A Robbins-Monro procedure for estimation in semiparametric regression models

Bernard Bercu, Philippe Fraysse

A ROBBINS-MONRO PROCEDURE FOR ESTIMATION IN SEMIPARAMETRIC REGRESSION MODELS

BERNARD BERCU AND PHILIPPE FRAYSSE

Université Bordeaux 1

ABSTRACT. This paper is devoted to the parametric estimation of a shift together with the nonparametric estimation of a regression function in a semiparametric regression model. We implement a Robbins-Monro procedure very efficient and easy to handle. On the one hand, we propose a stochastic algorithm similar to that of Robbins-Monro in order to estimate the shift parameter. A preliminary evaluation of the regression function is not necessary for estimating the shift parameter. On the other hand, we make use of a recursive Nadaraya-Watson estimator for the estimation of the regression function. This kernel estimator takes in account the previous estimation of the shift parameter. We establish the almost sure convergence for both Robbins-Monro and Nadaraya-Watson estimators. The asymptotic normality of our estimates is also provided.

1. INTRODUCTION

Our purpose is to investigate the parametric estimation of a shift parameter θ together with the nonparametric estimation of a regression function f in the semiparametric regression model given, for all $n \geq 0$, by

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

where (X_n) and (ε_n) are two independent sequences of independent and identically distributed random variables. First of all, we implement a Robbins-Monro procedure in order to estimate the unknown parameter θ without any preliminary evaluation of the regression function f. Our approach is very easy to handle and it performs very well. Moreover, our approach is totally different from the one recently proposed by Dalalyan, Golubev and Tsybakov [6] in the Gaussian white noise case. Firstly, a penalized maximum likelihood estimator of θ is proposed in [6] with an appropriately chosen penalty based on a Fourier series approximation of the function f. Secondly, the asymptotic behavior of the mean square risk of this estimator is investigated. One can observe that our estimator is much more easy to calculate. In addition, we do not require any assumption on the derivatives of the function f. In the situation where the parameter θ is random, Castillo and Loubes [3] propose a plug-in version of the Parzen-Rosenblatt density estimator of θ. The construction of this estimate also relies on the penalized maximum likelihood estimator of θ given in [6]. Furthermore,

2010 *Mathematics Subject Classification.* Primary: 62G05, Secondary: 62G20.

Key words and phrases. Semiparametric estimation, estimation of a shift, estimation of a regression function, asymptotic properties.

in the case where one observe several Gaussian functions differing from each other by a translation parameter, Gamboa and Loubes [10] propose to transform the starting model by using a discrete Fourier transform. Hence, from the resulting model, they estimate the shift parameters by minimizing a quadratic functional. This approach is very interesting by the few assumptions made on the regression function. Our alternative approach to estimate θ is associated to a stochastic recursive algorithm similar to that of Robbins-Monro [26], [27].

Assume that one can find a function ϕ, free of the parameter θ, such that $\phi(\theta) = 0$. Then, it is possible to estimate θ by the Robbins-Monro algorithm

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + \gamma_n T_{n+1} \tag{1.2}$$

where (γ_n) is a positive sequence of real numbers decreasing towards zero and (T_n) is a sequence of random variables such that

$$\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n)$$

where $\mathcal{F}_n$ stands for the σ-algebra of the events occurring up to time n. Under standard conditions on the function ϕ and on the sequence (γ_n), it is well-known [9], [18] that $\widehat{\theta}_n$ tends to θ almost surely. The asymptotic normality of $\widehat{\theta}_n$ together with the quadratic strong law may also be found in [12], [20] and [25]. A randomly truncated version of the Robbins-Monro algorithm is also given in [4], [19].

Our second goal is the estimation of the unknown regression function f. A wide range of literature is available on nonparametric estimation of a regression function. We refer the reader to [8], [22], [30], [32] for some excellent books on density and regression function estimation. Here, we focus our attention on the Nadaraya-Watson estimator of f. The almost sure convergence of the Nadaraya-Watson estimator [21], [33] without the shift θ was established by Noda [23], see also Härdle et al [14], [15] for the law of iterated logarithm and the uniform strong law. A nice extension of the previous results may be found in [16]. The asymptotic normality of the Nadaraya-Watson estimator was proved by Schuster [29]. Moreover, Choi, Hall and Rousson [5] propose three data-sharpening versions of the Nadaraya-Watson estimator in order to reduce the asymptotic variance in the central limit theorem. Furthermore, in the situation where the regression function is monotone, Hall and Huang [13] provide a method for monotonizing the Nadaraya-Watson estimator. For n large enough, their alternative estimator coincides with the standard Nadaraya-Watson estimator on a compact interval where the regression function f is monotone. In our situation, we propose to make use of a recursive Nadaraya-Watson estimator [9] of f which takes into account the previous estimation of the shift parameter θ. It is given, for all $x \in \mathbb{R}$, by

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^n W_k(x) Y_k}{\sum_{k=1}^n W_k(x)} \tag{1.3}$$

with

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

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. The main difficulty arising here is that we have to deal with the additional term $\widehat{\theta}_n$ inside the kernel K. Consequently, we are led to analyse a double stochastic algorithm with, at the same time, the study of the asymptotic behavior of the Robbins-Monro estimator $\widehat{\theta}_n$ of θ, and the Nadaraya-Watson estimator $\widehat{f}_n$ of f.

The paper is organized as follows. Section 2 is devoted to the parametric estimation of θ. We establish the almost sure convergence of $\widehat{\theta}_n$ as well as a law of iterated logarithm and the asymptotic normality. Section 3 deals with the nonparametric estimation of f. Under standard regularity assumptions on the kernel K, we prove the almost sure pointwise convergence of $\widehat{f}_n$ to f. In addition, we also establish the asymptotic normality of $\widehat{f}_n$. The proofs of the parametric results are given is Section 4, while those concerning the nonparametric results are postponed in Section 5.

2. ESTIMATION OF THE SHIFT

First of all, we focus our attention on the estimation of the shift parameter θ in the semiparametric regression model given by (1.1). We assume that (ε_n) is a sequence of independent and identically distributed random variables with zero mean and positive variance σ^2. Moreover, it is necessary to make several hypothesis similar to that of [6].

