n x)$'s are small and the mean is very close to the true value $f(\theta' x)$. One can also notice that the larger is the sample size n, the greater is the quality measure. As it was expected, the quality of the estimation decreases for large values of $f(\theta' x)$ since the number of observations around x decreases, see the scatterplots of Figure 4.1 to be convinced.

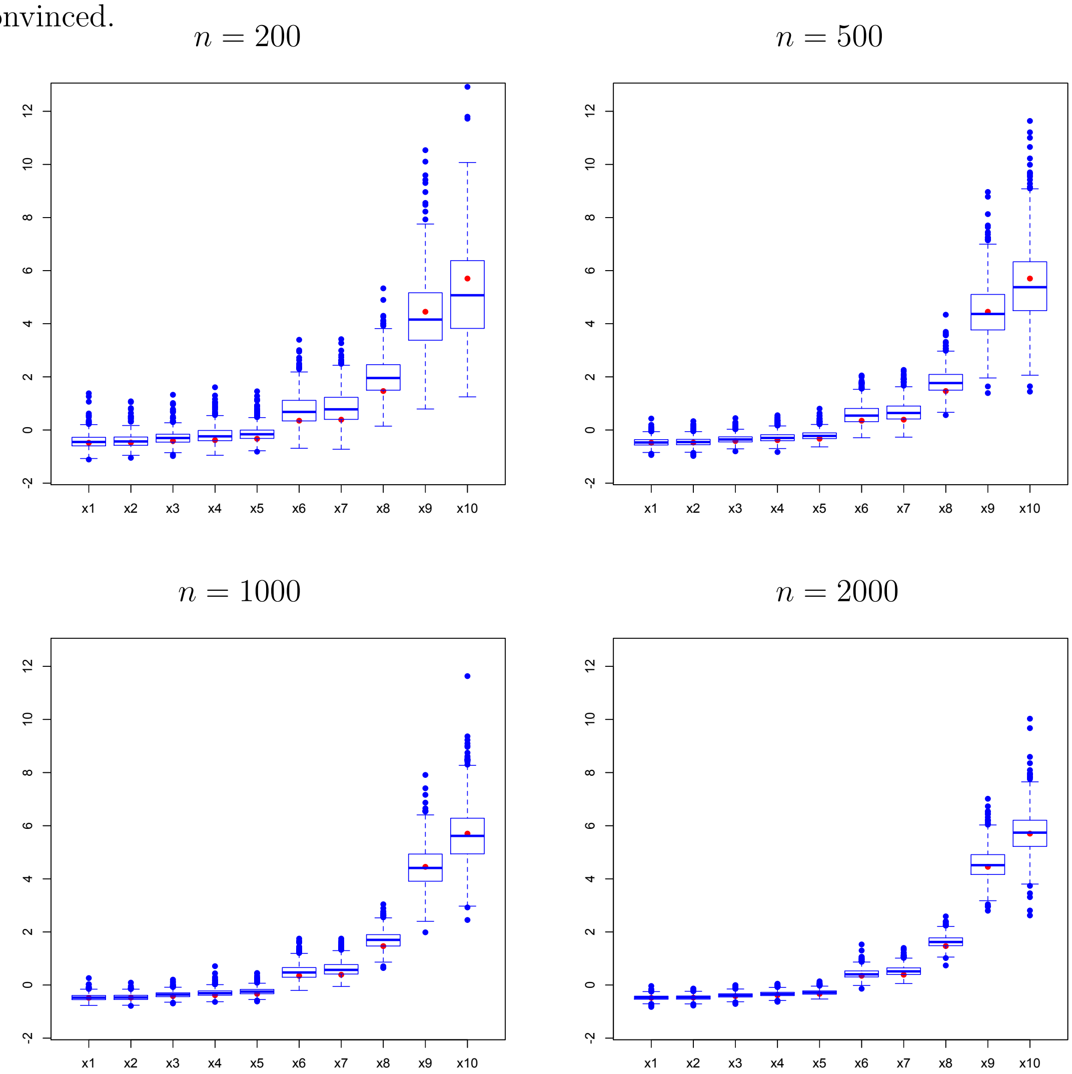

Figure 4.2.
Almost sure convergence of $\widehat{f}_n(\widehat{\theta}'_n x)$ to $f(\theta' x)$ for 10 different values of x.

