The Annals of Statistics
2012, Vol. 40, No. 2, 666–693
DOI: 10.1214/12-AOS969

A ROBBINS–MONRO PROCEDURE FOR ESTIMATION IN SEMIPARAMETRIC REGRESSION MODELS

BY BERNARD BERCU AND PHILIPPE FRAYSSE

Université de Bordeaux

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 very efficient and easy to handle Robbins–Monro procedure. 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 to estimate 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 into 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. Finally, we illustrate our semiparametric estimation procedure on simulated and real data.

1. Introduction. A wide range of real-life phenomena occur periodically. One can think about meteorology with daily or annual cycles of temperature [19], astronomy with the famous 11-year cycles of solar geomagnetic activity [23], medicine with human circadian rhythms [39] or ECG signals [36], econometry [2, 26], communication [11], etc. Statistical analysis of periodic data is of great interest in order to design suitable models for those cyclic phenomena. An important literature is available on the so-called periodic shape-invariant model introduced by Lawton, Sylvestre and Maggio [24]. Theoretical advances on shape-invariant models together with statistical applications may be found in [17, 19–21, 38, 39]. A periodic shape-invariant model is a semiparametric regression model with an unknown periodic shape function. It is given, for all $n \geq 0$, by

$$Y_n = h(X_n) + \varepsilon_n, \tag{1.1}$$

where the inputs (X_n) are known observation times, the output (Y_n) are the observations, and (ε_n) are unknown random errors. The function h is periodic and takes the form

$$h(x) = m + \sum_{k=1}^{p} a_k f(x - \theta_k),$$

Received February 2011; revised January 2012.
MSC2010 subject classifications. Primary 62G05; secondary 62G20.
Key words and phrases. Semiparametric estimation, estimation of a shift, estimation of a regression function, asymptotic properties.

where f represents the unknown characteristic shape function, m is the overall mean, while $\theta = (\theta_1, \dots, \theta_p)$ and $a = (a_1, \dots, a_p)$ are unknown shift and scale parameters.

In this paper, we shall focus our attention on the particular case $p = 1$, $m = 0$ and $a = 1$ by studying the semiparametric regression model given, for all $n \geq 0$, by

$$Y_n = f(X_n - \theta) + \varepsilon_n, \tag{1.2}$$

where (X_n) and (ε_n) are two independent sequences of independent and identically distributed random variables. We are dealing with random observation times in contrast with the previous literature where (X_n) are assumed to be known and equidistributed over a given interval. We are interested in the parametric estimation of the shift parameter θ together with the nonparametric estimation of the shape function f. However, one has to keep in mind that our main interest lies in the estimation of the parameter θ. We are also motivated by a statistical application on the detection of atrial fibrillation using ECG analysis [6, 36].

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 [7] in the Gaussian white-noise case. First, a penalized maximum likelihood estimator of θ is proposed in [7] with an appropriately chosen penalty based on a Fourier series approximation of the function f. Second, the asymptotic behavior of the mean squared risk of this estimator is investigated. One can observe that our estimator is much easier 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 [30, 33] density estimator of θ. The construction of this estimate also relies on the penalized maximum likelihood estimator of θ given in [7]. Furthermore, in the case where one observes several Gaussian functions differing from each other by a translation parameter, Gamboa, Loubes and Maza [11] 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. In a more general framework, Vimond [38] makes use of a truncated Fourier approximation of f in order to evaluate the profile log-likelihood score function associated with the shift and scale parameters. This two-step strategy requires, as in [11], the estimation of the Fourier coefficients of f. However, it performs pretty well as it leads to consistent and asymptotically efficient estimators of the shift and scale parameters. Our alternative approach to estimate θ is associated to a stochastic recursive algorithm similar to that of Robbins–Monro [32, 33].

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

$$\widehat{\theta}_{n+1}=\widehat{\theta}_n+\gamma_n T_{n+1}, \tag{1.3}$$

where (γ_n) is a positive sequence of real numbers decreasing toward 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, 22] 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 [13, 27] and [31]. A randomly truncated version of the Robbins–Monro algorithm is also given in [4, 25].

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, 37] for two 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 [28, 40] without the shift θ was established by Noda [29]; see also Härdle et al. [15, 16] for the law of iterated logarithm and the uniform strong law. A nice extension of the previous results may be found in [18]. The asymptotic normality of the Nadaraya–Watson estimator was proved by Schuster [34]. 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 [14] 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

$$\widehat{f}_n(x)=\frac{\sum_{k=1}^n W_k(x)Y_k}{\sum_{k=1}^n W_k(x)} \tag{1.4}$$

with

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

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 analyze 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 $\widehat{f}_n$. Section 4 contains some numerical experiments on simulated and real ECG data, illustrating the performances of our semiparametric estimation procedure. The proofs of the parametric results are given in Section 5, while those concerning the nonparametric results are postponed to Section 6.

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 (1.2). We assume that (ε_n) is a sequence of independent and identically distributed random variables with zero mean and unknown positive variance σ^2. Moreover, it is necessary to make several hypotheses similar to those of [7].

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

($\mathcal{H}_2$) The shape 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

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

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

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

Consequently, the symmetry of f leads to

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

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 [32, 33]. 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

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

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

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

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* $\widehat{\theta}_n + \operatorname{sign}(f_1)\gamma_{n+1}T_{n+1}$ *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

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

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

$$\xi^2(\theta) = \frac{\varphi(\theta)}{4\pi|f_1| - 1}. \tag{2.7}$$

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*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2(\theta)). \tag{2.8}$$

REMARK 2.1. We clearly have $\phi'(t) = -2\pi f_1 \cos(2\pi(\theta - t))$. Consequently, the value $\phi'(\theta) = -2\pi f_1$ does not depend upon the unknown parameter θ. On the one hand, if the first Fourier coefficient f_1 of f is known, it is possible to provide, via a slight modification of (2.4), an asymptotically efficient estimator $\widehat{\theta}_n$ of θ. More precisely, it is only necessary to replace $\gamma_n = 1/n$ in (2.4) by $\gamma_n = \gamma/n$ where

$$\gamma = \frac{1}{2\pi|f_1|}.$$

Then, we deduce from the original work of Fabian [10] that $\widehat{\theta}_n$ is an asymptotically efficient estimator of θ with