- $(\mathcal{H}_1)$ The sequence (X_n) is independent and identically distributed with symmetric probability density function g, positive on its support $[-1/2, 1/2]$. In addition, the function g is continuous, twice differentiable with bounded derivatives.
- $(\mathcal{H}_2)$ The function f is symmetric, bounded, periodic with period 1.

Let X be a random variable sharing the same distribution as (X_n). In all the sequel, the auxiliary function ϕ defined, for all $t \in \mathbb{R}$, by

$$\phi(t) = \mathbb{E}\Big[\frac{\sin(2\pi(X-t))}{g(X)} f(X-\theta)\Big] \tag{2.1}$$

will play a prominent role. More precisely, it follows from the periodicity of f that

$$\begin{aligned}\phi(t) &= \int_{-1/2}^{1/2} \sin(2\pi(x-t)) f(x-\theta)\,dx = \int_{-1/2}^{1/2} \sin(2\pi(y+\theta-t)) f(y)\,dy, \\ &= \sin(2\pi(\theta-t)) \int_{-1/2}^{1/2} \cos(2\pi y) f(y)\,dy + \cos(2\pi(\theta-t)) \int_{-1/2}^{1/2} \sin(2\pi y) f(y)\,dy.\end{aligned}$$

Consequently, the symmetry of f leads to

$$\phi(t) = \sin(2\pi(\theta-t)) f_1 \tag{2.2}$$

where f_1 is the first Fourier coefficient of f

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x)\,dx.$$

Throughout the paper, we assume that $f_1 \neq 0$. Obviously, ϕ is a continuous and bounded function such that $\phi(\theta) = 0$. In addition, one can easily verify that for all $t \in \mathbb{R}$ such that $|t - \theta| < 1/2$, the product $(t - \theta)\phi(t)$ has a constant sign. It is negative if $f_1 > 0$, while it is positive if $f_1 < 0$. Therefore, we are in position to implement our Robbins-Monro procedure [26], [27]. Let $K = [-1/4, 1/4]$ and denote by π_K the projection on the compact set K defined, for all $x \in \mathbb{R}$, by

$$\pi_K(x) = \begin{cases} x & \text{if } |x| \leq 1/4, \\ 1/4 & \text{if } x \geq 1/4, \\ -1/4 & \text{if } x \leq -1/4. \end{cases}$$

Let (γ_n) be a decreasing sequence of positive real numbers satisfying

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \qquad \text{and} \qquad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty. \tag{2.3}$$

For the sake of clarity, we shall make use of $\gamma_n = 1/n$. We estimate the shift parameter θ via the projected Robbins-Monro algorithm

$$\widehat{\theta}_{n+1} = \pi_K\Big(\widehat{\theta}_n + \text{sign}(f_1)\gamma_{n+1}T_{n+1}\Big) \tag{2.4}$$

where the initial value $\widehat{\theta}_0 \in K$ and the random variable T_{n+1} is defined by

$$T_{n+1} = \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))}{g(X_{n+1})} Y_{n+1}. \tag{2.5}$$

Our first result concerns the almost sure convergence of the estimator $\widehat{\theta}_n$.

Theorem 2.1. *Assume that $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ hold and that $|\theta| < 1/4$. Then, $\widehat{\theta}_n$ converges almost surely to θ. In addition, the number of times that the random variable $\widehat{\theta}_n + sign(f_1)\gamma_{n+1}T_{n+1}$ goes outside of K is almost surely finite.*

In order to establish the asymptotic normality of $\widehat{\theta}_n$, it is necessary to introduce a second auxiliary function φ defined, for all $t \in \mathbb{R}$, by

$$\begin{aligned} \varphi(t) &= \mathbb{E}\Big[\frac{\sin^2(2\pi(X - t))}{g^2(X)}(f^2(X - \theta) + \sigma^2)\Big], \\ &= \int_{-1/2}^{1/2} \frac{\sin^2(2\pi(x - t))}{g(x)}(f^2(x - \theta) + \sigma^2)\, dx. \end{aligned} \tag{2.6}$$

As soon as $4\pi|f_1| > 1$, denote

$$\xi^2(\theta) = \frac{\varphi(\theta)}{4\pi|f_1| - 1}.$$

Theorem 2.2. *Assume that $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ hold and that $|\theta| < 1/4$. In addition, suppose that (ε_n) has a finite moment of order > 2 and that $4\pi|f_1| > 1$. Then, we have the asymptotic normality*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2(\theta)). \tag{2.7}$$

Remark 2.1. *In the particular case where $4\pi|f_1| = 1$, it is also possible to show that*

$$\sqrt{\frac{n}{\log(n)}}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \varphi(\theta)).$$

Asymptotic results are also available when $0 < 4\pi|f_1| < 1$. However, we have chosen to focus our attention on the more attractive case $4\pi|f_1| > 1$.

Theorem 2.3. *Assume that $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ hold and that $|\theta| < 1/4$. In addition, suppose that (ε_n) has a finite moment of order > 2 and that $4\pi|f_1| > 1$. Then, we have the law of iterated logarithm*

$$\begin{aligned} \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\widehat{\theta}_n - \theta) &= -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\widehat{\theta}_n - \theta) \\ &= \xi(\theta) \qquad a.s. \end{aligned} \tag{2.8}$$

In particular,

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) (\widehat{\theta}_n - \theta)^2 = \xi^2(\theta) \qquad a.s. \tag{2.9}$$

In addition, we also have the quadratic strong law

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} (\widehat{\theta}_k - \theta)^2 = \xi^2(\theta) \qquad a.s. \tag{2.10}$$

Proof. The proofs are given in Section 4. □

3. ESTIMATION OF THE REGRESSION FUNCTION

This section is devoted to the nonparametric estimation of the regression function f via a recursive Nadaraya-Watson estimator. On the one hand, we add the standard hypothesis

$(\mathcal{H}_3)$ The regression function f is Lipschitz.