4.3. **Asymptotic normality.** In order to illustrate the asymptotic normality of our recursive Nadaraya-Watson estimator, we generate $N = 1000$ realizations of $\widehat{f}_n(\widehat{\theta}_n' x)$ for $n = 1000$ from model (M) with $p = 10$. In Figure 4.3, we plot the histogram of the standardized values of the $\widehat{f}_n(\widehat{\theta}_n' x)$'s for 2 different values of $x \in \mathbb{R}^p$. We add the density of the standard normal density on each histogram. One can clearly see that the normal density coincides pretty well with all the histograms, which visually illustrates the asymptotic normality of our recursive Nadaraya-Watson estimator $\widehat{f}_n$ of f.

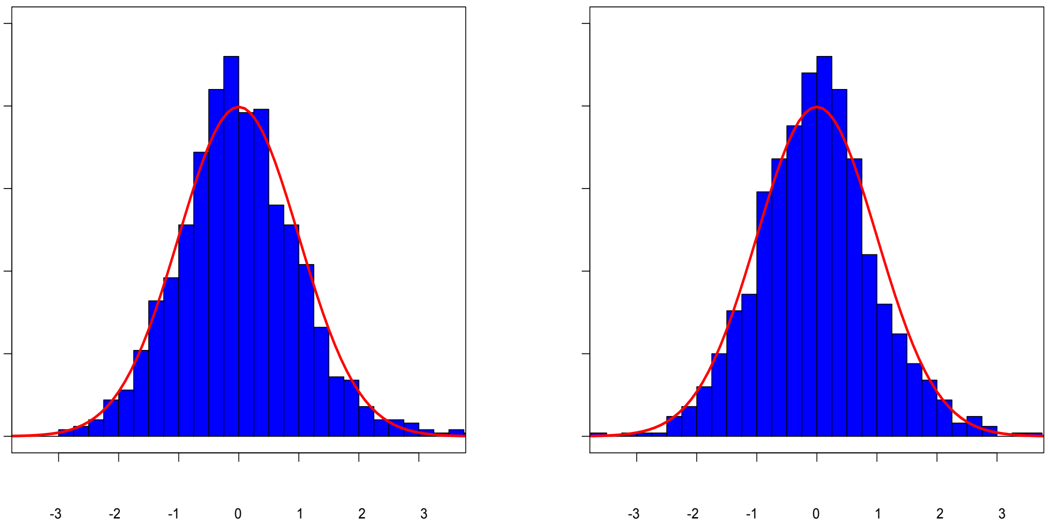

Figure 4.3.
Asymptotic normality of $\widehat{f}_n(\widehat{\theta}_n' x)$ to $f(\theta' x)$ for 2 different values of x.

Appendix A

PROOF OF THEOREM 3.1

In order to prove the almost sure pointwise convergence of Theorem 3.1, we shall denote for all $x \in \mathbb{R}$

$$P_n(x) = \sum_{k=1}^{n} W_k(x)\varepsilon_k, \qquad N_n(x) = \sum_{k=1}^{n} W_k(x),$$

and

$$Q_n(x) = \sum_{k=1}^{n} W_k(x)(f(\Phi_k) - f(x))$$

where $\Phi_n = \theta' X_n$. We clearly obtain from (1.1) the main decomposition

$$\widehat{f}_n(x) - f(x) = \frac{P_n(x) + Q_n(x)}{N_n(x)}. \tag{A.1}$$

We shall establish the asymptotic behavior of each sequence $(P_n(x))$, $(Q_n(x))$ and $(N_n(x))$. Let $(\mathcal{F}_n)$ be the filtration given by $\mathcal{F}_n = \sigma(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$. First of all, we can split $N_n(x)$ into two terms,

$$N_n(x) = M_n^{(N)}(x) + R_n^{(N)}(x) \tag{A.2}$$

where

$$M_n^{(N)}(x) = \sum_{k=1}^{n} \Big(W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] \Big) \quad \text{and} \quad R_n^{(N)}(x) = \sum_{k=1}^{n} \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}].$$