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\varphi(\theta)}{4\pi^2 f_1^2}\right). \tag{2.9}$$

On the other hand, if f_1 is unknown, it is also possible to provide by the same procedure an asymptotically efficient estimator $\widehat{\theta}_n$ of θ replacing f_1 by its natural estimate

$$\widehat{f}_{1,n} = \frac{1}{n}\sum_{k=1}^{n} \frac{Y_k \cos(2\pi(X_k - \widehat{\theta}_{k-1}))}{g(X_k)}.$$

REMARK 2.2. In the particular case where $4\pi|f_1| = 1$, it is also possible to show [9] 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*

$$\begin{aligned}\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)^{1/2}(\widehat{\theta}_n - \theta) &= -\liminf_{n\to\infty}\left(\frac{n}{2\log\log n}\right)^{1/2}(\widehat{\theta}_n - \theta)\\ &= \xi(\theta) \qquad a.s.\end{aligned} \tag{2.10}$$

In particular,

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)(\widehat{\theta}_n-\theta)^2=\xi^2(\theta)\qquad a.s. \tag{2.11}$$

In addition, we also have the quadratic strong law

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}(\widehat{\theta}_k-\theta)^2=\xi^2(\theta)\qquad a.s. \tag{2.12}$$

PROOF. The proofs are given in Section 5. □

REMARK 2.3. It is also possible to get rid of the symmetry assumption on f. However, it requires the knowledge of the first Fourier coefficients of f:

$$f_1=\int_{-1/2}^{1/2}\cos(2\pi x)f(x)\,dx\quad\text{and}\quad g_1=\int_{-1/2}^{1/2}\sin(2\pi x)f(x)\,dx.$$

On the one hand, it is necessary to assume that $f_1\neq 0$ or $g_1\neq 0$, and to replace the first auxiliary function ϕ defined in (2.1) by

$$\begin{aligned}\Phi(t)&=f_1\mathbb{E}\left[\frac{\sin(2\pi(X-t))}{g(X)}f(X-\theta)\right]-g_1\mathbb{E}\left[\frac{\cos(2\pi(X-t))}{g(X)}f(X-\theta)\right]\\&=(f_1^2+g_1^2)\sin\bigl(2\pi(\theta-t)\bigr).\end{aligned}$$

Then, Theorem 2.1 is true for the projected Robbins–Monro algorithm

$$\widehat{\theta}_{n+1}=\pi_K(\widehat{\theta}_n+\gamma_{n+1}T_{n+1}),$$

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

$$T_{n+1}=\frac{f_1\sin(2\pi(X_{n+1}-\widehat{\theta}_n))}{g(X_{n+1})}Y_{n+1}-\frac{g_1\cos(2\pi(X_{n+1}-\widehat{\theta}_n))}{g(X_{n+1})}Y_{n+1}.$$

On the other hand, we also have to replace the second function φ defined in (2.6) by

$$\begin{aligned}\Psi(t)&=\mathbb{E}\left[\frac{(f_1\sin(2\pi(X-t))-g_1\cos(2\pi(X-t)))^2}{g^2(X)}\bigl(f^2(X-\theta)+\sigma^2\bigr)\right]\\&=\int_{-1/2}^{1/2}\frac{(f_1\sin(2\pi(x-t))-g_1\cos(2\pi(x-t)))^2}{g(x)}\bigl(f^2(x-\theta)+\sigma^2\bigr)\,dx.\end{aligned}$$

Then, as soon as $4\pi(f_1^2+g_1^2)>1$, Theorems 2.2 and 2.3 hold with

$$\xi^2(\theta)=\frac{\Psi(\theta)}{4\pi(f_1^2+g_1^2)-1}.$$

In the rest of the paper, we shall not go in that direction as our strategy is to make very few assumptions on the Fourier coefficients of f.

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 [35] for the estimation of a symmetric probability density function replacing the estimator (1.4) 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

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

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

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

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 $|\theta| < 1/4$ and the sequence (ε_n) has a finite moment of order > 2. Then, for any $x \in \mathbb{R}$ such that $|x| \leq 1/2$,*

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

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 $|\theta| < 1/4$ and 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 $x \in \mathbb{R}$ such that $|x| \leq 1/2$ with $x \neq 0$, the pointwise asymptotic normality*

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

In addition, for $x = 0$,

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

PROOF. The proofs are given in Section 6. □

4. Simulations. The goal of this section is to illustrate via some numerical experiments the good performances of our estimation strategy. The first subsection is devoted to simulated data created according to the model (1.2) while the second one deals with real ECG data taken from the MIT-BIH database. Our aim is to propose an efficient and easy to handle procedure in order to detect atrial fibrillation using ECG records. An interesting study on ECG analysis in order to detect cardiac arrhythmia may also be found in [36].

4.1. *Simulated data.* Consider the semiparametric regression model

$$Y_n = f(X_n - \theta) + \varepsilon_n,$$

where $\theta = 1/10$ and the periodic shape function f is given, for $p \geq 1$ and for all $x \in \mathbb{R}$, by

$$f(x) = \sum_{k=1}^{p} \cos(2k\pi x)$$

with $p = 8$. We have chosen (X_n) and (ε_n) as two independent sequences of independent random variables with $\mathcal{U}[-1/2, 1/2]$ and $\mathcal{N}(0, 1)$ distributions, respectively. The simulated data are given in the left-hand side of Figure 1.

