

A ROBBINS MONRO PROCEDURE FOR THE ESTIMATION OF PARAMETRIC DEFORMATIONS ON RANDOM VARIABLES

Philippe Fraysse¹, Helene Lescornel² and Jean-Michel Loubes³

¹*Université de Bordeaux*, ²*Inria Saclay* and ³*Université Paul Sabatier*

Abstract: The paper is devoted to the study of a parametric deformation model of independent and identically random variables. We construct an efficient and easy-to-compute recursive estimate of the parameter. Our stochastic estimator is similar to the Robbins-Monro procedure where the contrast function is the Wasserstein distance. We then propose a recursive estimator similar to that of Parzen-Rosenblatt kernel density estimator in order to estimate the density of the random variables. This estimate takes into account the previous estimation of the parameter of the model. Finally, we illustrate the performance of our estimation procedure on simulations for the Box-Cox transformation and the arcsinh transformation.

Key words and phrases: Asymptotic properties, estimation of a regression function, estimation of shifts, semiparametric estimation.

1. Introduction

In many situations, random variables are observed only through their image by a deformation. Hence, finding the *mean behaviour* of a data sample is a difficult task since the usual notion of Euclidean mean is too rough when the information conveyed by the data possesses an inner geometry far from Euclidean. Indeed, such deformations on the data as translations, scale location models, or more general warping procedures prevent the use of the usual methods in data analysis.

Deformations may result from variations that are not directly correlated to the studied phenomenon. This situation occurs often in biology, for example, when considering gene expression data obtained from microarray technologies to measure genome-wide expression levels of genes in a given organism, as described in Bolstad et al. (2003). A natural way to handle this phenomenon is to remove the variations in order to align the measured densities, but this is difficult to implement since the densities are unknown. In bioinformatics and computational biology, a method to reduce this kind of variability is known as normalization (see Gallón, Loubes, and Maza (2013) and references therein). In epidemiology, removing variations is important in medical studies where one observes age-at-death of several cohorts. Indeed, the individuals of the cohort enjoy different life

conditions which means that time-variation is likely to exist between the cohort densities and hazard rates due to the effects of the different biotopes on aging. Synchronization of the different observations is a crucial point.

Variations on the observations are often due to transformations that have been conducted by the statisticians themselves. In econometric science, transformations have been used to aid interpretability as well as to improve statistical performance of some indicators. An important contribution to this methodology was made by Box and Cox in Box and Cox (1964) who proposed a parametric power family of transformations that nested the logarithm and the level. Estimation in this framework is achieved in Linton, Sperlich, and Van Keilegom (2008).

In this work, we concentrate on the case where the data and their transformation are observed in a sequence model defined, for all $n \geq 0$, by

$$X_n = \varphi_\theta(\varepsilon_n) \tag{1.1}$$

where, for all $t \in \mathbb{R}$, the family of parametric functions (φ_t) is known and (ε_n) is a sequence of independent and identically distributed random variables. Our approach to estimating θ is associated with a stochastic recursive algorithm similar to that of Robbins-Monro described in Robbins and Monro (1951) and Robbins and Siegmund (1971).

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

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

where (γ_n) is a positive sequence of real numbers decreasing to zero and (T_n) is a sequence of random variables such that $\mathbb{E}[T_{n+1} | \mathcal{F}_n] = \phi(\hat{\theta}_n)$, 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 (see Duflo (1997) and Kushner and Yin (2003)) that $\hat{\theta}_n$ tends to θ almost surely. The asymptotic normality of $\hat{\theta}_n$ together with the quadratic strong law can be found in Hall and Heyde (1980). A randomly truncated version of the Robbins-Monro algorithm is given in Chen, Lei, and Gao (1988), Lelong (2008), whereas we can find in Bercu and Fraysse (2012) an application of the Robbins-Monro algorithm in semiparametric regression models. In our framework, if we assume that φ_t is invertible, then one can consider

$$Z_n(t) = \varphi_t^{-1}(X_n).$$

Hence, a natural registration criterion is to minimize with respect to t the quadratic distance between $Z_n(t)$ and ε_n ,

$$M(t) = \mathbb{E} \left[|Z_n(t) - \varepsilon_n|^2 \right].$$

It is then obvious that the parameter θ is a global minimum of M and one can implement a Robbins-Monro procedure for the contrast function M' , which is the differential of the L^2 function M .

The second part of the paper concerns the estimation of the density f of the random variables (ε_n) . More precisely, we focus our attention on the Parzen-Rosenblatt estimator of f described for instance in Parzen (1962) or Rosenblatt (1956). Under reasonable conditions on the function f , Parzen established in Parzen (1962) the pointwise convergence in probability and the asymptotic normality of the estimator without the parameter θ . In Silverman (1978), Silverman obtained uniform consistency properties of the estimator. Contributions to the L^1 -integrated risk have been obtained by Devroye in Devroye (1988), whereas Hall has studied the L^2 -integrated risk in Hall (1982) and Hall (1984). We propose to make use of a recursive Parzen-Rosenblatt estimator of f that takes into account the previous estimate of θ . It is given, for all $x \in \mathbb{R}$, by

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n W_i(x) \quad (1.3)$$

with

$$W_i(x) = \frac{1}{h_i} K \left(\frac{x - Z_i(\widehat{\theta}_{i-1})}{h_i} \right),$$

where the kernel K is a chosen probability density function and the bandwidth (h_i) is a sequence of positive real numbers decreasing to zero. The main difficulty arising here is that we have to deal with the term $Z_i(\widehat{\theta}_{i-1})$ inside the kernel K .

We proceed as follows. Section 2 is devoted to the description of the model. Section 3 deals with the estimation of θ ; we establish the almost sure convergence of $\widehat{\theta}_n$ as well as its asymptotic normality. In Section 4, under standard regularity assumptions on the kernel K , we prove the almost sure pointwise and quadratic convergences of $\widehat{f}_n(x)$ to $f(x)$. Section 5 contains some numerical experiments on the well-known Box-Cox transformation and on the arcsinh transformation illustrating the performances of our parametric estimation procedure. The proofs of the parametric results are given in Section 6, while those concerning the non-parametric results are in Section 7.