On the one hand, we have

$$\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] = \frac{1}{h_n} \int_{\mathbb{R}^p} K\Big(\frac{x - \widehat{\theta}_{n-1}' x_n}{h_n}\Big) g(x_n)\, dx_n.$$

We can assume without loss of generality that, for n large enough, at least one component of $\widehat{\theta}_n$ is different from zero a.s. As a matter of fact, we already saw from Lemma 2.1 that $\widehat{\theta}_n$ converges a.s. to θ which is different from zero. For the sake of simplicity, suppose that the first component $\widehat{\theta}_{n-1,1} \neq 0$ a.s. We can make the change of variables

$$z = \frac{x - \widehat{\theta}_{n-1}' x_n}{h_n}$$

and $z_2 = x_{n,2}, \ldots, z_p = x_{n,p}$. The Jacobian of this linear transformation is given by

$$J = -\frac{h_n}{\widehat{\theta}_{n-1,1}}.$$

Consequently, we obtain that

$$\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] = \int_{\mathbb{R}} K(z) h(\widehat{\theta}_{n-1}, x - z h_n) dz \tag{A.3}$$

where

$$h(\widehat{\theta}_{n-1}, x) = \frac{1}{|\widehat{\theta}_{n-1,1}|} \int_{\mathbb{R}^{p-1}} g\Big(\frac{1}{\widehat{\theta}_{n-1,1}} \Big(x - \sum_{k=2}^{p} \widehat{\theta}_{n-1,k} z_k\Big), z_2, \ldots, z_p\Big) dz_2 \ldots dz_p.$$

One can observe that $h(\theta, x)$ is exactly the probability density function associated with the identically distributed sequence $(\theta' X_n)$. Therefore, as the probability density function g is continuous, twice differentiable with bounded derivatives, we deduce from (A.3) togheter with Taylor's formula that

$$\begin{aligned}
\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] &= \int_{\mathbb{R}} K(z) \Big(h(\widehat{\theta}_{n-1}, x) - z h_n h'(\widehat{\theta}_{n-1}, x) \\
&\qquad + \frac{z^2 h_n^2}{2} h''(\widehat{\theta}_{n-1}, x - z h_n \xi) \Big) dz, \\
&= h(\widehat{\theta}_{n-1}, x) + \frac{h_n^2}{2} \int_{\mathbb{R}} z^2 K(z) h''(\widehat{\theta}_{n-1}, x - z h_n \xi) dz
\end{aligned}$$

where $0 < \xi < 1$. Consequently, for n large enough,

$$\Big|\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - h(\widehat{\theta}_{n-1}, x)\Big| \leq M_h \tau^2 h_n^2 \qquad \text{a.s.} \tag{A.4}$$

where

$$M_h = \sup_{x\in\mathbb{R}} \Big|h''(\widehat{\theta}_{n-1}, x)\Big| \qquad \text{and} \qquad \tau^2 = \frac{1}{2}\int_{\mathbb{R}} x^2 K(x)dx.$$

Hence, we find from (A.4) that

$$\sum_{k=1}^{n} \Big|\mathbb{E}[W_k(x) \mid \mathcal{F}_{k-1}] - h(\widehat{\theta}_{k-1}, x)\Big| = \mathcal{O}\Big(\sum_{k=1}^{n} h_k^2\Big) \qquad \text{a.s.}$$

It follows from the continuity of h together with the fact that $\widehat{\theta}_n$ converges to θ a.s. and h_n goes to zero that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] = h(\theta, x) \qquad \text{a.s.} \tag{A.5}$$

which of course immediately implies that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{R_n^{(N)}(x)}{n} = h(\theta, x) \qquad \text{a.s.} \tag{A.6}$$

On the other hand, $(M_n^{(N)}(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

$$\begin{aligned} < M^{(N)}(x) >_n &= \sum_{k=1}^{n} \mathbb{E}[(M_k^{(N)}(x) - M_{k-1}^{(N)}(x))^2|\mathcal{F}_{k-1}], \\ &= \sum_{k=1}^{n} \Big(\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}]\Big). \end{aligned}$$