On the other hand, we recall that under $(\mathcal{H}_2)$, the function f is assumed to be symmetric. Consequently, we follow the same approach as the one developed by Stone [31] for the estimation of a symmetric probability density function replacing the estimator (1.3) by its symmetrized version

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^n (W_k(x) + W_k(-x))Y_k}{\sum_{k=1}^n (W_k(x) + W_k(-x))} \tag{3.1}$$

where

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

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 $\alpha \in]0, 1[$. Moreover, we shall assume in all the sequel that the kernel K is a positive symmetric function, bounded with compact support, twice differentiable with bounded derivatives, satisfying

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

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

Theorem 3.1. *Assume that $(\mathcal{H}_1)$, $(\mathcal{H}_2)$ and $(\mathcal{H}_3)$ hold and that the sequence (ε_n) has a finite moment of order > 2. Then, for any $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \qquad a.s. \tag{3.2}$$

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

Theorem 3.2. *Assume that $(\mathcal{H}_1)$, $(\mathcal{H}_2)$ and $(\mathcal{H}_3)$ hold and that the sequence (ε_n) has a finite moment of order > 2. Then, as soon as the bandwidth (h_n) satisfies $h_n = 1/n^{\alpha}$ with $\alpha > 1/3$, we have for any $x \in \mathbb{R}$ with $x \neq 0$, the pointwise asymptotic normality*

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma^2\nu^2}{(1+\alpha)(g(\theta+x)+g(\theta-x))}\Big). \tag{3.3}$$

In addition, for $x = 0$,

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

Proof. The proofs are given in Section 5. □

4. PROOFS OF THE PARAMETRIC RESULTS

4.1. **Proof of Theorem 2.1.** We can assume without loss of generality that $f_1 > 0$ inasmuch as the proof for $f_1 < 0$ follows exactly the same lines. Denote by $\mathcal{F}_n$ the σ-algebra of the events occurring up to time n, $\mathcal{F}_n = \sigma(X_0, \varepsilon_0, \ldots, X_n, \varepsilon_n)$. First of all, we shall calculate the two first conditional moments of the random variable T_n given by (2.5). It follows from (1.1) that

$$\begin{aligned}
\mathbb{E}[T_{n+1}|\mathcal{F}_n] &= \mathbb{E}\Big[\frac{\sin(2\pi(X_{n+1}-\widehat{\theta}_n))Y_{n+1}}{g(X_{n+1})}|\mathcal{F}_n\Big], \\
&= \mathbb{E}\Big[\frac{\sin(2\pi(X_{n+1}-\widehat{\theta}_n))(f(X_{n+1}-\theta)+\varepsilon_{n+1})}{g(X_{n+1})}|\mathcal{F}_n\Big].
\end{aligned}$$

On the one hand, as (X_n) is a sequence of independent random variables sharing the same distribution as a random variable X, we have

$$\mathbb{E}\Big[\frac{\sin(2\pi(X_{n+1}-\widehat{\theta}_n))f(X_{n+1}-\theta)}{g(X_{n+1})}|\mathcal{F}_n\Big] = \phi(\widehat{\theta}_n) \qquad \text{a.s.} \tag{4.1}$$

where ϕ is the function given by (2.1). On the other hand, as (X_n) and (ε_n) are two independent sequences and (ε_n) is a sequence of independent and square integrable random variables with zero mean, we also have

$$\mathbb{E}\Big[\frac{\sin(2\pi(X_{n+1}-\widehat{\theta}_n))\varepsilon_{n+1}}{g(X_{n+1})}|\mathcal{F}_n\Big] = \mathbb{E}\Big[\frac{\sin(2\pi(X-\widehat{\theta}_n))}{g(X)}\Big]\mathbb{E}[\varepsilon_{n+1}] = 0.$$

Hence, (4.1) leads to

$$\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n) \qquad \text{a.s.} \tag{4.2}$$

On the other hand,

$$\begin{aligned} T_{n+1}^2 &= \frac{\sin^2(2\pi(X_{n+1}-\widehat{\theta}_n))Y_{n+1}^2}{g^2(X_{n+1})}, \\ &= \frac{\sin^2(2\pi(X_{n+1}-\widehat{\theta}_n))(f^2(X_{n+1}-\theta)+2\varepsilon_{n+1}f(X_{n+1}-\theta)+\varepsilon_{n+1}^2)}{g^2(X_{n+1})}. \end{aligned}$$

Consequently, as the function f is bounded, the density g is positive on $[-1/2, 1/2]$, and $\mathbb{E}[\varepsilon_{n+1}^2|\mathcal{F}_n] = \mathbb{E}[\varepsilon_{n+1}^2] = \sigma^2$, we obtain that

$$\mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = \mathbb{E}\Big[\frac{\sin^2(2\pi(X-\widehat{\theta}_n))}{g^2(X)}(f^2(X-\theta)+\sigma^2)\Big] = \varphi(\widehat{\theta}_n) \tag{4.3}$$

where φ is given by (2.6). Therefore, as f is bounded and g does not vanish on its support $[-1/2, 1/2]$, we deduce from (4.3) that for some constant $M > 0$

$$\sup_{n\geq 0} \mathbb{E}[T_{n+1}^2|\mathcal{F}_n] \leq M \qquad \text{a.s.} \tag{4.4}$$

Furthermore, for all $n \geq 0$, let $V_n = (\widehat{\theta}_n - \theta)^2$. We clearly have

$$\begin{aligned} V_{n+1} &= (\widehat{\theta}_{n+1}-\theta)^2, \\ &= (\pi_K(\widehat{\theta}_n+\gamma_{n+1}T_{n+1})-\theta)^2, \\ &= (\pi_K(\widehat{\theta}_n+\gamma_{n+1}T_{n+1})-\pi_K(\theta))^2 \end{aligned}$$

as we have assumed that θ belongs to K. Since π_K is a Lipschitz function with Lipschitz constant 1, we obtain that

$$\begin{aligned} V_{n+1} &\leq (\widehat{\theta}_n+\gamma_{n+1}T_{n+1}-\theta)^2, \\ &\leq V_n+\gamma_{n+1}^2T_{n+1}^2+2\gamma_{n+1}T_{n+1}(\widehat{\theta}_n-\theta). \end{aligned}$$

Hence, it follows from (4.2) and (4.4) that

$$\begin{aligned} \mathbb{E}[V_{n+1}|\mathcal{F}_n] &\leq V_n+\gamma_{n+1}^2\mathbb{E}[T_{n+1}^2|\mathcal{F}_n]+2\gamma_{n+1}(\widehat{\theta}_n-\theta)\mathbb{E}[T_{n+1}|\mathcal{F}_n], \\ &\leq V_n+\gamma_{n+1}^2M+2\gamma_{n+1}(\widehat{\theta}_n-\theta)\phi(\widehat{\theta}_n) \qquad \text{a.s.} \end{aligned} \tag{4.5}$$