2. Description of the Model and the Criterion

Suppose that we observe independent and identically distributed random variables ε_n and a deformation X_n of ε_n defined, for all $n \geq 0$, by

$$X_n = \varphi_\theta(\varepsilon_n),$$

where $\theta \in \Theta \subset \mathbb{R}$. Throughout, we denote by ε and X random variables sharing the same distribution as ε_n and X_n , respectively.

Assume that for all $t \in \mathbb{R}$, the family of parametric functions (φ_t) is known but that the parameter θ is unknown. This situation corresponds to the case where the warping operator can be modeled by a parametric shape. Estimating the parameter is the key to understanding the amount of deformation in the chosen deformation class. This model has been widely used in the regression case, see for instance in Gamboa, Loubes, and Maza (2007). Assume also that, for all $t \in \mathbb{R}$, φ_t is invertible on an interval to be made precise in the next section. Then, one can consider the random variable

$$Z_n(t) = \varphi_t^{-1}(X_n) = \varphi_t^{-1}(\varphi_\theta(\varepsilon_n)). \quad (2.1)$$

We denote by $Z(t)$ a random variable sharing the same distribution as $Z_n(t)$. In order to estimate θ , we choose to evaluate the L^2 distance between ε and $Z(t)$,

$$M(t) = \mathbb{E} \left[|Z(t) - \varepsilon|^2 \right]. \quad (2.2)$$

If F^{-1} is the quantile function associated with ε , we can write

$$M(t) = \mathbb{E} \left[|\varphi_t^{-1}(\varphi_\theta(\varepsilon)) - \varepsilon|^2 \right] = \int_0^1 (\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) - F^{-1}(x))^2 dx.$$

Indeed (see for instance Van der Vaart (2000) p.305), if Y is a random variable with distribution function G , then for $U \sim \mathcal{U}_{[0;1]}$, $Y \sim G^{-1}(U)$.

If we assume that φ_t is increasing for all t , then one has an expression for the quantile function associated with $Z(t)$: $F_{Z(t)}^{-1} = \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}$, and so

$$M(t) = \int_0^1 \left(F_{Z(t)}^{-1}(x) - F^{-1}(x) \right)^2 dx.$$

This quantity corresponds to the Wasserstein distance between the laws of $Z(t)$ and ε , defined and studied for instance in Cuesta and Matrán (1989) in the general case. Using Wasserstein metrics to align distributions is rather natural since it corresponds to the transportation cost between two probability laws. It is also a proper criterion for studying similarities between point distributions (see for instance Munk and Czado (1998)), and is already used for density registration in Agullo et al. (2014) or Gallón, Loubes, and Maza (2013) in a non-sequential way.

Here, considering the L^2 distance between the starting point and the registered point is equivalent to investigating the Wasserstein distance between their laws.

As $M(\theta) = 0$ and the function M defined by (2.2) is non-negative, it is clear that M admits at least a global minimum at θ which allows us a characterization of the parameter of interest.

3. Estimation of the Parameter θ

We focus attention on the estimation of the parameter $\theta \in \Theta$ with Θ an interval of \mathbb{R} . We require some assumptions.

For all $t \in \Theta$, φ_t is invertible, increasing from I_1 to I_2 , some subsets of \mathbb{R} . (A1)

For all $x \in I_2$, $\varphi_t^{-1}(x)$ is continuously differentiable with respect to $t \in \Theta$, with derivative $\partial\varphi_t^{-1}(x)$. (A2)

For all $t \in \Theta$, $\varphi_t^{-1} \circ \varphi_\theta \in L^2(\varepsilon)$. (A3)

For all compact B in Θ , $\mathbb{E} \left[\sup_{t \in B} |\partial\varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^4 \right] < +\infty$. (A4)

From (A1), the distribution function of X is $F_X = F \circ \varphi_\theta^{-1}$ whereas that of $Z(t)$ is $F \circ \varphi_\theta^{-1} \circ \varphi_t$.

Lemma 1. *If (A1) to (A4) hold, then M is continuously differentiable on Θ .*

Using Lemma 1, the differential M' of M is, for all $t \in \Theta$,

$$\begin{aligned} M'(t) &= -2 \int_0^1 \partial\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx \\ &= -2\mathbb{E} [\partial\varphi_t^{-1}(X) (\varepsilon - \varphi_t^{-1}(X))]. \end{aligned} \tag{3.1}$$

It is then clear that $M'(\theta) = 0$. Then, we can assume that there exists $\{a, b\} \in \Theta^2$ with $a < b$ and $\theta \in]a; b[\subset \Theta$ such that, for all $t \in [a; b]$,

$$(t - \theta)M'(t) > 0. \tag{A5}$$

We can now implement our Robbins-Monro procedure. Let $\pi_{[a;b]}$ be the projection on the compact set $[a; b]$ defined for all $x \in [a; b]$ by

$$\pi_{[a;b]}(x) = x\mathbb{I}_{\{a \leq x \leq b\}} + a\mathbb{I}_{\{x \leq a\}} + b\mathbb{I}_{\{x \geq b\}}.$$

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

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty. \tag{3.2}$$

We estimate the parameter θ via the projected Robbins-Monro algorithm

$$\widehat{\theta}_{n+1} = \pi_{[a;b]}(\widehat{\theta}_n - \gamma_{n+1}T_{n+1}), \tag{3.3}$$

where the deterministic initial value $\widehat{\theta}_0 \in [a; b]$ and the random variable T_{n+1} is

$$T_{n+1} = -2\partial\varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \right). \tag{3.4}$$

Theorem 1. *Assume (A1) to (A5), with $\theta \in]a; b[$ where $a < b$. Then, $\widehat{\theta}_n$ converges almost surely to θ .*

To control the rate of convergence of $\widehat{\theta}_n$, we need slightly stronger condition of regularity on the deformation functions.