Via the same change of variables as in (A.3), we obtain that

$$\begin{aligned} \mathbb{E}[W_n^2(x)|\mathcal{F}_{n-1}] &= \frac{1}{h_n}\int_{\mathbb{R}} K^2(z) h(\widehat{\theta}_{n-1}, x - z h_n) dz, \\ &= \frac{1}{h_n}\int_{\mathbb{R}} K^2(z)\Big(h(\widehat{\theta}_{n-1}, x) - z h_n h'(\widehat{\theta}_{n-1}, x) \\ &\qquad + \frac{z^2 h_n^2}{2} h''(\widehat{\theta}_{n-1}, x - z h_n \xi)\Big) dz \end{aligned}$$

where $0 < \xi < 1$. Consequently, for n large enough,

$$\Big|\mathbb{E}[W_n^2(x)|\mathcal{F}_{n-1}] - \frac{\nu^2}{h_n} h(\widehat{\theta}_{n-1}, x)\Big| \leq M_h \mu^2 h_n \qquad \text{a.s.} \tag{A.7}$$

where

$$\nu^2 = \int_{\mathbb{R}} K^2(x)dx \qquad \text{and} \qquad \mu^2 = \frac{1}{2}\int_{\mathbb{R}} x^2 K^2(x)dx.$$

Hence, (A.7) ensures that

$$\sum_{k=1}^{n}\Big|\mathbb{E}[W_k^2(x) \mid \mathcal{F}_{k-1}] - \frac{\nu^2}{h_k}h(\widehat{\theta}_{k-1},x)\Big| = \mathcal{O}\Big(\sum_{k=1}^{n} h_k\Big) \qquad \text{a.s.}$$

However, it is not hard to see that

$$\lim_{n\to\infty}\frac{1}{n^{1+\alpha}}\sum_{k=1}^{n}\frac{1}{h_k} = \frac{1}{1+\alpha}.$$

Therefore, it follows from (A.7) together with the almost sure convergence of $h(\widehat{\theta}_n,x)$ to $h(\theta,x)$ and Toeplitz's lemma that

$$\lim_{n\to\infty}\frac{1}{n^{1+\alpha}}\sum_{k=1}^{n}\mathbb{E}[W_k^2(x) \mid \mathcal{F}_{k-1}] = \frac{\nu^2}{1+\alpha}h(\theta,x) \qquad \text{a.s.} \tag{A.8}$$

Furthermore, we also have from (A.4) that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}] = h^2(\theta,x) \qquad \text{a.s.} \tag{A.9}$$

Consequently, we deduce from (A.8) and (A.9) that for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty}\frac{<M^{(N)}(x)>_n}{n^{1+\alpha}} = \frac{\nu^2}{1+\alpha}h(\theta,x) \qquad \text{a.s.} \tag{A.10}$$

We are now in position to make use of the strong law of large numbers for martingales given e.g. by Theorem 1.3.15 of [8]. As the probability density function g is positive on its support, we have for all $x \in \mathbb{R}$, $h(\theta,x) > 0$, which implies that $<M^{(N)}(x)>_n$ goes to infinity a.s. Hence, for any $\gamma > 0$, $(M_n^{(N)}(x))^2 = o(n^{1+\alpha}(\log n)^{1+\gamma})$ a.s. which leads to

$$M_n^{(N)}(x) = o(n) \qquad \text{a.s.} \tag{A.11}$$

Then, we obtain from (A.2), (A.6) and (A.11) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty}\frac{N_n(x)}{n} = h(\theta,x) \qquad \text{a.s.} \tag{A.12}$$

We shall now investigate the asymptotic behavior of the sequence $(P_n(x))$. Since (X_n) and (ε_n) are independent, $(P_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

$$<P(x)>_n = \sum_{k=1}^{n}\mathbb{E}[(P_k(x)-P_{k-1}(x))^2|\mathcal{F}_{k-1}] = \sigma^2\sum_{k=1}^{n}\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}].$$