In addition, as $\widehat{\theta}_n \in K$, $|\widehat{\theta}_n| < 1/4$, $|\widehat{\theta}_n - \theta| < 1/2$ which implies that $(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n) < 0$. Then, we deduce from (4.5) together with Robbins-Siegmund Theorem, see Duflo

[9] page 18, that the sequence (V_n) converges a.s. to a finite random variable V and

$$\sum_{n=1}^{\infty} \gamma_{n+1}(\theta - \widehat{\theta}_n)\phi(\widehat{\theta}_n) < +\infty \qquad \text{a.s.} \tag{4.6}$$

Assume by contradiction that $V \neq 0$ a.s. Then, one can find $0 < a < b < 1/2$ such that, for n large enough, $a < |\widehat{\theta}_n - \theta| < b$. However, on this annulus, one can also find some constant $c > 0$ such that $(\theta - \widehat{\theta}_n)\phi(\widehat{\theta}_n) > c$ which, by (4.6), implies that

$$\sum_{n=1}^{\infty} \gamma_n < +\infty$$

This is of course in contradiction with assumption (2.3). Consequently, it follows that $V = 0$ a.s. leading to the almost sure convergence of $\widehat{\theta}_n$ to θ.
It remains to show that $\widehat{\theta}_n + \gamma_{n+1}T_{n+1}$ goes almost surely outside of K a finite number of times. For all $n \geq 1$, denote

$$N_n = \sum_{k=0}^{n-1} \mathrm{I}_{\{|\widehat{\theta}_k + \gamma_{k+1}T_{k+1}| > 1/4\}}.$$

The random sequence (N_n) is nondecreasing. Assume by contradiction that N_n goes to infinity a.s. Then, one can find a subsequence (n_k) such that (N_{n_k}) is increasing. Consequently, for all $n_k > 0$,

$$|\widehat{\theta}_{n_k} + \gamma_{n_k+1}T_{n_k+1}| > \frac{1}{4} \qquad \text{a.s.}$$

which implies that $|\widehat{\theta}_{n_k+1}| = 1/4$ a.s. Hence,

$$\lim_{n_k \to \infty} |\widehat{\theta}_{n_k}| = |\theta| = \frac{1}{4} \qquad \text{a.s.}$$

leading to a contradiction as $|\theta| < 1/4$. Finally, (N_n) converges to a finite limiting value a.s. which completes the proof of Theorem 2.1. □

4.2. **Proof of Theorem 2.2.** We assume without loss of generality that $f_1 > 0$. Our goal is to apply Theorem 2.1 of Kushner and Yin [18] page 330. The only assumption that is not immediate to check is that the sequence (W_n) given by

$$W_n = \frac{(\widehat{\theta}_n - \theta)^2}{\gamma_n}$$

is tight. It follows from (4.5) that for some constant $M > 0$ and for all $n \geq 1$,

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] \leq (1 + \gamma_n)W_n + \gamma_{n+1}M + 2(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n). \tag{4.7}$$

Moreover, we have for all $x \in \mathbb{R}$, $\phi(x) = 2\pi f_1(\theta - x) + f_1(\theta - x)v(x)$ where

$$v(x) = \frac{\sin(2\pi(\theta - x)) - 2\pi(\theta - x)}{(\theta - x)}.$$

By the continuity of the function v, one can find $0 < \varepsilon < 1/2$ such that, if $|x-\theta| < \varepsilon$,

$$\frac{q}{2f_1} < v(x) < 0. \tag{4.8}$$

We also deduce from (4.7) that for all $n \geq 1$,

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] \leq W_n + 2\gamma_n W_n(q - f_1 v(\widehat{\theta}_n)) + \gamma_n M \tag{4.9}$$

with $2q = 1 - 4\pi f_1$ which means that $q < 0$. Moreover, let A_n and B_n be the sets $A_n = \{|\widehat{\theta}_n - \theta| \leq \varepsilon\}$ and

$$B_n = \bigcap_{k=m}^{n} A_k$$

with $1 \leq m \leq n$. Then, it follows from (4.8) that

$$0 < -f_1 v(\widehat{\theta}_n)\mathrm{I}_{B_n} < -\Big(\frac{q}{2}\Big)\mathrm{I}_{B_n}. \tag{4.10}$$

Hence, we deduce from the conjunction of (4.9) and (4.10) that for all $n \geq m$,

$$\begin{aligned} \mathbb{E}[W_{n+1}\mathrm{I}_{B_n}|\mathcal{F}_n] &\leq W_n\mathrm{I}_{B_n} + 2\gamma_n W_n \mathrm{I}_{B_n}\Big(q - \frac{q}{2}\Big) + \gamma_n M, \\ &\leq W_n\mathrm{I}_{B_n}(1 + q\gamma_n) + \gamma_n M. \end{aligned} \tag{4.11}$$

Since $B_{n+1} = B_n \cap A_{n+1}$, $B_{n+1} \subset B_n$, and we obtain by taking the expectation on both sides of (4.11) that for all $n \geq m$,

$$\mathbb{E}[W_{n+1}\mathrm{I}_{B_{n+1}}] \leq (1 + q\gamma_n)\mathbb{E}[W_n\mathrm{I}_{B_n}] + \gamma_n M. \tag{4.12}$$

From now on, denote $\alpha_n = \mathbb{E}[W_n\mathrm{I}_{B_n}]$. We infer from (4.12) that for all $n \geq m$,

$$\alpha_{n+1} \leq \beta_n \alpha_m + M\beta_n \sum_{k=m}^{n} \frac{\gamma_k}{\beta_k} \qquad \text{where} \qquad \beta_n = \prod_{k=m}^{n}(1 + q\gamma_k). \tag{4.13}$$