For all $x \in I_2$, $\varphi_t^{-1}(x)$ is twice differentiable with respect to $t \in \Theta$ and, for all compact B in Θ , $\mathbb{E} \left[\sup_{t \in B} |\partial^2 \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^2 \right] < +\infty$. (A6)

Lemma 2. *If (A1) to (A6) hold, then M is twice continuously differentiable on Θ .*

We compute the second differential of M'' of M for all $t \in \Theta$ as

$$\begin{aligned} M''(t) &= 2 \int_0^1 [\partial\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)]^2 dx \\ &\quad - 2 \int_0^1 \partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx \end{aligned} \tag{3.5}$$

so

$$M''(t) = 2\mathbb{E} \left[(\partial\varphi_t^{-1}(X))^2 \right] - 2\mathbb{E} \left[\partial^2 \varphi_t^{-1}(X) (\varepsilon - \varphi_t^{-1}(X)) \right]. \tag{3.6}$$

For the sake of clarity, we make use of $\gamma_n = 1/n$ for the following theorem.

Theorem 2. *Assume (A1) to (A6), with $\theta \in]a; b[$ where $a < b$. Suppose $M''(\theta) > 1/2$ and that there exists $\alpha > 4$ such that for all compact B in Θ ,*

$$\mathbb{E} \left[\sup_{t \in B} |\partial\varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^\alpha \right] < +\infty.$$

Then, as $n \rightarrow \infty$,

$$\sqrt{n} (\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \delta_0. \tag{3.7}$$

If for all $t \in [a; b]$,

$$M''(t) \geq \frac{1}{2}, \tag{A7}$$

then for all $n \geq 0$,

$$\mathbb{E} \left[\left(\widehat{\theta}_n - \theta \right)^2 \right] \leq \left(\widehat{\theta}_0 - \theta \right)^2 \frac{\exp \left(C_1 \pi^2 / 6 \right)}{n + 1}, \tag{3.8}$$

where

$$C_1 = 4\mathbb{E} \left[\sup_{t \in [a; b]} \left| \partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon) \right|^4 \right]. \tag{3.9}$$

Proof. Proofs are in Section 6.

Remark 1. One can observe that

$$M''(\theta) = 2 \int_0^1 \left[\partial \varphi_\theta^{-1} \circ \varphi_\theta \circ F^{-1}(x) \right]^2 dx = 2\mathbb{E} \left[\left(\partial \varphi_\theta^{-1}(X) \right)^2 \right].$$

Hence $M''(\theta) > 0$ holds in the general case. Moreover, replacing M by λM where λ is real and positive does not change any results. Then, the condition $M''(t) \geq 1/2$ can be verified with little modifications.

Remark 2. If one replaces the algorithm (3.3) by its “excited” version

$$\widetilde{\theta}_{n+1} = \pi_{[a; b]} \left(\widetilde{\theta}_n - \gamma_{n+1} \widetilde{T}_{n+1} \right), \tag{3.10}$$

where the initial deterministic value $\widetilde{\theta}_0 \in [a; b]$ and

$$\widetilde{T}_{n+1} = -2\partial \varphi_{\widetilde{\theta}_n}^{-1} \left(X_{n+1} \right) \left(\varepsilon_{n+1} - \varphi_{\widetilde{\theta}_n}^{-1} \left(X_{n+1} \right) \right) + V_{n+1} \tag{3.11}$$

with (V_n) a sequence of independent and identically distributed simulated random variables with mean 0 and variance $\sigma^2 > 0$, then Theorem 1 and Theorem 2 are still true for $\widetilde{\theta}_n$ with (3.7) replaced by

$$\sqrt{n} \left(\widetilde{\theta}_n - \theta \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2}{2M''(\theta) - 1} \right). \tag{3.12}$$

4. Estimation of the Density

In this section, we suppose that the random variable ε has a density f and focus on the non-parametric estimation of this density. A natural way to estimate f is to consider the recursive Parzen-Rosenblatt estimator

$$\widetilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K \left(\frac{x - \varepsilon_i}{h_i} \right), \tag{4.1}$$

where K is a standard kernel function. While \widetilde{f}_n is a good approximation of f for large values of n , it may not be for small values of n . To improve matters,

we consider the prior estimation of θ to construct a Parzen-Rosenblatt estimator of f which is of length $2n$. We need more assumptions for this. Let ∂ be the differential operator with respect to t , and d the differential operator with respect to x .

f is bounded, twice continuously differentiable on I_1 , with bounded derivatives. (AD1)

For all $t \in \Theta$, φ_t is three times continuously differentiable on I_1 . (AD2)

φ_θ^{-1} is three times continuously differentiable on I_2 , with bounded derivatives. (AD3)

$d\varphi, d^2\varphi, d^3\varphi$ are bounded. (AD4)

Denote by K a positive kernel which is a symmetric, integrable and bounded function, such that

$$\int_{\mathbb{R}} K(u)du = 1, \quad \lim_{|x| \rightarrow +\infty} |x| K(x) = 0, \quad \text{and} \quad \int_{\mathbb{R}} u^2 K(u)du < +\infty.$$

Then consider the recursive estimate

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{x - Z_i(\widehat{\theta}_{i-1})}{h_i}\right), \quad (4.2)$$

where $\widehat{\theta}_{i-1}$ is given by (3.3), and where the bandwidths h_n are positive, decrease to zero, and such that nh_n tends to infinity when n goes to infinity. For simplicity, we make use of $h_n = 1/n^\alpha$ with $0 < \alpha < 1$.

Theorem 3. *Assume (A1) to (A5) with $\theta \in]a; b[$ where $a < b$ and (AD1) to (AD4). Then for all $x \in I_1$,*

$$\widehat{f}_n(x) \xrightarrow{n \rightarrow \infty} f(x) \quad \text{a.s..} \quad (4.3)$$

It follows from Theorem 3 that for small values of n , the averaged estimator

$$\bar{f}_n = \frac{1}{2} (\widetilde{f}_n + \widehat{f}_n),$$

where \widetilde{f}_n and \widehat{f}_n are given by (4.1) and (4.2), performs better than \widetilde{f}_n or \widehat{f}_n .