Then, it follows from convergence (A.8) that

$$\lim_{n\to\infty}\frac{<P(x)>_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2}{1+\alpha}h(\theta,x) \qquad \text{a.s.} \tag{A.13}$$

Consequently, we obtain from the strong law of large numbers for martingales that for any $\gamma > 0$ and that for all $x \in \mathbb{R}$,

$$P_n(x) = o\Big(\sqrt{n^{1+\alpha}(\log n)^{1+\gamma}}\Big) = o(n) \qquad \text{a.s.} \tag{A.14}$$

It remains to study the asymptotic behavior of the sequence $(Q_n(x))$. We can split $Q_n(x)$ into two terms,

$$Q_n(x) = \Sigma_n(x) + \Delta_n(x) \tag{A.15}$$

where $\widehat{\Phi}_n = \widehat{\theta}_{n-1}' X_n$,

$$\Sigma_n(x) = \sum_{k=1}^{n} W_k(x)(f(\Phi_k) - f(\widehat{\Phi}_k)) \quad \text{and} \quad \Delta_n(x) = \sum_{k=1}^{n} W_k(x)(f(\widehat{\Phi}_k) - f(x)).$$

The right-hand side of (A.15) is easy to handle. As a matter of fact, the kernel K is compactly supported which means that one can find a positive constant A such that K vanishes outside the interval $[-A, A]$. Thus, for all $n \geq 1$ and all $x \in \mathbb{R}$,

$$W_n(x) = \frac{1}{h_n} K\Big(\frac{x - \widehat{\theta}_{n-1}' X_n}{h_n}\Big) \mathrm{I}_{\{|\widehat{\theta}_{n-1}' X_n - x| \leq A h_n\}}.$$

In addition, the function f is Lipschitz, so it exists a positive constant C_f such that for all $n \geq 1$

$$|f(\widehat{\Phi}_n) - f(x)| \leq C_f |\widehat{\Phi}_n - x| \leq C_f |\widehat{\theta}_{n-1}' X_n - x|. \tag{A.16}$$

Consequently, we obtain from (A.16) that for all $x \in \mathbb{R}$

$$\begin{aligned} |\Delta_n(x)| &\leq C_f \sum_{k=1}^{n} W_k(x) |\widehat{\theta}_{k-1}' X_k - x|, \\ &\leq A C_f \sum_{k=1}^{n} h_k W_k(x). \end{aligned} \tag{A.17}$$

Moreover, via the same lines as in the proof of (A.5), we find that

$$\lim_{n \to \infty} \frac{1}{n^{1-\alpha}} \sum_{k=1}^{n} h_k \mathbb{E}[W_k(x) | \mathcal{F}_{k-1}] = \frac{1}{1-\alpha} h(\theta, x) \qquad \text{a.s.} \tag{A.18}$$

Furthermore, denote

$$M_n^{(\Delta)}(x) = \sum_{k=1}^{n} h_k \Big(W_k(x) - \mathbb{E}[W_k(x) | \mathcal{F}_{k-1}]\Big).$$

One can observe that $(M_n^{(\Delta)}(x))$ is a square integrable martingale difference sequence with bounded increments and predictable quadratic variation given by

$$\begin{aligned} <M^{(\Delta)}(x)>_n &= \sum_{k=1}^n \mathbb{E}[(M_k^{(\Delta)}(x) - M_{k-1}^{(\Delta)}(x))^2 | \mathcal{F}_{k-1}], \\ &= \sum_{k=1}^n h_k^2 \Big(\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}] \Big). \end{aligned}$$

Hence, it follows from (A.4) and (A.7) together with the almost sure convergence of $h(\widehat{\theta}_n, x)$ to $h(\theta, x)$ and Toeplitz's lemma that

$$\lim_{n\to\infty} \frac{<M^{(\Delta)}(x)>_n}{n^{1-\alpha}} = \frac{\nu^2}{1-\alpha} h(\theta, x) \qquad \text{a.s.} \tag{A.19}$$

Consequently, we obtain from the strong law of large numbers for martingales that

$$\Big(M_n^{(\Delta)}(x) \Big)^2 = \mathcal{O}\Big(n^{1-\alpha} \log n \Big) \qquad \text{a.s.} \tag{A.20}$$