For the estimation of the shift parameter θ, we implement our Robbins–Monro procedure with $n = 1000$ iterations. We obtain the estimate $\widehat{\theta}_n = 0.1014$ which shows the good asymptotic behavior of the estimator $\widehat{\theta}_n$ comparing to the true value $\theta = 1/10$. Moreover, using convergence (2.8), one can obtain confidence intervals for the shift parameter. More precisely, they are given, for all $n \geq 1$, by

$$I_n(\theta) = \left[\widehat{\theta}_n - q_\beta \frac{\widehat{\xi}_n(\theta)}{\sqrt{n}}, \widehat{\theta}_n + q_\beta \frac{\widehat{\xi}_n(\theta)}{\sqrt{n}}\right],$$

where q_β stands for the quantile of order $0 < \beta < 1$ of the $\mathcal{N}(0, 1)$ distribution and $\widehat{\xi}_n(\theta)$ is a consistent estimator of $\xi(\theta)$ given by (2.7). In our particular case, it

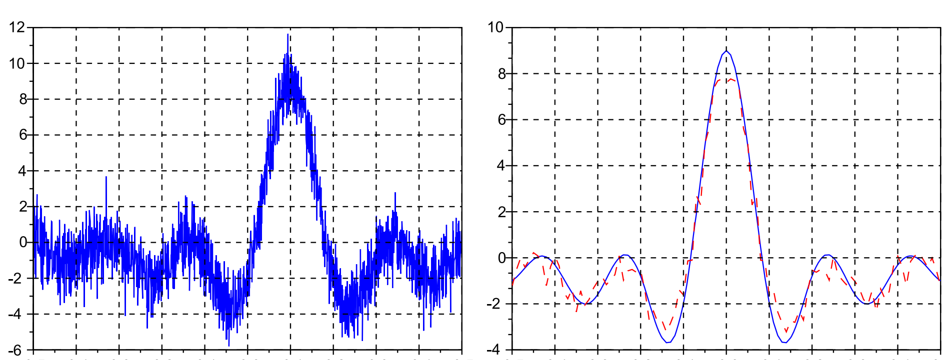

FIG. 1. *Simulated data and almost sure convergence.*

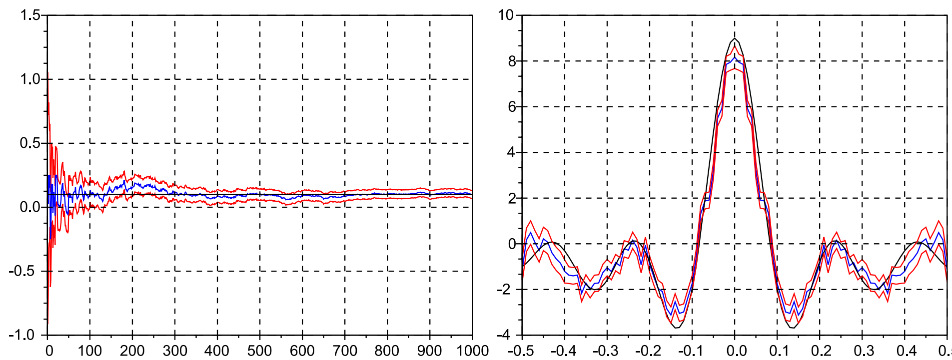

FIG. 2. *Confidence intervals for θ and f.*

is not necessary to estimate $\xi^2(\theta)$ since via straightforward calculations, $f_1 = 1/2$ and

$$\xi^2(\theta) = \frac{7}{8(2\pi - 1)}.$$

Moreover, for $n = 1000$ and for a risk $\beta = 5\%$, the confidence interval is precisely $I_n(\theta) = [0.0762; 0.1266]$. The length of $I_n(\theta)$ is 0.0504, which is rather small, so our Robbins–Monro procedure performs pretty well. All confidence intervals $I_n(\theta)$, for $n = 1, \dots, 1000$, are drawn in red in the left-hand side of Figure 2.

For the estimation of the regression function f, we make use of the uniform kernel K on the interval $[-1, 1]$, and the bandwidth $h_n = 1/n^\alpha$ with $\alpha = 9/10$. In addition, it follows from convergences (3.3) and (3.4) that for $n = 1000$ and for all $x \in [-1/2, 1/2]$, a confidence interval for $f(x)$ is given by

$$J_n(x) = \left[\widehat{f}_n(x) - q_\beta \frac{\widehat{v}_n(x, \widehat{\theta}_n)}{\sqrt{n h_n}}, \widehat{f}_n(x) + q_\beta \frac{\widehat{v}_n(x, \widehat{\theta}_n)}{\sqrt{n h_n}} \right],$$

where q_β stands for the quantile of order $0 < \beta < 1$ of the $\mathcal{N}(0, 1)$ distribution and $\widehat{v}_n^2(x, \widehat{\theta}_n)$ is a consistent estimator of the asymptotic variance $v^2(x, \theta)$ in Theorem 3.2. In our particular case, $\nu^2 = 1/2$ and

$$v^2(x, \theta) = \begin{cases} 5/19, & \text{if } -1/2 \le x < -2/ \text{ or } 2/5 < x \le 1/2, \\ 5/38, & \text{if } -2/5 \le x \le 2/5 \text{ and } x \ne 0, \\ 5/19, & \text{if } x = 0. \end{cases}$$

All confidence intervals $J_n(x)$, for all $x \in [-1/2, 1/2]$, are drawn in red in the right-hand side of Figure 2. On the one hand, the simulations show that the largest length of the confidence intervals $J_n(x)$ is for $x = -0.47$ and $x = 0.47$ and the length is precisely equal to 1.0066. On the other hand, the smallest length of the confidence intervals $J_n(x)$ is for $x = -0.04$ and $x = 0.04$ and is equal to 0.7118. The fact that there are two values of x for the largest and the smallest length of confidence intervals is due to the symmetry of the estimator $\widehat{f}_n$. Then, one can observe on this first set of simulated data that the Robbins–Monro estimator $\widehat{\theta}_n$ of θ as well as the Nadaraya–Watson estimator $\widehat{f}_n$ of f perform pretty well.

Our second experiment deals with 30 curves according to the model

$$Y_n = f(X_n - \theta) + \varepsilon_n,$$

where $\theta = -1/5$ for the first 10 curves and $\theta' = 1/10$ for the last 20 curves. The periodic shape function f is given, for all $x \in [-1/2, 1/2]$, by

$$f(x) = \cos(2\pi x) + \sin(2\pi x) + \cos(2\pi x)\sin(2\pi x).$$

Our goal is to propose a statistical procedure in order to detect a lag between the first 10 curves with $\theta = -1/5$ and the last 20 curves with $\theta' = 1/10$. In other words, we want to observe whether or not the value $\Delta = \theta' - \theta$ is far away from zero. We have chosen (X_n) and (ε_n) as two independent sequences of independent and Gaussian random variables with uniform distribution on $[-1/2, 1/2]$ and $\mathcal{N}(0, 1/5)$ distribution, respectively. Each curve is drawn with $n = 200$ points. The different curves are given in Figure 3.