For the convergence in quadratic mean of $\widehat{f}_n(x)$ to $f(x)$, we need another assumption.

φ is twice continuously differentiable on $\Theta \times I_1$ and $\partial\varphi_t(x)$, $\partial d\varphi_t(x)$ are bounded with respect to t . (AD5)

Theorem 4. Assume (A1) to (A7) with $\theta \in]a; b[$ where $a < b$ and (AD1) to (AD5). Then, for all $x \in I_1$,

$$\mathbb{E} \left[\left| \widehat{f}_n(x) - f(x) \right|^2 \right] \xrightarrow{n \rightarrow \infty} 0. \tag{4.4}$$

Proof. Proofs are postponed in Section 7.

5. Simulations

This section is devoted to the numerical illustration of the asymptotic properties of our estimator $\widehat{\theta}_n$ defined by (3.3). For the model (1.1) when φ_θ is not invertible with respect to θ , it is not possible to use a direct expression for the estimator and our procedure is useful for estimating θ . We focus on two transformations that are used in econometrics : the Box-Cox transformation φ_t^1 and the arcsinh transformation φ_t^2 . Here, for all $x \in \mathbb{R}_*^+$,

$$\varphi_t^1(x) = \begin{cases} \frac{x^t - 1}{t} & \text{if } t \neq 0, \\ \log(x) & \text{if } t = 0 \end{cases} \tag{5.1}$$

and, for all $x \in \mathbb{R}$,

$$\varphi_t^2(x) = \begin{cases} \frac{1}{t} \sinh^{-1}(tx) & \text{if } t \neq 0, \\ x & \text{if } t = 0. \end{cases} \tag{5.2}$$

Throughout this section, we suppose that $\theta > 0$, and take $\theta \in]a; b[$ with $a = 1/10$ and $b = 2$. Then, the Box-Cox transform φ_t^1 is invertible from $]1; +\infty[$ to \mathbb{R}_*^+ and the arcsinh transformation is invertible from \mathbb{R} to \mathbb{R} , with

$$\forall x \in \mathbb{R}_*^+, \quad (\varphi_t^1)^{-1}(x) = (1 + tx)^{1/t}, \tag{5.3}$$

$$\forall x \in \mathbb{R}, \quad (\varphi_t^2)^{-1}(x) = \frac{1}{t} \sinh(tx). \tag{5.4}$$

Then, for all $t \in [a; b]$, $(\varphi_t^1)^{-1}(x)$ and $(\varphi_t^2)^{-1}(x)$

$$\forall x \in \mathbb{R}_*^+, \quad \partial (\varphi_t^1)^{-1}(x) = \frac{1}{t} \left(\frac{x}{1 + tx} - \frac{1}{t} \log(1 + tx) \right) (1 + tx)^{1/t}, \tag{5.5}$$

$$\forall x \in \mathbb{R}, \quad \partial (\varphi_t^2)^{-1}(x) = -\frac{1}{t} \left(\frac{1}{t} \sinh(tx) - x \cosh(tx) \right). \tag{5.6}$$

Denote by M^1 , respectively M^2 , the function M given by (2.2) associated with φ_t^1 and φ_t^2 . For the simulations, we chose $\theta = 1$. The functions M^1 and M^2 are

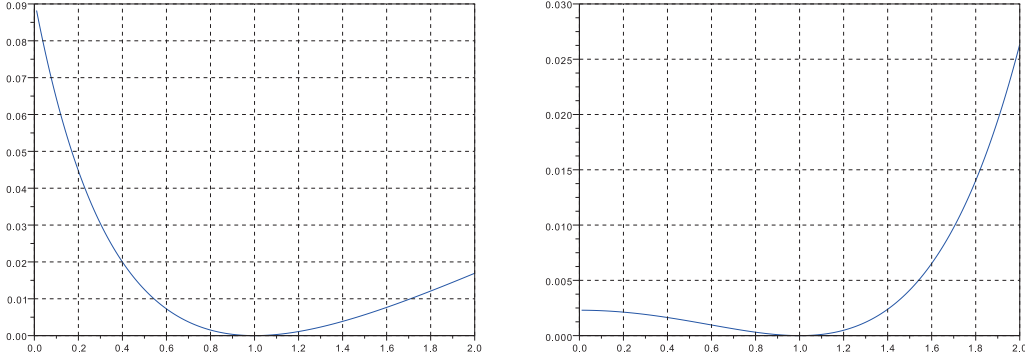


Figure 1. The functions M^1 and M^2

represented in Figure 1. One can see that θ is effectively a global minimum of M^1 and M^2 . For the estimation of θ in both models, we took (ε_n^1) as independent uniform on $[1; 2]$ and (ε_n^2) independent uniform on $[0; 1]$. We simulated random variables X_n^1 and X_n^2 according to $X_n^i = \varphi_\theta^i(\varepsilon_n^i)$, for $i = 1, 2$. Then, for $i = 1, 2$ and for the choice of step $\gamma_n = 1/n$, we computed

$$\widehat{\theta}_{n+1}^i = \pi_{[a;b]} \left(\widehat{\theta}_n^i - \gamma_n T_{n+1}^i \right),$$

where

$$T_{n+1}^i = -2\partial \left(\varphi_{\widehat{\theta}_n^i}^i \right)^{-1} \left(X_{n+1}^i \right) \left(\varepsilon_{n+1}^i - \left(\varphi_{\widehat{\theta}_n^i}^i \right)^{-1} \left(X_{n+1}^i \right) \right),$$

and with $(\varphi_{\widehat{\theta}_n^i}^i)^{-1}$ given by (5.3) and (5.4) and $\partial(\varphi_{\widehat{\theta}_n^i}^i)^{-1}$ given by (5.5) and (5.6).