Then, we infer from the conjunction of (A.17), (A.18) and (A.20) that for all $x \in \mathbb{R}$

$$|\Delta_n(x)| = \mathcal{O}\Big(n^{1-\alpha} \Big) \qquad \text{a.s.} \tag{A.21}$$

The left-hand side of (A.15) is much more difficult to handle. We can use once again the assumption that the function f is Lipschitz to deduce that it exists a positive constant C_f such that for all $n \geq 1$

$$|f(\widehat{\Phi}_n) - f(\Phi_n)| \leq C_f |\pi_n| \tag{A.22}$$

where $\pi_n = (\widehat{\theta}_{n-1} - \theta)' X_n$. Hence, it immediately follows from (A.22) that for all $x \in \mathbb{R}$

$$|\Sigma_n(x)| \leq C_f \sum_{k=1}^n W_k(x) |\pi_k|. \tag{A.23}$$

Denote

$$\mathcal{A}_n = \Big\{ |\widehat{\theta}'_{n-1} X_n - x| \leq A h_n \Big\} \quad \text{and} \quad \mathcal{B}_n = \Big\{ |\theta' X_n - x| \leq A h_n + b_n \Big\}$$

where (b_n) is a sequence of positive real numbers which will be explicitely given later. On the one hand, we immediately have from the triangle inequality that on the set $\mathcal{A}_n \cap \mathcal{B}_n$,

$$|\pi_n| \leq 2Ah_n + b_n.$$

On the other hand, we also have on the set $\mathcal{A}_n \cap \overline{\mathcal{B}_n}$,

$$Ah_n + b_n < |\theta' X_n - x| \leq |\pi_n| + |\widehat{\theta}'_{n-1} X_n - x| \leq |\pi_n| + Ah_n$$

which implies that $|\pi_n| > b_n$. Consequently, we obtain from (A.23) that

$$|\Sigma_n(x)| \leq 2AC_f \sum_{k=1}^n h_k W_k(x) + C_f \sum_{k=1}^n b_k W_k(x) + C_f \sum_{k=1}^n W_k(x) |\pi_k| \mathrm{I}_{\{|\pi_k| > b_k\}}. \tag{A.24}$$

We already saw from (A.21) that

$$\sum_{k=1}^{n} h_k W_k(x) = \mathcal{O}\Big(n^{1-\alpha}\Big) \qquad \text{a.s.} \tag{A.25}$$

Moreover, it is assumed that the sequence (X_n) has a finite moment of order $a > 2$ which ensures that

$$\sup_{1\leq k\leq n} ||X_k|| = o(n^{1/a}) \qquad \text{a.s.}$$

Consequently, we find from Lemma (2.1) that

$$|\pi_n| = o(b_n) \qquad \text{a.s.} \tag{A.26}$$

where we can choose

$$b_n = n^{1/a}\sqrt{\frac{\log(\log n)}{n}}.$$

Therefore, we clearly have

$$\sum_{k=1}^{n} W_k(x)|\pi_k|\mathrm{I}_{\{|\pi_k|>b_k\}} < +\infty \qquad \text{a.s.} \tag{A.27}$$

Furthermore, it is not hard to see that

$$\sum_{k=1}^{n} b_k = \mathcal{O}\Big(n^{1/a}\sqrt{n\log(\log n)}\Big).$$

Hence, via the same lines as in the proof of (A.21), we obtain that

$$\sum_{k=1}^{n} b_k W_k(x) = \mathcal{O}\Big(n^{1/a}\sqrt{n\log(\log n)}\Big) \qquad \text{a.s.} \tag{A.28}$$

Then, we deduce from the conjunction of (A.24), (A.25), (A.27), and (A.28) that

$$|\Sigma_n(x)| = \mathcal{O}\Big(n^{1-\alpha}\Big) + \mathcal{O}\Big(n^{1/a}\sqrt{n\log(\log n)}\Big) \qquad \text{a.s.} \tag{A.29}$$