On the one hand, we estimate the first value $\theta = -1/5$ from the first 10 curves. We implement our Robbins–Monro procedure with $n = 200$ iterations for the first estimate $\widehat{\theta}_n$ of θ evaluated on the first curve, then with $n = 400$ iterations for the second estimate $\widehat{\theta}_n$ of θ evaluated on the two first curves, and so on, until the calculation of the last estimate $\widehat{\theta}_n$ of θ with $n = 2000$. Therefore, we obtain -0.1950 for the arithmetic mean of the first 10 estimates $\widehat{\theta}_n$ of θ. We continue with the same procedure on all the set of curves. The value of the eleven estimates with $n = 2200$ is 0.0986. This value is significantly different from the first 10 estimates. It corresponds to the first curve simulated with $\theta' = 1/10$. Furthermore, we obtain 0.0998 for the arithmetic mean of the last 20 estimates $\widehat{\theta}_n$ of θ'. Finally, our statistical procedure allows us to detect a change of parameterization from the value $\theta = -1/5$

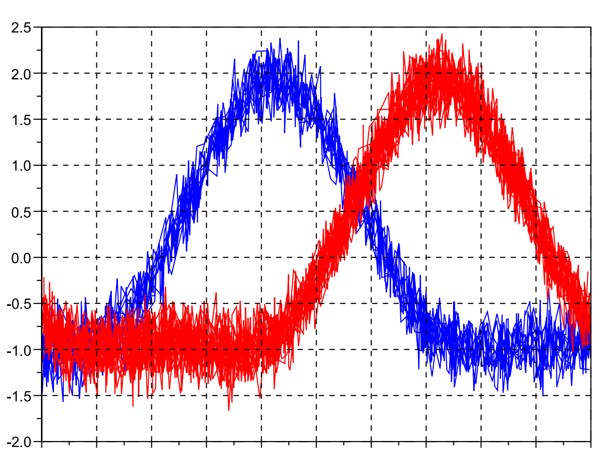

FIG. 3. *Simulated data with two different values θ and θ'.*

to the value $\theta' = 1/10$ as $\widehat{\Delta}_n = 0.0998 + 0.1950 = 0.2948$. In order to compute more accurate values of $\widehat{\theta}_n$, one can replace $\gamma_n = 1/n$ in (2.4) by $\gamma_n = 1/n^a$ where $1/2 < a < 1$. This will be done for the implementation of our Robbins–Monro procedure on real ECG data.

4.2. *Real ECG data.* We shall now focus our attention on real ECG data. We have chosen the record 04015 in the Atrial Fibrillation (AF) database provided by MIT-BIH database. Each recording consists in a continuous digitized ECG signal measured over 1 hour in order to detect AF which is the most common cardiac arrythmia. A stronger indicator of AF is the absence of P waves or the irregularities of RR interval on an electrocardiogram. We refer the reader to [6] for an interesting book on statistical methods and tools for ECG data analysis. Our aim is to propose a statistical procedure in order to detect irregularities of RR interval on the ECG record 04015. The record and its projection on the interval $[-1/2, 1/2]$ are given in Figure 4. The size of the data set is 2038. We assume that the model

$$Y_n = f(X_n - \theta) + \varepsilon_n$$

fits the data, where the sequence (X_n) is uniformly distributed over the interval $[-1/2, 1/2]$. The periodic shape function f is clearly not symmetric. However, we already saw in Remark 2.3 that our Robbins–Monro procedure still holds for nonsymmetric regression function.

As for simulated data, in view of the signal, we would find two different values θ and θ'. The first value θ is associated with the first part of the signal, while the second value θ' corresponds to the second part. The difference $\Delta = \theta' - \theta$ between the two parameters would explain the lag between the two parts of the signal. A value of Δ far away from zero could be interpreted as the detection of irregularities of RR interval which confirms the diagnostic of atrial fibrillation. On this record, our Robbins–Monro procedure with $n = 800$ iterations leads to the first estimate $\widehat{\theta}_n = 0.1734$ for θ and the last estimate $\widehat{\theta}_n = -0.0092$ for θ' with

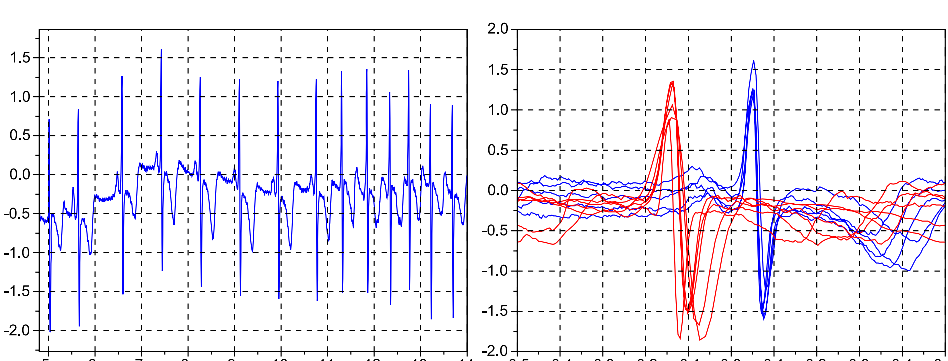

FIG. 4. *Original data.*

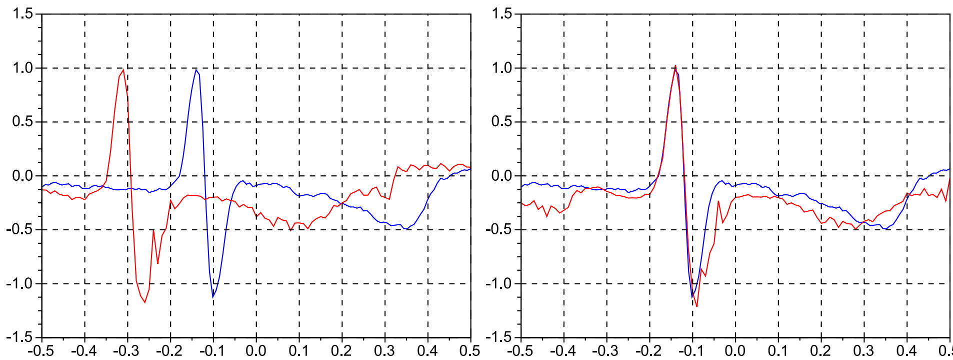

FIG. 5. *Reconstruction of the ECG.*

$n = 1238$. The value $\widehat{\Delta}_n = -0.0092 - 0.1734 = -0.1826$ explains the lag in Figure 4. Figure 5 shows that our Nadaraya–Watson procedure for the reconstruction of ECG signals works pretty well.