The values of $\widehat{\theta}_n^i$ were computed until $n = 1,000$. We represent on the left-hand side (respectively on the right-hand side) of Figure 2 the difference between $\widehat{\theta}_{1,000}^1$ and θ (respectively $\widehat{\theta}_{1,000}^2$ and θ) for $1 \leq n \leq 1,000$. In particular, we found $|\widehat{\theta}_{1,000}^1 - \theta| = 0.00239$ and $|\widehat{\theta}_{1,000}^2 - \theta| = 0.0042$, showing that our procedure performs well for both models.

On the left-hand side of Figure 3, one has represented the degenerated asymptotic normality given by (3.7) for the data generated according for φ_θ^1 . For that, we made 200 realizations of $\sqrt{2,000} \left(\widehat{\theta}_{2,000}^1 - \theta \right)$. We also considered the excited version (3.10) of algorithm (3.3) for the first deformation φ_θ^1 ,

$$\widetilde{\theta}_{n+1}^1 = \pi_{[a;b]} \left(\widetilde{\theta}_n^1 - \gamma_n \widetilde{T}_{n+1}^1 \right),$$

with

$$\widetilde{T}_{n+1}^1 = -2\partial \left(\varphi_{\widetilde{\theta}_n^1}^1 \right)^{-1} \left(X_{n+1}^1 \right) \left(\varepsilon_{n+1}^1 - \left(\varphi_{\widetilde{\theta}_n^1}^1 \right)^{-1} \left(X_{n+1}^1 \right) \right) + V_{n+1},$$

where V_n were independent $\mathcal{N}(0, 1/2)$. For the degenerated asymptotic normality, we made 200 realizations of $\sqrt{2,000} \left(\widetilde{\theta}_{2,000}^1 - \theta \right)$ to illustrate the asymptotic

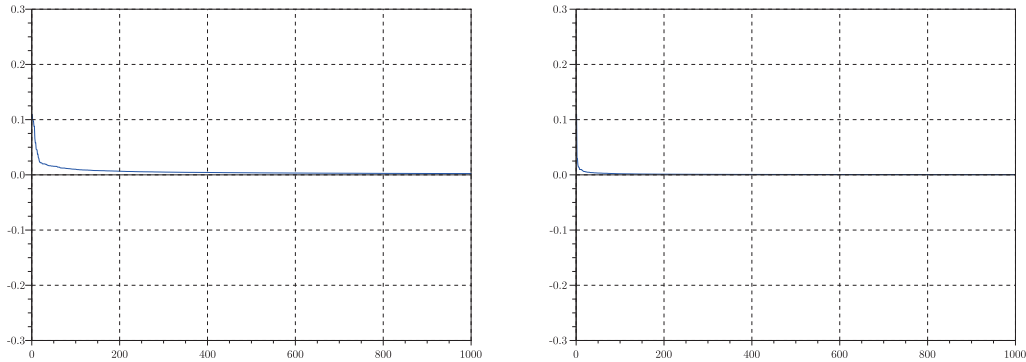


Figure 2. Difference $\hat{\theta}_n^1 - \theta$ and $\hat{\theta}_n^2 - \theta$.

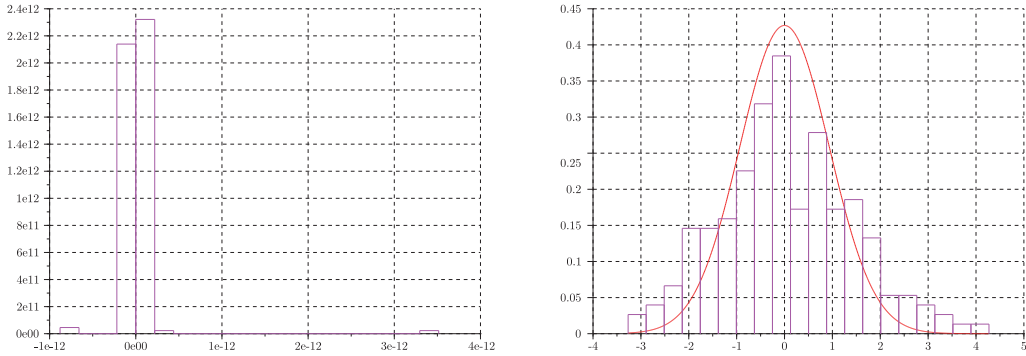


Figure 3. Asymptotic normalities of $\sqrt{n}(\hat{\theta}_n^1 - \theta)$ and $\sqrt{n}(\tilde{\theta}_n^1 - \theta)$.

normality given by (3.12). This numerical result is represented on the right-hand side of Figure 3. In these cases, to verify the condition $M''(\theta) > 1/2$, one has to modify the criterion M as described in Remark 1, with $\lambda = 10$.

6. Proofs of the Parametric Results

6.1. Proof of Lemma 1

From (A4), for all compact B in Θ ,

$$\mathbb{E} \left[\sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^2 \right] < +\infty.$$

Consequently,

$$\mathbb{E} \left[\sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^2 \right] = \int_0^1 \sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))|^2 dx < +\infty. \quad (6.1)$$

Now, it follows from (A2) that for all $x \in I_2$,

$$\begin{aligned} & \partial \left[(F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))^2 \right] \\ &= -2\partial\varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) \end{aligned} \quad (6.2)$$

is a continuous function with respect to t . In addition, if B is a compact set containing θ , from (A2) and the Mean Value Theorem there exists a constant $C_B > 0$ such that

$$\sup_{t \in B} |F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))| \leq C_B \sup_{t \in B} |\partial\varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))|. \quad (6.3)$$

Hence, we deduce from (6.2) and the previous inequality that

$$\sup_{t \in B} \left| \partial \left[(F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))^2 \right] \right| \leq 2C_B \sup_{t \in B} |\partial\varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))|^2$$

which implies by (6.1) that

$$\sup_{t \in B} \left| \partial \left[(F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))^2 \right] \right|$$

is integrable with respect to x . Finally, M is continuously differentiable on Θ and for all $t \in \Theta$,

$$M'(t) = \int_0^1 -2\partial\varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx.$$

6.2. Proof of Lemma 2

From (A6),

$$-2\partial\varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) \quad (6.4)$$

is continuously differentiable with respect to t . In addition, we have

$$\begin{aligned} & \partial \left[\partial\varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) \right] \\ &= -\left[\partial\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) \right]^2 + \partial^2\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)). \end{aligned}$$