As $\gamma_n = 1/n$, it follows from straightforward calculations that $\beta_n = O(n^q)$ and

$$\sum_{k=1}^{n} \frac{\gamma_k}{\beta_k} = O(n^{-q}).$$

Consequently, (4.13) immediately leads to

$$\sup_{n \geq m} \alpha_n < +\infty. \tag{4.14}$$

We are now in position to prove the tightness of the sequence (W_n). Indeed, it was already proved in Theorem 2.1 that $\widehat{\theta}_n$ converges to θ a.s. Consequently, if

$$C_n = \bigcup_{k \geq n} \overline{A}_k,$$

then $\mathbb{P}(C_n)$ converges to zero as n tends to infinity. Moreover, for $n \geq m$, $\overline{B}_n \subset C_m$ which implies that as m, n tend to infinity, $\mathbb{P}(\overline{B}_n)$ goes to zero. For all $\xi, K > 0$ and for all $n \geq m$ with m large enough,

$$\begin{aligned} \mathbb{P}(W_n > K) &\leq \mathbb{P}(W_n \mathrm{I}_{B_n} > K/2) + \mathbb{P}(W_n \mathrm{I}_{\overline{B}_n} > K/2), \\ &\leq \frac{2}{K}\mathbb{E}[W_n \mathrm{I}_{B_n}] + \mathbb{P}(\overline{B}_n). \end{aligned} \tag{4.15}$$

We deduce from (4.14) that one can find K depending on ξ such that the first term on the right-hand side of (4.15) is smaller than $\xi/2$. It is also the case for the second term as $\mathbb{P}(\overline{B}_n)$ goes to zero. Finally, for all $\xi > 0$, it exists $K > 0$ such that for m large enough,

$$\sup_{n\geq m} \mathbb{P}(W_n > K) < \xi$$

which implies the tightness of (W_n) and completes the proof of Theorem 2.2. □

4.3. **Proof of Theorem 2.3.** As the number of times that the random variable $\widehat{\theta}_n + \gamma_{n+1}T_{n+1}$ goes outside of K is almost surely finite, the sequence $(\widehat{\theta}_n)$ shares the same almost sure asymptotic properties than the classical Robbins-Monro algorithm. Consequently, we deduce the law of iterated logarithm given by (2.8) from Theorem 1 of [11], see also Hall and Heyde [12] page 240, and the quadratic strong law given by (2.10) from Theorem 3 of [25]. □

5. PROOFS OF THE NONPARAMETRIC RESULTS

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

$$\widehat{h}_n(x) = \frac{1}{n}\sum_{k=1}^{n} W_k(x)Y_k \qquad \text{and} \qquad \widehat{g}_n(x) = \frac{1}{n}\sum_{k=1}^{n} W_k(x).$$

We obtain from (1.1) the decomposition

$$n\widehat{h}_n(x) = M_n(x) + P_n(x) + Q_n(x) + n\widehat{g}_n(x)f(x), \tag{5.1}$$

$$n\widehat{g}_n(x) = N_n(x) + R_n(x) + ng(\theta + x) \tag{5.2}$$

where

$$M_n(x) = \sum_{k=1}^{n} W_k(x)\varepsilon_k, \tag{5.3}$$

$$N_n(x) = \sum_{k=1}^{n} W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}], \tag{5.4}$$

and

$$P_n(x) = \sum_{k=1}^{n} W_k(x)(f(X_k - \widehat{\theta}_{k-1}) - f(x)), \tag{5.5}$$

$$Q_n(x) = \sum_{k=1}^{n} W_k(x)(f(X_k - \theta) - f(X_k - \widehat{\theta}_{k-1})), \tag{5.6}$$

$$R_n(x) = \sum_{k=1}^{n} (\mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] - g(\theta + x)). \tag{5.7}$$

On the one hand,

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

After the change of variables $z = h_n^{-1}(x_n - \widehat{\theta}_{n-1} - x)$, as the density function g is continuous, twice differentiable with bounded derivatives, we infer from the Taylor formula that

$$\begin{aligned}(5.8)\quad \mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] &= \int_{\mathbb{R}} K(z) g(\widehat{\theta}_{n-1} + x + h_n z)\, dz, \\ &= \int_{\mathbb{R}} K(z) \Big(g(\widehat{\theta}_{n-1} + x) + h_n z g'(\widehat{\theta}_{n-1} + x) \\ &\qquad + \frac{h_n^2 z^2}{2} g''(\widehat{\theta}_{n-1} + x + h_n z \xi)\Big) dz, \\ &= g(\widehat{\theta}_{n-1} + x) + \frac{h_n^2}{2} \int_{\mathbb{R}} z^2 K(z) g''(\widehat{\theta}_{n-1} + x + h_n z \xi) dz\end{aligned}$$

where $0 < \xi < 1$. Consequently, for all $n \geq 1$,

$$(5.9)\qquad |\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - g(\widehat{\theta}_{n-1} + x)| \leq M_g \tau^2 h_n^2 \qquad \text{a.s.}$$

where $M_g = \sup_{x \in \mathbb{R}} |g''(x)|$ and

$$\tau^2 = \frac{1}{2} \int_{\mathbb{R}} x^2 K(x) dx.$$

The continuity of g together with the fact that $\widehat{\theta}_n$ converges to θ a.s. leads to

$$(5.10)\qquad \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] = g(\theta + x) \qquad \text{a.s.}$$

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

$$(5.11)\qquad R_n(x) = o(n) \qquad \text{a.s.}$$

On the other hand, $(N_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

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

It follows from the same calculation as in (5.8) that

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

where $0 < \xi < 1$, which leads to

$$|\mathbb{E}[W_n^2(x)|\mathcal{F}_{n-1}] - \frac{\nu^2}{h_n} g(\widehat{\theta}_{n-1} + x)| \leq M_g \mu^2 h_n \qquad \text{a.s.} \tag{5.12}$$

with

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

Hence, since

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

we deduce from (5.9) and (5.12) together with Toeplitz lemma and the almost sure convergence of $g(\widehat{\theta}_n + x)$ to $g(\theta + x)$ that

$$\lim_{n\to\infty} \frac{< N(x) >_n}{n^{1+\alpha}} = \frac{\nu^2 g(\theta + x)}{1+\alpha} \qquad \text{a.s.} \tag{5.13}$$

Consequently, we obtain from the strong law of large numbers for martingales given e.g. by Theorem 1.3.15 of [9] that for any $\gamma > 0$, $(N_n(x))^2 = o(n^{1+\alpha}(\log n)^{1+\gamma})$ a.s. which ensures that, for all $x \in \mathbb{R}$,

$$N_n(x) = o(n) \qquad \text{a.s.} \tag{5.14}$$

Therefore, it follows from (5.2), (5.11) and (5.14) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \widehat{g}_n(x) = g(\theta + x) \qquad \text{a.s.} \tag{5.15}$$