5. Proofs of the parametric results.

5.1. *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, $\mathcal{F}_n = \sigma(X_0, \varepsilon_0, \dots, X_n, \varepsilon_n)$. First of all, we shall calculate the two first conditional moments of the random variable T_n given by (2.5). It follows from (1.2) that

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

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

$$\mathbb{E}\left[\frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))f(X_{n+1} - \theta)}{g(X_{n+1})}\Big|\mathcal{F}_n\right] = \phi(\widehat{\theta}_n) \qquad \text{a.s.,} \tag{5.1}$$

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

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

Hence, (5.1) leads to

$$\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n) \qquad \text{a.s.} \tag{5.2}$$

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

$$\mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = \mathbb{E}\biggl[\frac{\sin^2(2\pi(X-\widehat{\theta}_n))}{g^2(X)}\bigl(f^2(X-\theta)+\sigma^2\bigr)\biggr] = \varphi(\widehat{\theta}_n), \tag{5.3}$$

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 (5.3) that for some constant $M > 0$

$$\sup_{n\geq 0} \mathbb{E}[T_{n+1}^2|\mathcal{F}_n] \leq M \qquad \text{a.s.} \tag{5.4}$$

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

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

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

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

Hence, it follows from (5.2) and (5.4) that

$$\begin{aligned} \mathbb{E}[V_{n+1}|\mathcal{F}_n] &\leq V_n + \gamma_{n+1}^2\mathbb{E}[T_{n+1}^2|\mathcal{F}_n] + 2\gamma_{n+1}(\widehat{\theta}_n - \theta)\mathbb{E}[T_{n+1}|\mathcal{F}_n] \\ &\leq V_n + \gamma_{n+1}^2 M + 2\gamma_{n+1}(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n) \qquad \text{a.s.} \end{aligned} \tag{5.5}$$

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 (5.5) together with the Robbins–Siegmund theorem (see Duflo [9], page 18) that the sequence (V_n) converges a.s. to a finite random variable V and

$$\sum_{n=1}^{\infty} \gamma_{n+1}(\theta - \widehat{\theta}_n)\phi(\widehat{\theta}_n) < +\infty \qquad \text{a.s.} \tag{5.6}$$

Assume by contradiction that $V \neq 0$ a.s. Then, one can find $0 < a < b < 1/2$ such that, for n large enough, the event $\{a < |\widehat{\theta}_n - \theta| < b\}$ is not negligible. 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 (5.6), 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 $n_k > 0$,

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

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.

5.2. *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 ([22], page 330). First of all, as $\gamma_n = 1/n$, the condition on the decreasing step is satisfied. Moreover, we already saw that $\widehat{\theta}_n$ converges almost surely to θ. Consequently, all the local assumptions of Theorem 2.1 of [22] are satisfied. In addition, it follows from (5.2) that $\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n)$ a.s. and the function ϕ is continuously differentiable since $\phi(t) = f_1 \sin(2\pi(\theta - t))$. Hence, $\phi(\theta) = 0$ and $\phi'(\theta) = -2\pi f_1$ and $4\pi f_1 > 1$. Furthermore, we deduce from (5.3) that

$$\mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = \varphi(\widehat{\theta}_n) \qquad \text{a.s.,}$$

which leads to

$$\lim_{n \to \infty} \mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = \varphi(\theta) \qquad \text{a.s.}$$

Consequently, if we are able to prove that the sequence (W_n) given by

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

is tight, then we shall deduce from Theorem 2.1 of [22] that

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

where

$$\xi^2(\theta) = \varphi(\theta) \int_0^{+\infty} \exp\bigl((1 - 4\pi f_1)t\bigr)\, dt = \frac{\varphi(\theta)}{4\pi f_1 - 1}.$$

Therefore, it remains to prove the tightness of the sequence (W_n). It follows from (5.5) that for some constant $M > 0$ and for all $n \geq 1$,

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] \leq (1 + \gamma_n)W_n + \gamma_{n+1}M + 2(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n). \tag{5.7}$$

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 $|x - \theta| < \varepsilon$,

$$\frac{q}{2 f_1} < v(x) < 0. \tag{5.8}$$

We also deduce from (5.7) that for all $n \geq 1$,

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] \leq W_n + 2\gamma_n W_n\bigl(q - f_1 v(\widehat{\theta}_n)\bigr) + \gamma_n M \tag{5.9}$$

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 (5.8) that

$$0 < -f_1 v(\widehat{\theta}_n)\mathrm{I}_{B_n} < -\left(\frac{q}{2}\right)\mathrm{I}_{B_n}. \tag{5.10}$$

Hence, we deduce from the conjunction of (5.9) and (5.10) that for all $n \geq m$,

$$\begin{aligned} \mathbb{E}[W_{n+1}\mathrm{I}_{B_n}|\mathcal{F}_n] &\leq W_n\mathrm{I}_{B_n} + 2\gamma_n W_n\mathrm{I}_{B_n}\left(q - \frac{q}{2}\right) + \gamma_n M \\ &\leq W_n\mathrm{I}_{B_n}(1 + q\gamma_n) + \gamma_n M. \end{aligned} \tag{5.11}$$

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 (5.11) that for all $n \geq m$,

$$\mathbb{E}[W_{n+1}\mathrm{I}_{B_{n+1}}] \leq (1 + q\gamma_n)\mathbb{E}[W_n\mathrm{I}_{B_n}] + \gamma_n M. \tag{5.12}$$

From now on, denote $\alpha_n = \mathbb{E}[W_n\mathrm{I}_{B_n}]$. We infer from (5.12) that for all $n \geq m$,

$$\alpha_{n+1} \leq \beta_n\alpha_m + M\beta_n \sum_{k=m}^{n} \frac{\gamma_k}{\beta_k} \qquad \text{where } \beta_n = \prod_{k=m}^{n} (1 + q\gamma_k). \tag{5.13}$$

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, (5.13) immediately leads to

$$\sup_{n\geq m} \alpha_n < +\infty. \tag{5.14}$$

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,

$$\begin{aligned} \mathbb{P}(W_n > K) &\leq \mathbb{P}(W_n\mathrm{I}_{B_n} > K/2) + \mathbb{P}(W_n\mathrm{I}_{\overline{B}_n} > K/2) \\ &\leq \frac{2}{K}\mathbb{E}[W_n\mathrm{I}_{B_n}] + \mathbb{P}(\overline{B}_n). \end{aligned} \tag{5.15}$$

We deduce from (5.14) that one can find K depending on ξ such that the first term on the right-hand side of (5.15) 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.