It follows from (6.3) that for every compact set B containing t and θ ,

$$\begin{aligned} & \left| \partial^2\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) \right| \\ & \leq C_B \sup_{t \in B} \left| \partial^2\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) \right| \sup_{t \in B} \left| \partial\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) \right|. \end{aligned}$$

Then, from (A6), (6.1) and the Cauchy-Schwartz inequality,

$$\partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))$$

is integrable with respect to x . Hence, we have

$$\int_0^1 \sup_{t \in B} |\partial [\partial \varphi_t^{-1} (\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))]| dx < +\infty,$$

which enables us to conclude that M is twice continuously differentiable on Θ and for all $t \in \Theta$,

$$\begin{aligned} M''(t) &= 2 \int_0^1 [\partial \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)]^2 dx \\ &\quad - 2 \int_0^1 \partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx. \end{aligned}$$

6.3. Proof of Theorem 1.

Denote by \mathcal{F}_n the σ -algebra of the events occurring up to time n , $\mathcal{F}_n = \sigma(\varepsilon_0, \dots, \varepsilon_n)$. We calculate the first two conditional moments of T_n given by (3.4). One has

$$\begin{aligned} \mathbb{E}[T_{n+1} | \mathcal{F}_n] &= -2\mathbb{E} \left[\partial \varphi_{\hat{\theta}_n}^{-1} (X_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\hat{\theta}_n}^{-1} (X_{n+1}) \right) | \mathcal{F}_n \right], \\ &= -2\mathbb{E} \left[\partial \varphi_{\hat{\theta}_n}^{-1} \circ \varphi_\theta (\varepsilon_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\hat{\theta}_n}^{-1} \circ \varphi_\theta (\varepsilon_{n+1}) \right) | \mathcal{F}_n \right]. \end{aligned}$$

Moreover, as ε_{n+1} is independent of \mathcal{F}_n and $\hat{\theta}_n \in \mathcal{F}_n$, one can deduce from (3.1) that

$$\begin{aligned} &-2\mathbb{E} \left[\partial \varphi_{\hat{\theta}_n}^{-1} \circ \varphi_\theta (\varepsilon_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\hat{\theta}_n}^{-1} \circ \varphi_\theta (\varepsilon_{n+1}) \right) | \mathcal{F}_n \right] \\ &= -2 \int_0^1 \partial \varphi_{\hat{\theta}_n}^{-1} \circ \varphi_\theta \circ F^{-1}(x) \left(F^{-1}(x) - \varphi_{\hat{\theta}_n}^{-1} \circ \varphi_\theta \circ F^{-1}(x) \right) dx \\ &= M'(\hat{\theta}_n) \quad \text{a.s.,} \end{aligned}$$

which immediately leads to

$$\mathbb{E}[T_{n+1} | \mathcal{F}_n] = M'(\hat{\theta}_n) \quad \text{a.s..} \tag{6.5}$$

As well,

$$\begin{aligned} \mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] &= 4\mathbb{E} \left[\partial \varphi_{\hat{\theta}_n}^{-1} (X_{n+1})^2 \left(\varepsilon_{n+1} - \varphi_{\hat{\theta}_n}^{-1} (X_{n+1}) \right)^2 | \mathcal{F}_n \right], \\ &= 4\mathbb{E} \left[\partial \varphi_{\hat{\theta}_n}^{-1} (X_{n+1})^2 \left(\varphi_\theta^{-1} (X_{n+1}) - \varphi_{\hat{\theta}_n}^{-1} (X_{n+1}) \right)^2 | \mathcal{F}_n \right]. \end{aligned} \tag{6.6}$$

It follows from the Mean Value Theorem that

$$|\varphi_{\theta}^{-1}(X_{n+1}) - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1})| \leq \sup_{t \in [a; b]} |\partial \varphi_t^{-1}(X_{n+1})| \times |\widehat{\theta}_n - \theta|. \quad (6.7)$$

Consequently, (6.6) and (6.7) lead to

$$\mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] \leq 4 (\widehat{\theta}_n - \theta)^2 \mathbb{E} \left[\sup_{t \in [a; b]} |\partial \varphi_t^{-1}(X)|^4 \right]. \quad (6.8)$$

Hence, there exists a positive constant C_1 given by (3.9) such that

$$\mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] \leq C_1 (\widehat{\theta}_n - \theta)^2 \quad \text{a.s.} \quad (6.9)$$

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

$$\begin{aligned} V_{n+1} &= (\widehat{\theta}_{n+1} - \theta)^2 \\ &= \left(\pi_{[a; b]}(\widehat{\theta}_n - \gamma_{n+1} T_{n+1}) - \theta \right)^2 \\ &= \left(\pi_{[a; b]}(\widehat{\theta}_n - \gamma_{n+1} T_{n+1}) - \pi_{[a; b]}(\theta) \right)^2 \end{aligned}$$

as we have assumed that θ belongs to $]a; b[$. Since $\pi_{[a; b]}$ 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 (6.5) together with (6.9) that

$$\mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n (1 + C_1 \gamma_{n+1}^2) - 2\gamma_{n+1} (\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) \quad \text{a.s.} \quad (6.10)$$

In addition, as $\widehat{\theta}_n \in [a; b]$, (A5) implies that $(\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) > 0$. Then, we deduce from (6.10) together with Robbins-Siegmund Theorem, see Duflo Duflo (1997) page 18, that the sequence (V_n) converges a.s. to a finite random variable V and

$$\sum_{n=1}^{\infty} \gamma_{n+1} (\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) < +\infty \quad \text{a.s.} \quad (6.11)$$

Assume by contradiction that $V \neq 0$ a.s. Then, one can find two constants c and d such that $0 < c < d < 2 \max(|a|, |b|)$, and for n large enough, the event

$\{c < |\hat{\theta}_n - \theta| < d\}$ is not negligible. However, on this annulus, one can also find some constant $e > 0$ such that $(\hat{\theta}_n - \theta)M'(\hat{\theta}_n) \geq e$ which, by (6.11), implies that

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

This contradicts (3.2).