Moreover, 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 $x \in \mathbb{R}$,

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

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

$$|f(X_n - \widehat{\theta}_{n-1}) - f(x)| \leq C_f |X_n - \widehat{\theta}_{n-1} - x|.$$

Consequently, we obtain from (5.5) that for all $x \in \mathbb{R}$

$$\begin{aligned} |P_n(x)| &\leq C_f \sum_{k=1}^n W_k(x) |X_k - \widehat{\theta}_{k-1} - x|, \\ &\leq A C_f \sum_{k=1}^n h_k W_k(x). \end{aligned} \tag{5.16}$$

Hence, it follows from convergence (5.10) together with (5.14) and (5.16) that for all $x \in \mathbb{R}$

$$P_n(x) = o(n) \qquad \text{a.s.} \tag{5.17}$$

Furthermore, we obtain from (5.6) that for all $x \in \mathbb{R}$

$$|Q_n(x)| \leq C_f \sum_{k=1}^{n} W_k(x)|\widehat{\theta}_{k-1} - \theta|. \tag{5.18}$$

Then, it follows from the Cauchy-Schwarz inequality that

$$Q_n^2(x) \leq C_f^2 \sum_{k=1}^{n} W_k^2(x) \sum_{k=1}^{n} |\widehat{\theta}_{k-1} - \theta|^2. \tag{5.19}$$

We can split the first sum at the right-hand side of (5.19) into two terms,

$$\sum_{k=1}^{n} W_k^2(x) = I_n(x) + J_n(x)$$

where

$$\begin{aligned} I_n(x) &= \sum_{k=1}^{n} W_k^2(x) - \mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}], \\ J_n(x) &= \sum_{k=1}^{n} \mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}]. \end{aligned}$$

Following the same lines as in the proof of (5.14), it is not hard to see that

$$I_n(x) = o(n^{1+\alpha}) \qquad \text{a.s.}$$

We also deduce from convergence (5.13) that

$$J_n(x) = O(n^{1+\alpha}) \qquad \text{a.s.}$$

Consequently, we obtain that for all $x \in \mathbb{R}$

$$\sum_{k=1}^{n} W_k^2(x) = O(n^{1+\alpha}) \qquad \text{a.s.} \tag{5.20}$$

Therefore, we infer from the quadratic strong law given by (2.10) together with (5.19) and (5.20) that $Q_n^2(x) = O(n^{1+\alpha}\log n)$ a.s. which implies that for all $x \in \mathbb{R}$

$$Q_n(x) = o(n) \qquad \text{a.s.} \tag{5.21}$$

It now remains to study the asymptotic behavior of $M_n(x)$ given by (5.3). As (X_n) and (ε_n) are two independent sequences of independent and identically distributed random variables, $(M_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

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

Then, it follows from convergence (5.13) that

$$\lim_{n\to\infty} \frac{< M(x) >_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2 g(\theta+x)}{1+\alpha} \qquad \text{a.s.} \tag{5.22}$$

Consequently, we obtain from the strong law of large numbers for martingales that for any $\gamma > 0$, $(M_n(x))^2 = o(n^{1+\alpha}(\log n)^{1+\gamma})$ a.s. which leads to

$$M_n(x) = o(n) \qquad \text{a.s.} \tag{5.23}$$

Therefore, we deduce from (5.1) and (5.15) together with the conjunction of (5.17), (5.21) and (5.23) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \widehat{h}_n(x) = f(x)g(\theta+x) \qquad \text{a.s.} \tag{5.24}$$

Finally, we can conclude from the identity

$$\widehat{f}_n(x) = \frac{\widehat{h}_n(x) + \widehat{h}_n(-x)}{\widehat{g}_n(x) + \widehat{g}_n(-x)} \tag{5.25}$$

and the parity of the function f that, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \qquad \text{a.s.} \tag{5.26}$$

5.2. **Proof of Theorem 3.2.** We shall now proceed to the proof of the asymptotic normality of $\widehat{f}_n$. It follows from (5.1), (5.2) and (5.25) that for all $x \in \mathbb{R}$

$$\widehat{f}_n(x) - f(x) \;=\; \frac{\mathcal{M}_n(x) + \mathcal{P}_n(x) + \mathcal{Q}_n(x)}{n\mathcal{G}_n(x)} \tag{5.27}$$

where $\mathcal{G}_n(x) = \widehat{g}_n(x) + \widehat{g}_n(-x)$ and

$$\begin{aligned}
\mathcal{M}_n(x) &= M_n(x) + M_n(-x), \\
\mathcal{P}_n(x) &= P_n(x) + P_n(-x), \\
\mathcal{Q}_n(x) &= Q_n(x) + Q_n(-x),
\end{aligned}$$

with $M_n(x)$, $P_n(x)$ and $Q_n(x)$ given by (5.3), (5.5) and (5.6), respectively. We already saw from (5.15) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathcal{G}_n(x) = g(\theta+x) + g(\theta-x) \qquad \text{a.s.} \tag{5.28}$$

In order to establish the asymptotic normality, it is now necessary to be more precise in the almost sure rates of convergence given in (5.17) and (5.21). It follows from (5.16) that for all $x \in \mathbb{R}$

$$|P_n(x)| \le AC_f(L_n(x) + \Lambda_n(x)) \tag{5.29}$$

where

$$\begin{aligned}
L_n(x) &= \sum_{k=1}^{n} h_k(W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]), \\
\Lambda_n(x) &= \sum_{k=1}^{n} h_k \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}].
\end{aligned}$$

On the one hand, we infer from (5.9) that

$$\Lambda_n(x) = O\Big(\sum_{k=1}^{n} h_k\Big) = O(n^{1-\alpha}) \qquad \text{a.s.} \tag{5.30}$$

On the other hand, $(L_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

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

We deduce from (5.9) and (5.12) together with Toeplitz lemma that

$$\lim_{n\to\infty} \frac{< L(x) >_n}{n^{1-\alpha}} = \frac{\nu^2 g(\theta + x)}{1-\alpha} \qquad \text{a.s.} \tag{5.31}$$