Consequently, we infer from (A.21) and (A.29) that for all $x \in \mathbb{R}$

$$Q_n(x) = \mathcal{O}\Big(n^{1-\alpha}\Big) + \mathcal{O}\Big(n^{1/a}\sqrt{n\log(\log n)}\Big) \qquad \text{a.s.} \qquad \text{a.s.} \tag{A.30}$$

Finally, we can conclude from (A.1) together with (A.12), (A.14) and (A.30) that

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \qquad \text{a.s.}$$

with the almost sure rates of convergence given by (3.2) and (3.3), which completes the proof of Theorem 3.1. $\square$

Appendix B

PROOF OF THEOREM 3.2

We already saw that $(P_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation satisfying

$$\lim_{n\to\infty} \frac{<P(x)>_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2}{1+\alpha} h(\theta, x) \qquad \text{a.s.}$$

In order to establish the asymptotic normality of Theorem 3.2, it is necessary to prove that the sequence $(P_n(x))$ satisfies the Lindeberg condition, that is for all $\varepsilon > 0$,

$$\mathcal{P}_n(x) = \frac{1}{n^{1+\alpha}} \sum_{k=1}^{n} \mathbb{E}\left[|\Delta P_k(x)|^2 \mathrm{I}_{\{|\Delta P_k(x)| \geq \varepsilon\sqrt{n^{1+\alpha}}\}} |\mathcal{F}_{k-1}\right] \xrightarrow{\mathcal{P}} 0 \tag{B.1}$$

where $\Delta P_n(x) = P_n(x) - P_{n-1}(x)$. We have assumed that the sequence (ε_n) has a finite conditional moment of order $b > 2$ which means that

$$\sup_{n\geq 0} \mathbb{E}[|\varepsilon_n|^b|\mathcal{F}_{n-1}] < +\infty \qquad \text{a.s.}$$

Consequently, for all $\varepsilon > 0$, we have

$$\begin{aligned}
\mathcal{P}_n(x) &\leq \frac{1}{\varepsilon^{b-2}n^c} \sum_{k=1}^{n} \mathbb{E}[|\Delta P_k(x)|^b|\mathcal{F}_{k-1}], \\
&\leq \frac{1}{\varepsilon^{b-2}n^c} \sum_{k=1}^{n} \mathbb{E}[W_k^b(x)|\mathcal{F}_{k-1}]\mathbb{E}[|\varepsilon_k|^b|\mathcal{F}_{k-1}], \\
&\leq \frac{1}{\varepsilon^{b-2}n^c} \sup_{1\leq k\leq n} \mathbb{E}[|\varepsilon_k|^b|\mathcal{F}_{k-1}] \sum_{k=1}^{n} \mathbb{E}[W_k^b(x)|\mathcal{F}_{k-1}] \qquad \text{(B.2)}
\end{aligned}$$

where $c = b(1+\alpha)/2$. In addition, via the same lines as in the proof of (A.8), we obtain that

$$\lim_{n\to\infty} \frac{1}{n^{1+\alpha(b-1)}} \sum_{k=1}^{n} \mathbb{E}[W_k^b(x) \mid \mathcal{F}_{k-1}] = \frac{\xi^b}{1+\alpha(b-1)} h(\theta, x) \qquad \text{a.s.} \tag{B.3}$$

where

$$\xi^b = \int_{\mathbb{R}} K^b(x)\,dx.$$

Therefore, we deduce from (B.1) together with (B.2) and (B.3) that, for all $\varepsilon > 0$,

$$\mathcal{P}_n(x) = \mathcal{O}(n^d) \qquad \text{a.s.}$$

where $d = (2-b)(1-\alpha)/2$. We recall that $b > 2$ which means that $d < 0$. It ensures that the Lindeberg condition is satisfied. Hence, it follows from the central limit theorem for martingales given e.g. by Corollary 2.1.10 of [8] that for all $x \in \mathbb{R}$,

$$\frac{P_n(x)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma^2\nu^2}{1+\alpha} h(\theta, x)\Big). \tag{B.4}$$

Furthermore, as soon as $a \geq 6$ and $1/3 < \alpha < 1$, we clearly obtain from (A.30) that

$$\lim_{n\to\infty} \frac{Q_n(x)}{\sqrt{n^{1+\alpha}}} = 0 \qquad \text{a.s.} \tag{B.5}$$

Finally, we find from (A.1) together with (A.12), (B.4), (B.5) and Slutsky's lemma that, for all $x \in \mathbb{R}$,

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma^2\nu^2}{(1+\alpha)h(\theta,x)}\Big)$$

which acheives the proof of Theorem 3.2. □

Acknowledgements. The first author would like to thanks Bruno Portier for helpful remarks made on a preliminary version of the paper.