6.4. Proof of Theorem 2.

Our goal is to apply Theorem 2.1 of Kushner and Yin Kushner and Yin (2003) page 330. First of all, as $\gamma_n = 1/n$, the conditions on the decreasing step is satisfied. Moreover, $\hat{\theta}_n$ converges almost surely to θ . Consequently, the local assumptions of Theorem 2.1 of Kushner and Yin (2003) are satisfied. In addition, it follows from (6.5) that $\mathbb{E}[T_{n+1}|\mathcal{F}_n] = M'(\hat{\theta}_n)$ a.s. and M is two times continuously differentiable. Hence, $M(\theta) = 0$, $M'(\theta) = 0$ and $M''(\theta) > 1/2$. It follows from (6.9) and the almost sure convergence of $\hat{\theta}_n$ to θ that

$$\lim_{n \rightarrow \infty} \mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = 0 \quad \text{a.s..}$$

Finally, Theorem 4.1 of Kushner and Yin (2003) page 341 ensures that the sequence (W_n) given by

$$W_n = \sqrt{n}(\hat{\theta}_n - \theta)$$

is tight. Then, from Theorem 2.1 of Kushner and Yin (2003),

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \delta_0.$$

Taking expectation on both sides of (6.10) leads, for all $n \geq 0$, to

$$v_{n+1} \leq v_n(1 + C_1\gamma_{n+1}^2) - 2\gamma_{n+1}\mathbb{E}\left[(\hat{\theta}_n - \theta)M'(\hat{\theta}_n)\right], \tag{6.12}$$

where

$$v_n = \mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right].$$

In addition, as $M'(\theta) = 0$,

$$M'(\hat{\theta}_n) = (\hat{\theta}_n - \theta) \int_0^1 M''(\theta + x(\hat{\theta}_n - \theta))dx \quad \text{a.s..} \tag{6.13}$$

From (6.12) and (6.13),

$$v_{n+1} \leq v_n(1 + C_1\gamma_{n+1}^2) - 2\gamma_{n+1}\mathbb{E}\left[(\hat{\theta}_n - \theta)^2 \int_0^1 M''(\theta + x(\hat{\theta}_n - \theta))dx\right]. \tag{6.14}$$

As $\theta \in]a; b[$ and $\widehat{\theta}_n \in [a; b]$, $\theta + x(\widehat{\theta}_n - \theta) \in [a; b]$ for all $x \in [0; 1]$. Then, since $M''(t) \geq 1/2$ for all $t \in [a; b]$, we can write

$$\int_0^1 M''(\theta + x(\widehat{\theta}_n - \theta)) dx \geq \frac{1}{2}.$$

From (6.14), for all $n \geq 0$,

$$v_{n+1} \leq v_n(1 + C_1\gamma_{n+1}^2 - \gamma_{n+1}). \quad (6.15)$$

Moreover, the inequality $1 - x \leq \exp(-x)$ implies that

$$v_{n+1} \leq v_n \exp(C_1\gamma_{n+1}^2 - \gamma_{n+1}). \quad (6.16)$$

An immediate recurrence in (6.16) leads to

$$\begin{aligned} v_n &\leq v_0 \prod_{k=1}^n \exp(C_1\gamma_k^2 - \gamma_k) \\ &\leq v_0 \exp\left(C_1 \sum_{k=1}^n \gamma_k^2 - \sum_{k=1}^n \gamma_k\right) \\ &\leq v_0 \exp\left(C_1 \sum_{k=1}^{+\infty} \gamma_k^2 - \sum_{k=1}^n \gamma_k\right). \end{aligned} \quad (6.17)$$

As $\gamma_k = 1/k$, it follows from (6.17) with

$$\begin{aligned} \sum_{k=1}^{+\infty} \gamma_k^2 &= \frac{\pi^2}{6}, \\ \sum_{k=1}^n \gamma_k &\geq \log(n+1) \end{aligned}$$

that, for all $n \geq 0$,

$$v_n \leq v_0 \frac{\exp(C_1\pi^2/6)}{n+1}.$$

This finishes the proof of Theorem 2.

7. Proofs of the Nonparametric Results

With f the density of ε , let f^t be the density of $Z(t)$. As the distribution of $Z(t)$ is $F \circ \varphi_\theta^{-1} \circ \varphi_t$, for all $x \in I_1$,

$$f^t(x) = f(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x).$$

Here $f^\theta = f$. We have

$$\begin{aligned} f^t(x) &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x) \\ &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d\varphi_t(x) d[\varphi_\theta^{-1}](\varphi_t(x)). \end{aligned}$$

Hence, (AD1), (AD2) and (AD3) imply that f^t is twice continuously differentiable with respect to x . Moreover, for all $x \in I_1$,

$$\begin{aligned} df^t(x) &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d^2[\varphi_\theta^{-1} \circ \varphi_t](x) + f'(\varphi_\theta^{-1} \circ \varphi_t(x)) (d[\varphi_\theta^{-1} \circ \varphi_t](x))^2, \\ d^2 f^t(x) &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d^3[\varphi_\theta^{-1} \circ \varphi_t](x) \\ &\quad + 3f'(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x) d^2[\varphi_\theta^{-1} \circ \varphi_t](x) \\ &\quad + f''(\varphi_\theta^{-1} \circ \varphi_t(x)) (d[\varphi_\theta^{-1} \circ \varphi_t](x))^3. \end{aligned}$$