Consequently, we obtain from the strong law of large numbers for martingales that for any $\gamma > 0$, $(L_n(x))^2 = o(n^{1-\alpha}(\log n)^{1+\gamma})$ a.s. which clearly implies that $(L_n(x))^2 = o(n^{1+\alpha})$ a.s. Therefore, we find from (5.29) and (5.30) that, as soon as $\alpha > 1/3$,

$$(P_n(x))^2 = O(n^{2-2\alpha}) + o(n^{1+\alpha}) = o(n^{1+\alpha}) \qquad \text{a.s.}$$

which immediately leads to

$$(\mathcal{P}_n(x))^2 = o(n^{1+\alpha}) \qquad \text{a.s.} \tag{5.32}$$

Proceeding as in the proof of (5.32), we obtain from (5.18) that for all $x \in \mathbb{R}$

$$|Q_n(x)| \le C_f (S_n(x) + \Sigma_n(x)) \tag{5.33}$$

where

$$\begin{aligned} S_n(x) &= \sum_{k=1}^{n} \ell_k (W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]), \\ \Sigma_n(x) &= \sum_{k=1}^{n} \ell_k \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] \end{aligned}$$

with $\ell_n = |\widehat{\theta}_{n-1} - \theta|$. We deduce from (5.9) together with the Cauchy-Schwarz inequality and the quadratic strong law given by (2.10) that

$$\Sigma_n(x) = O\Big(\sum_{k=1}^{n} \ell_k\Big) = O(\sqrt{n \log n}) \qquad \text{a.s.} \tag{5.34}$$

In addition, it follows from (5.12) that $(S_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation satisfying

$$< S(x) >_n = O(n^{\alpha} \log n) \qquad \text{a.s.}$$

Consequently, we obtain from the strong law of large numbers for martingales that for any $\gamma > 0$, $(S_n(x))^2 = o(n^{\alpha}(\log n)^{2+\gamma})$ a.s. so $(S_n(x))^2 = o(n^{1+\alpha})$ a.s. Hence, we find from (5.33) and (5.34) that

$$(Q_n(x))^2 = O(n \log n) + o(n^{1+\alpha}) = o(n^{1+\alpha}) \qquad \text{a.s.}$$

which obviously implies

$$(\mathcal{Q}_n(x))^2 = o(n^{1+\alpha}) \qquad \text{a.s.} \tag{5.35}$$

It remains to establish the asymptotic behavior of the dominating term $\mathcal{M}_n(x)$. We already saw that $(M_n(x))$ is a square integrable martingale difference sequence. Consequently, $(\mathcal{M}_n(x))$ is also a square integrable martingale difference sequence with predictable quadratic variation given by

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

Hence, it is necessary to evaluate the cross-term $\mathbb{E}[W_n(x)W_n(-x)|\mathcal{F}_{n-1}]$. It follows from the same calculation as in (5.8) that

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

with $0 < \xi < 1$. Consequently, we obtain that

$$\left|\mathbb{E}[W_n(x)W_n(-x)|\mathcal{F}_{n-1}] - \frac{1}{h_n} g(\widehat{\theta}_{n-1}+x)I_n(x) - g'(\widehat{\theta}_{n-1}+x)J_n(x)\right| \leq M_g H_n(x) h_n \quad \text{a.s.}$$

where

$$\begin{aligned}
I_n(x) &= \int_{\mathbb{R}} K(z)K(z+2h_n^{-1}x)dz, \\
J_n(x) &= \int_{\mathbb{R}} zK(z)K(z+2h_n^{-1}x)dz, \\
H_n(x) &= \int_{\mathbb{R}} z^2 K(z)K(z+2h_n^{-1}x)dz.
\end{aligned}$$

However, as the kernel K is compactly supported, we have for all $x \in \mathbb{R}$ with $x \neq 0$,

$$\lim_{n\to\infty} K(z+2h_n^{-1}x) = 0.$$

Then, we deduce from Lebesgue dominated convergence theorem that all the three integrals $I_n(x)$, $J_n(x)$, and $H_n(x)$ tend to zero as n goes to infinity, which implies that for all $x \in \mathbb{R}$ with $x \neq 0$,

$$\sum_{k=1}^{n} \mathbb{E}[W_k(x)W_k(-x)|\mathcal{F}_{k-1}] = o\Big(\sum_{k=1}^{n} h_k^{-1}\Big) = o(n^{1+\alpha}) \qquad \text{a.s.} \tag{5.36}$$

Therefore, we find from (5.22) together with (5.36) that for all $x \in \mathbb{R}$ with $x \neq 0$,

$$\lim_{n\to\infty} \frac{<\mathcal{M}(x)>_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2}{1+\alpha}(g(\theta+x)+g(\theta-x)) \qquad \text{a.s.} \tag{5.37}$$

If $x = 0$, it immediately follows from (5.22)

$$\lim_{n\to\infty} \frac{<\mathcal{M}(0)>_n}{n^{1+\alpha}} = \frac{4\sigma^2\nu^2 g(\theta)}{1+\alpha} \qquad \text{a.s.} \tag{5.38}$$

Furthermore, it is not hard to see that Lindeberg condition is satisfied. As a matter of fact, we have assumed that the sequence (ε_n) has a finite moment of order $a > 2$. If we denote $\Delta\mathcal{M}_n(x) = \mathcal{M}_n(x) - \mathcal{M}_{n-1}(x)$, we have

$$\mathbb{E}[|\Delta\mathcal{M}_n(x)|^a|\mathcal{F}_{n-1}] = \mathbb{E}[|\varepsilon_n|^a]\mathbb{E}[|W_n(x) - W_n(-x)|^a|\mathcal{F}_{n-1}],$$

which implies that

$$\mathbb{E}[|\Delta\mathcal{M}_n(x)|^a|\mathcal{F}_{n-1}] \leq 2^{a-1}\mathbb{E}[|\varepsilon_n|^a]\mathbb{E}[W_n^a(x) + W_n^a(-x)|\mathcal{F}_{n-1}].$$

However, it follows from the same calculation as in (5.8) that

$$\sum_{k=1}^{n} \mathbb{E}[W_k^a(x)|\mathcal{F}_{k-1}] = O\Big(\sum_{k=1}^{n} h_k^{1-a}\Big) = O(n^{1+\alpha(a-1)}) \qquad \text{a.s.} \tag{5.39}$$