5.3. *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 as the classical Robbins–Monro algorithm. Consequently, we deduce the law of iterated logarithm given by (2.10) from Theorem 1 of [12]; see also Hall and Heyde ([13], page 240), and the quadratic strong law given by (2.12) from Theorem 3 of [31].

6. Proofs of the nonparametric results.

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

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

As in [1], we obtain from (1.2) the decomposition

$$n\widehat{h}_n(x) = M_n(x) + P_n(x) + Q_n(x) + n\widehat{g}_n(x) f(x), \tag{6.1}$$

$$n\widehat{g}_n(x) = N_n(x) + R_n(x) + ng(\theta + x), \tag{6.2}$$

where

$$M_n(x) = \sum_{k=1}^{n} W_k(x)\varepsilon_k, \tag{6.3}$$

$$N_n(x) = \sum_{k=1}^{n} W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] \tag{6.4}$$

and

$$P_n(x) = \sum_{k=1}^{n} W_k(x)\big(f(X_k - \widehat{\theta}_{k-1}) - f(x)\big), \tag{6.5}$$

$$Q_n(x) = \sum_{k=1}^{n} W_k(x)\big(f(X_k - \theta) - f(X_k - \widehat{\theta}_{k-1})\big), \tag{6.6}$$

$$R_n(x) = \sum_{k=1}^{n} \big(\mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] - g(\theta + x)\big). \tag{6.7}$$

On the one hand,

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

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

$$
\begin{aligned}
&\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] \\
&\qquad = \int_{\mathbb{R}} K(z) g(\widehat{\theta}_{n-1} + x + h_n z)\, dz \\
&\qquad = \int_{\mathbb{R}} K(z)\Big(g(\widehat{\theta}_{n-1} + x) + h_n z g'(\widehat{\theta}_{n-1} + x) \\
&\qquad\qquad\qquad + \frac{h_n^2 z^2}{2} g''(\widehat{\theta}_{n-1} + x + h_n z\xi)\Big)\, dz \\
&\qquad = g(\widehat{\theta}_{n-1} + x) + \frac{h_n^2}{2}\int_{\mathbb{R}} z^2 K(z) g''(\widehat{\theta}_{n-1} + x + h_n z\xi)\, dz,
\end{aligned} \tag{6.8}
$$

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

$$
\big|\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - g(\widehat{\theta}_{n-1} + x)\big| \leq M_g \tau^2 h_n^2 \qquad \text{a.s.}, \tag{6.9}
$$

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

$$
\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] = g(\theta + x) \qquad \text{a.s.}, \tag{6.10}
$$

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

$$
R_n(x) = o(n) \qquad \text{a.s.} \tag{6.11}
$$

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

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

It follows from the same calculation as in (6.8) 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

$$\left|\mathbb{E}[W_n^2(x)|\mathcal{F}_{n-1}] - \frac{\nu^2}{h_n} g(\widehat{\theta}_{n-1} + x)\right| \le M_g \mu^2 h_n \qquad \text{a.s.} \tag{6.12}$$

with

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

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 (6.9) and (6.12) together with the Toeplitz lemma and the almost sure convergence of $g(\widehat{\theta}_n + x)$ to $g(\theta + x)$ that

$$\lim_{n\to\infty} \frac{\langle N(x)\rangle_n}{n^{1+\alpha}} = \frac{\nu^2 g(\theta + x)}{1+\alpha} \qquad \text{a.s.} \tag{6.13}$$

Consequently, we obtain from the strong law of large numbers for martingales given, for example, 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 $x \in \mathbb{R}$

$$N_n(x) = o(n) \qquad \text{a.s.} \tag{6.14}$$

Therefore, it follows from (6.2), (6.11) and (6.14) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \widehat{g}_n(x) = g(\theta + x) \qquad \text{a.s.} \tag{6.15}$$

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 $n \ge 1$ and all $x \in \mathbb{R}$,

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

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

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

Consequently, we obtain from (6.5) that for all $x \in \mathbb{R}$

$$\begin{aligned} |P_n(x)| &\le C_f \sum_{k=1}^{n} W_k(x) |X_k - \widehat{\theta}_{k-1} - x| \\ &\le A C_f \sum_{k=1}^{n} h_k W_k(x). \end{aligned} \tag{6.16}$$

Hence, it follows from convergence (6.10) together with (6.14) and (6.16) that for all $x \in \mathbb{R}$

$$P_n(x) = o(n) \qquad \text{a.s.} \tag{6.17}$$

Furthermore, we obtain from (6.6) that for all $x \in \mathbb{R}$

$$|Q_n(x)| \le C_f \sum_{k=1}^{n} W_k(x)|\widehat{\theta}_{k-1} - \theta|. \tag{6.18}$$

Then, it follows from the Cauchy–Schwarz inequality that

$$Q_n^2(x) \le C_f^2 \sum_{k=1}^{n} W_k^2(x) \sum_{k=1}^{n} |\widehat{\theta}_{k-1} - \theta|^2. \tag{6.19}$$

We can split the first sum at the right-hand side of (6.19) into two terms,

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

where

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

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

Following the same lines as in the proof of (6.14), it is not hard to see that

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

We also deduce from convergence (6.13) that

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

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

$$\sum_{k=1}^{n} W_k^2(x) = O(n^{1+\alpha}) \qquad \text{a.s.} \tag{6.20}$$

Therefore, we infer from the quadratic strong law given by (2.12) together with (6.19) and (6.20) that $Q_n^2(x) = O(n^{1+\alpha} \log n)$ a.s. which implies that for all $x \in \mathbb{R}$

$$Q_n(x) = o(n) \qquad \text{a.s.} \tag{6.21}$$

It now remains to study the asymptotic behavior of $M_n(x)$ given by (6.3). 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

$$\langle M(x)\rangle_n = \sum_{k=1}^{n} \mathbb{E}\big[\big(M_k(x) - M_{k-1}(x)\big)^2 | \mathcal{F}_{k-1}\big]$$