References

[1] Bercu, B., Nguyen, T.M.N., and Saracco, J., A new approach on recursive and non-recursive SIR methods, Journal of the Korean Statistical Society, 41(1), pp. 17-36, 2012.
[2] Bercu, B., and Portier, B., Kernel density estimation and goodness-of-fit test in adaptive tracking, SIAM J. Control Optim. 47, pp. 2440-2457, 2008.
[3] Chen, C.-H., and Li, K.-C., Can SIR be as popular as multiple linear regression?, Statistica Sinica, 8(2), pp. 289-316, 1998.
[4] Choi, E., Hall, P., and Rousson, V., Data sharpening methods for bias reduction in nonparametric regression, Ann. Statist. 28, pp. 1339-1355, 2000.
[5] Cook, R. D.,and Ni, L., Sufficient dimension reduction via inverse regression: a minimum discrepancy approach, Journal of American Statistical Association, 100, pp. 410-418, 2005.
[6] Devroye, L., and Lugosi, G., Combinatorial methods in density estimation, Springer Series in Statistics. Springer-Verlag, New York, 2001.
[7] Duan, N., and Li, K.-C., Slicing regression: A link-free regression method, Ann. Stat. 19(2), pp. 505-530,1991.
[8] Duflo, M., Random Iterative Models, Springer-Verlag, Berlin, 1997.
[9] Hall, P., and Heyde, C. C., Martingale limit theory and its application, Academic Press Inc. New York, 1980.
[10] Hall, P., and Li, K.-C., On almost linearity of low-dimensional projections from high-dimensional data, Annals of Statistics, 21(2), pp. 867-889, 1993.
[11] Li, K.-C., Sliced inverse regression for dimension reduction (with discussions), J. Am. Stat. Assoc. 86(414), pp. 316-342, 1991.
[12] Nadaraja, E. A., On a regression estimate, Teor. Verojatnost. i Primenen. 9, pp. 157-159, 1964.
[13] Nadaraya, E. A., Nonparametric estimation of probability densities and regression curves, Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1989.
[14] Nguyen, T.M.N., and Saracco, J., Recursive estimation for sliced inverse regression (in French), Journal de la Société Française de Statistique, Vol. 151(2), pp. 19-46, 2010.
[15] Noda, K., Estimation of a regression function by the Parzen kernel-type density estimators, Ann. Inst. Statist. Math. 28, pp. 221-234, 1976.
[16] Parzen, E., On estimation of a probability density function and mode, Ann. Math. Statist. 33, pp. 1065-1076, 1972.
[17] Rosenblatt, M., Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, pp. 832-837, 1956.
[18] Schuster, E. F., Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann. Math. Statist. 43, pp. 84-88, 1972.

[19] Silverman, B. W., Density estimation for statistics and data analysis, Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1986.

[20] Stout, W. F. , Almost sure convergence, Academic Press, New York, 1974.

[21] Tsybakov, A. B., Introduction à l'estimation non-paramétrique, Mathématiques et Applications, Springer-Verlag, Berlin, 2004.

[22] Watson, G. S., Smooth regression analysis, Sankhya Ser. A 26, pp. 359-372, 1964.

E-mail address: `Bernard.Bercu@math.u-bordeaux1.fr`
E-mail address: `Nguyen@math.unistra.fr`
E-mail address: `Jerome.Saracco@math.u-bordeaux1.fr`

Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, 351 cours de la libération, 33405 Talence cedex, France.

Université de Strasbourg, Institut de Recherche Mathématique Avancée, UMR CNRS 7501, 7 rue René Descartes 67084 Strasbourg cedex, France.

Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, 351 cours de la libération, 33405 Talence cedex, France.