In addition, for all $\varepsilon > 0$,

$$\frac{1}{n^{1+\alpha}} \sum_{k=1}^{n} \mathbb{E}[(\Delta\mathcal{M}_k(x))^2 \mathrm{I}_{|\Delta\mathcal{M}_k(x)|\geq\varepsilon\sqrt{n^{1+\alpha}}}|\mathcal{F}_{k-1}] \leq \frac{1}{\varepsilon^{a-2}n^b} \sum_{k=1}^{n} \mathbb{E}[|\Delta\mathcal{M}_k(x)|^a|\mathcal{F}_{k-1}]$$

where $b = a(1+\alpha)/2$. Consequently, it follows from (5.39) that for all $\varepsilon > 0$,

$$\frac{1}{n^{1+\alpha}} \sum_{k=1}^{n} \mathbb{E}[(\Delta\mathcal{M}_k(x))^2 \mathrm{I}_{|\Delta\mathcal{M}_k(x)|\geq\varepsilon\sqrt{n^{1+\alpha}}}|\mathcal{F}_{k-1}] = O(n^c) \qquad \text{a.s.}$$

where $c = (2-a)(1-\alpha)/2$. As $c < 0$, Lindeberg condition is clearly satisfied. We can conclude from the central limit theorem for martingales given e.g. by Corollary 2.1.10 of [9] that for all $x \in \mathbb{R}$ with $x \neq 0$,

$$\frac{\mathcal{M}_n(x)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma^2\nu^2}{1+\alpha}(g(\theta+x) + g(\theta-x))\Big) \tag{5.40}$$

while, for $x = 0$,

$$\frac{\mathcal{M}_n(0)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{4\sigma^2\nu^2}{1+\alpha}g(\theta)\Big). \tag{5.41}$$

Finally, it follows from (5.27) and (5.28) together with (5.32), (5.35), (5.40), (5.41) and Slutsky lemma that, for all $x \in \mathbb{R}$ with $x \neq 0$,

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

while, for $x = 0$,

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

which completes the proof of Theorem 3.2. □

References

[1] Bercu, B., and Portier, B. Kernel density estimation and goodness-of-fit test in adaptive tracking. *SIAM J. Control Optim. 47*, 5 (2008), 2440–2457.

[2] Bigot, J., and Gadat, S. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist. 38*, 4 (2010), 2422–2464.

[3] Castillo, I., and Loubes, J.-M. Estimation of the distribution of random shifts deformation. *Math. Methods Statist. 18*, 1 (2009), 21–42.

[4] Chen, H. F., Lei, G., and Gao, A. J. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stoch. Process. Appl. 27*, 2 (1988), 217–231.

[5] Choi, E., Hall, P., and Rousson, V. Data sharpening methods for bias reduction in nonparametric regression. *Ann. Statist. 28*, 5 (2000), 1339–1355.

[6] Dalalyan, A. S., Golubev, G. K., and Tsybakov, A. B. Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist. 34*, 1 (2006), 169–201.

[7] Dette, H., Neumeyer, N., and Pilz, K. F. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli 12*, 3 (2006), 469–490.

[8] Devroye, L., and Lugosi, G. *Combinatorial methods in density estimation.* Springer Series in Statistics. Springer-Verlag, New York, 2001.

[9] Duflo, M. *Random iterative models*, vol. 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997.

[10] Gamboa, F., Loubes, J.-M., and Maza, E. Semi-parametric estimation of shifts. *Electron. J. Stat. 1* (2007), 616–640.

[11] Gapoškin, V. F., and Krasulina, T. P. The law of the iterated logarithm in stochastic approximation processes. *Teor. Verojatnost. i Primenen. 19* (1974), 879–886.

[12] Hall, P., and Heyde, C. C. *Martingale limit theory and its application.* Academic Press Inc. New York, 1980.

[13] Hall, P., and Huang, L.-S. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist. 29*, 3 (2001), 624–647.

[14] Härdle, W. A law of the iterated logarithm for nonparametric regression function estimators. *Ann. Statist. 12*, 2 (1984), 624–635.

[15] Härdle, W., Janssen, P., and Serfling, R. Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist. 16*, 4 (1988), 1428–1449.

[16] Härdle, W., and Tsybakov, A. B. Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist. 16*, 1 (1988), 120–135.

[17] Kushner, H. J., and Clark, D. S. *Stochastic approximation methods for constrained and unconstrained systems*, vol. 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1978.

[18] Kushner, H. J., and Yin, G. G. *Stochastic approximation and recursive algorithms and applications*, second ed., vol. 35 of *Applications of Mathematics*. Springer-Verlag, New York, 2003.

[19] Lelong, J. Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett. 78*, 16 (2008), 2632–2636.

[20] Mokkadem, A., and Pelletier, M. A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist. 35*, 4 (2007), 1749–1772.

[21] Nadaraja, È. A. On a regression estimate. *Teor. Verojatnost. i Primenen. 9* (1964), 157–159.

[22] Nadaraya, È. A. *Nonparametric estimation of probability densities and regression curves*, vol. 20 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1989.

[23] Noda, K. Estimation of a regression function by the Parzen kernel-type density estimators. *Ann. Inst. Statist. Math. 28*, 2 (1976), 221–234.

[24] Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Statist. 33* (1962), 1065–1076.

[25] Pelletier, M. On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic Process. Appl. 78*, 2 (1998), 217–244.

[26] Robbins, H., and Monro, S. A stochastic approximation method. *Ann. Math. Statistics 22* (1951), 400–407.

[27] Robbins, H., and Siegmund, D. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics*. Academic Press, New York, 1971, pp. 233–257.

[28] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist. 27* (1956), 832–837.

[29] Schuster, E. F. Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist. 43* (1972), 84–88.

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

[31] Stone, C. J. Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist. 3* (1975), 267–284.

[32] Tsybakov, A. B. *Introduction à l'estimation non-paramétrique*, vol. 41 of *Mathématiques & Applications*. Springer-Verlag, Berlin, 2004.

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

E-mail address: bernard.bercu@math.u-bordeaux1.fr
E-mail address: philippe.fraysse@math.u-bordeaux1.fr

Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, and INRIA Bordeaux, team ALEA, 351 cours de la libération, 33405 Talence cedex, France.