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

Then, it follows from convergence (6.13) that

$$\lim_{n\to\infty} \frac{\langle M(x)\rangle_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2 g(\theta+x)}{1+\alpha} \qquad \text{a.s.} \tag{6.22}$$

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

$$M_n(x) = o(n) \qquad \text{a.s.} \tag{6.23}$$

Therefore, we deduce from (6.1) and (6.15) together with the conjunction of (6.17), (6.21) and (6.23) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \widehat{h}_n(x) = f(x)g(\theta+x) \qquad \text{a.s.} \tag{6.24}$$

Finally, we can conclude from the identity

$$\widehat{f}_n(x) = \frac{\widehat{h}_n(x) + \widehat{h}_n(-x)}{\widehat{g}_n(x) + \widehat{g}_n(-x)} \tag{6.25}$$

and the parity of the function f that, for all $x \in \mathbb{R}$ such that $|x| \le 1/2$,

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \qquad \text{a.s.} \tag{6.26}$$

6.2. *Proof of Theorem* 3.2. We shall now proceed to the proof of the asymptotic normality of $\widehat{f}_n$. It follows from (6.1), (6.2) and (6.25) that for all $x \in \mathbb{R}$

$$\widehat{f}_n(x) - f(x) = \frac{\mathcal{M}_n(x) + \mathcal{P}_n(x) + \mathcal{Q}_n(x)}{n\mathcal{G}_n(x)}, \tag{6.27}$$

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

$$\mathcal{M}_n(x) = M_n(x) + M_n(-x),$$

$$\mathcal{P}_n(x) = P_n(x) + P_n(-x),$$

$$\mathcal{Q}_n(x) = Q_n(x) + Q_n(-x)$$

with $M_n(x)$, $P_n(x)$ and $Q_n(x)$ given by (6.3), (6.5) and (6.6), respectively. We already saw from (6.15) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathcal{G}_n(x) = g(\theta+x) + g(\theta-x) \qquad \text{a.s.} \tag{6.28}$$

In order to establish the asymptotic normality, it is now necessary to be more precise in the almost sure rates of convergence given in (6.17) and (6.21). It follows from (6.16) that for all $x \in \mathbb{R}$

$$|P_n(x)| \leq AC_f\big(L_n(x) + \Lambda_n(x)\big), \tag{6.29}$$

where

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

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

On the one hand, we infer from (6.9) that

$$\Lambda_n(x) = O\left(\sum_{k=1}^{n} h_k\right) = O(n^{1-\alpha}) \qquad \text{a.s.} \tag{6.30}$$

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

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

We deduce from (6.9) and (6.12) together with the Toeplitz lemma that

$$\lim_{n\to\infty} \frac{\langle L(x)\rangle_n}{n^{1-\alpha}} = \frac{\nu^2 g(\theta + x)}{1-\alpha} \qquad \text{a.s.} \tag{6.31}$$

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 (6.29) and (6.30) 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

$$(\mathcal{P}_n(x))^2 = o(n^{1+\alpha}) \qquad \text{a.s.} \tag{6.32}$$

Proceeding as in the proof of (6.32), we obtain from (6.18) that for all $x \in \mathbb{R}$

$$|Q_n(x)| \leq C_f\big(S_n(x) + \Sigma_n(x)\big), \tag{6.33}$$

where

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

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

with $\ell_n = |\widehat{\theta}_{n-1} - \theta|$. We deduce from (6.9) together with the Cauchy–Schwarz inequality and the quadratic strong law given by (2.12) that

$$\Sigma_n(x) = O\left(\sum_{k=1}^{n} \ell_k\right) = O\big(\sqrt{n \log n}\big) \qquad \text{a.s.} \tag{6.34}$$

In addition, it follows from (6.12) that $(S_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation satisfying

$$\langle S(x)\rangle_n = O(n^{\alpha} \log n) \qquad \text{a.s.}$$

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 (6.33) and (6.34) that

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

which obviously implies

$$(\mathcal{Q}_n(x))^2 = o(n^{1+\alpha}) \qquad \text{a.s.} \tag{6.35}$$

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

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

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 (6.8) that

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

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

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

where

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

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 the 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 $x \neq 0$,

$$\sum_{k=1}^{n} \mathbb{E}[W_k(x)W_k(-x)|\mathcal{F}_{k-1}] = o\left(\sum_{k=1}^{n} h_k^{-1}\right) = o(n^{1+\alpha}) \qquad \text{a.s.} \tag{6.36}$$

Therefore, we find from (6.22) together with (6.36) that for all $x \in \mathbb{R}$ with $x \neq 0$,

$$\lim_{n\to\infty} \frac{\langle \mathcal{M}(x)\rangle_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2}{1+\alpha}\big(g(\theta + x) + g(\theta - x)\big) \qquad \text{a.s.} \tag{6.37}$$

If $x = 0$, it immediately follows from (6.22)

$$\lim_{n\to\infty} \frac{\langle \mathcal{M}(0)\rangle_n}{n^{1+\alpha}} = \frac{4\sigma^2\nu^2 g(\theta)}{1+\alpha} \qquad \text{a.s.} \tag{6.38}$$

Furthermore, it is not hard to see that the 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 (6.8) that

$$\sum_{k=1}^{n} \mathbb{E}[W_k^a(x)|\mathcal{F}_{k-1}] = O\left(\sum_{k=1}^{n} h_k^{1-a}\right) = O\big(n^{1+\alpha(a-1)}\big) \qquad \text{a.s.} \tag{6.39}$$

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

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

where $b = a(1+\alpha)/2$. Consequently, it follows from (6.39) 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 $c < 0$, the Lindeberg condition is clearly satisfied. We can conclude from the central limit theorem for martingales given, for example, by Corollary 2.1.10 of [9] that for all $x \in \mathbb{R}$ with $x \neq 0$,

$$\frac{\mathcal{M}_n(x)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{1+\alpha}(g(\theta+x) + g(\theta-x))\right), \tag{6.40}$$

while, for $x = 0$,

$$\frac{\mathcal{M}_n(0)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{4\sigma^2 \nu^2}{1+\alpha} g(\theta)\right). \tag{6.41}$$

Finally, it follows from (6.27) and (6.28) together with (6.32), (6.35), (6.40), (6.41) and the Slutsky lemma that, for all $x \in \mathbb{R}$ such that $|x| \leq 1/2$ with $x \neq 0$,

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

while, for $x = 0$,

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

which completes the proof of Theorem 3.2.

Acknowledgments. The authors would like to thank the Associate Editor and the two anonymous reviewers for their suggestions and constructive comments which helped to improve the paper substantially.

REFERENCES

[1] BERCU, B. and PORTIER, B. (2008). Kernel density estimation and goodness-of-fit test in adaptive tracking. *SIAM J. Control Optim.* **47** 2440–2457. MR2448468

[2] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. MR1623559

[3] CASTILLO, I. and LOUBES, J. M. (2009). Estimation of the distribution of random shifts deformation. *Math. Methods Statist.* **18** 21–42. MR2508947

[4] CHEN, H. F., LEI, G. and GAO, A. J. (1988). Convergence and robustness of the Robbins–Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27** 217–231. MR0931029

[5] CHOI, E., HALL, P. and ROUSSON, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *Ann. Statist.* **28** 1339–1355. MR1805786

[6] CLIFFORD, G. D., AZUAJE, F. and MCSHARRY, P. (2006). *Advanced Methods and Tools for ECG Data Analysis*. Artech House, Boston.

[7] DALALYAN, A. S., GOLUBEV, G. K. and TSYBAKOV, A. B. (2006). Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist.* **34** 169–201. MR2275239

[8] DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation*. Springer, New York. MR1843146

[9] DUFLO, M. (1997). *Random Iterative Models*. *Applications of Mathematics* (*New York*) **34**. Springer, Berlin. MR1485774

[10] FABIAN, V. (1973). Asymptotically efficient stochastic approximation; the RM case. *Ann. Statist.* **1** 486–495. MR0381189

[11] GAMBOA, F., LOUBES, J.-M. and MAZA, E. (2007). Semi-parametric estimation of shifts. *Electron. J. Stat.* **1** 616–640. MR2369028

[12] GAPOŠKIN, V. F. and KRASULINA, T. P. (1974). The law of the iterated logarithm in stochastic approximation processes. *Teor. Verojatnost. i Primenen.* **19** 879–886. MR0365954

[13] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press [Harcourt Brace Jovanovich Publishers], New York. MR0624435

[14] HALL, P. and HUANG, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **29** 624–647. MR1865334

[15] HÄRDLE, W. (1984). A law of the iterated logarithm for nonparametric regression function estimators. *Ann. Statist.* **12** 624–635. MR0740916

[16] HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449. MR0964932

[17] HÄRDLE, W. and MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89. MR1041386

[18] HÄRDLE, W. and TSYBAKOV, A. B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* **16** 120–135. MR0924860

[19] HÜRTGEN, H. and GERVINI, D. (2009). Semiparametric shape-invariant models for periodic data. *J. Appl. Stat.* **36** 1055–1065. MR2744122

[20] KNEIP, A. and ENGEL, J. (1995). Model estimation in nonlinear regression under shape invariance. *Ann. Statist.* **23** 551–570. MR1332581

[21] KNEIP, A. and GASSER, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.* **16** 82–112. MR0924858

[22] KUSHNER, H. J. and YIN, G. G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. *Applications of Mathematics* (*New York*) **35**. Springer, New York.

[23] LASSEN, K. and FRIIS-CHRISTENSEN, E. (1995). Variability of the solar cycle length during the past five centuries and the apparent association with terrestrial climate. *J. Atmospheric and Terrestrial Physics* **57** 835–845.

[24] LAWTON, W. H., SYLVESTRE, E. A. and MAGGIO, M. S. (1972). Self modeling nonlinear regression. *Technometrics* **14** 513–532.

[25] LELONG, J. (2008). Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett.* **78** 2632–2636. MR2542461

[26] MCDONALD, J. A. (1986). Periodic smoothing of time series. *SIAM J. Sci. Statist. Comput.* **7** 665–688. MR0833929

[27] MOKKADEM, A. and PELLETIER, M. (2007). A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm. *Ann. Statist.* **35** 1749–1772. MR2351104

[28] NADARAJA, È. A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.* **9** 157–159. MR0166874

[29] NODA, K. (1976). Estimation of a regression function by the Parzen kernel-type density estimators. *Ann. Inst. Statist. Math.* **28** 221–234. MR0426278

[30] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076. MR0143282

[31] PELLETIER, M. (1998). On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic Process. Appl.* **78** 217–244. MR1654569

[32] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407. MR0042668

[33] ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics* (*Proc. Sympos., Ohio State Univ., Columbus, Ohio*, 1971) 233–257. Academic Press, New York. MR0343355

[34] SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88. MR0301845

[35] STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284. MR0362669

[36] TRIGANO, T., ISSERLES, U. and RITOV, Y. (2011). Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Process.* **59** 1970–1984. MR2816476

[37] TSYBAKOV, A. B. (2004). *Introduction à L'estimation Non-Paramétrique*. *Mathématiques and Applications* (*Berlin*) **41**. Springer, Berlin. MR2013911

[38] VIMOND, M. (2010). Efficient estimation for a subclass of shape invariant models. *Ann. Statist.* **38** 1885–1912. MR2662362

[39] WANG, Y. and BROWN, M. B. (1996). A flexible model for human circadian rhythms. *Biometrics* **52** 588–596.

[40] WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372. MR0185765

INSTITUT DE MATHÉMATIQUES DE BORDEAUX
UMR CNRS
UNIVERSITÉ DE BORDEAUX
AND
INRIA BORDEAUX, TEAM ALEA
F-33400 TALENCE
FRANCE
E-MAIL: Bernard.Bercu@math.u-bordeaux1.fr
Philippe.Fraysse@math.u-bordeaux1.fr