Hence, from (AD1) to (AD4), $f^t(x)$, $df^t(x)$ and $d^2 f^t(x)$ are bounded on $\Theta \times I_1$. Now (AD5) implies that $f^t(x)$ is also continuously differentiable with respect to (t, x) , and we have for all $t \in \Theta$ and for all $x \in I_1$,

$$\begin{aligned} \partial f^t(x) &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) \partial d[\varphi_\theta^{-1} \circ \varphi_t](x) + f'(\varphi_\theta^{-1} \circ \varphi_t(x)) \\ &\quad d[\varphi_\theta^{-1} \circ \varphi_t](x) \partial[\varphi_\theta^{-1} \circ \varphi_t](x), \end{aligned}$$

where

$$\begin{aligned} \partial[\varphi_\theta^{-1} \circ \varphi_t](x) &= \partial\varphi_t(x) d[\varphi_\theta^{-1}](\varphi_t(x)), \\ \partial d[\varphi_\theta^{-1} \circ \varphi_t](x) &= \partial d\varphi_t(x) d[\varphi_\theta^{-1}](\varphi_t(x)) + \partial\varphi_t(x) d\varphi_t(x) d^2[\varphi_\theta^{-1}](\varphi_t(x)). \end{aligned}$$

Hence, under (AD4) and (AD5)

$$\sup_{t \in \Theta} |\partial f^t(x)| < +\infty. \tag{7.1}$$

7.1. Proof of Theorem 3

With $\mathcal{F}_n = \sigma\{\varepsilon_0, \dots, \varepsilon_n\}$, $\widehat{\theta}_{n-1}$ is measurable with respect to \mathcal{F}_{n-1} . Let for all $x \in I_1$,

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

Then, we have the decomposition, for all $x \in I_1$, $n\widehat{f}_n(x) = M_n(x) + N_n(x)$, where

$$M_n(x) = \sum_{i=1}^n \mathbb{E}[W_i(x) | \mathcal{F}_{i-1}], \tag{7.2}$$

$$N_n(x) = \sum_{i=1}^n (W_i(x) - \mathbb{E}[W_i(x) | \mathcal{F}_{i-1}]). \tag{7.3}$$

For a fixed $\widehat{\theta}_{n-1}$, $f^{\widehat{\theta}_{n-1}}$ is the density of $Z_n(\widehat{\theta}_{n-1})$, so, with $v = (x - u)/h_i$, we have

$$\begin{aligned} \mathbb{E}[W_i(x)|\mathcal{F}_{i-1}] &= \int_{\mathbb{R}} \frac{1}{h_i} K\left(\frac{x-u}{h_i}\right) f^{\widehat{\theta}_{i-1}}(u) du \\ &= \int_{\mathbb{R}} K(v) f^{\widehat{\theta}_{i-1}}(x - h_i v) dv. \end{aligned}$$

Hence,

$$\mathbb{E}[W_i(x)|\mathcal{F}_{i-1}] - f^{\widehat{\theta}_{i-1}}(x) = \int_{\mathbb{R}} \left(f^{\widehat{\theta}_{i-1}}(x - v h_i) - f^{\widehat{\theta}_{i-1}}(x) \right) K(v) dv.$$

As f^t is twice continuously differentiable, for all $t \in \Theta$, there exists a real $z_i = x - v h_i y$, with $0 < y < 1$, such that

$$f^t(x - v h_i) - f^t(x) = -v h_i df^t(x) + \frac{(v h_i)^2}{2} d^2 f^t(z_i). \tag{7.4}$$

Using the parity of K and preliminary remarks on $d^2 f^t$, we obtain that

$$\int_{\mathbb{R}} (f^t(x - v h_i) - f^t(x)) K(v) dv = \int_{\mathbb{R}} \frac{(v h_i)^2}{2} d^2 f^t(z_i) K(v) dv$$

which implies that

$$\sup_{t \in \Theta} \left| \int_{\mathbb{R}} (f^t(x - v h_i) - f^t(x)) K(v) dv \right| \leq \frac{h_i^2}{2} \sup_{t \in \Theta, z \in I_1} |d^2 f^t(z)| \int_{\mathbb{R}} v^2 K(v) dv.$$

Consequently, there exists $C_2 > 0$ such that

$$\left| \mathbb{E}[W_i(x)|\mathcal{F}_{i-1}] - f^{\widehat{\theta}_{i-1}}(x) \right| \leq C_2 h_i^2. \tag{7.5}$$

Since f^t is a continuous function with respect to t , and $\widehat{\theta}_n$ converges to θ almost surely, we have for all $x \in I_1$,

$$f^{\widehat{\theta}_{n-1}}(x) \xrightarrow{i \rightarrow \infty} f(x) \quad \text{a.s.} \tag{7.6}$$

Consequently, Cesaro's Theorem and (7.5) imply that

$$\frac{1}{n} M_n(x) \xrightarrow{n \rightarrow \infty} f(x) \quad \text{a.s.} \tag{7.7}$$

Since K is bounded, $(N_n(x))$ is a square integrable martingale whose predictable quadratic variation is given by

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

We have

$$\mathbb{E} [W_i^2(x)|\mathcal{F}_{i-1}] = \frac{1}{h_i} \int K^2(v) f^{\widehat{\theta}_{i-1}}(x - h_i v) dv.$$

However, (7.4), together with the regularity of $f^t(x)$ and the parity of K , implies that

$$\sup_{t \in \Theta} \left| \int_{\mathbb{R}} \frac{1}{h_i} (f^t(x - v h_i) - f^t(x)) K^2(v) dv \right| \leq \frac{h_i}{2} \sup_{t \in \Theta, z \in I_1} |d^2 f^t(z)| \int_{\mathbb{R}} v^2 K^2(v) dv.$$

Consequently, there exists $C_3 > 0$ such that

$$\left| \mathbb{E} [W_i^2(x)|\mathcal{F}_{i-1}] - \frac{\nu^2}{h_i} f^{\widehat{\theta}_{i-1}}(x) \right| \leq C_3 h_i, \tag{7.8}$$

where $\nu^2 = \int_{\mathbb{R}} K^2(u) du$. It follows from (7.6) and the Toeplitz Lemma that

$$\lim_{n \rightarrow \infty} \frac{1}{\sum_{i=1}^n h_i^{-1}} \sum_{i=1}^n \frac{1}{h_i} f^{\widehat{\theta}_{i-1}}(x) = f(x) \quad \text{a.s..}$$

From the elementary equivalence

$$\sum_{i=1}^n \frac{1}{h_i} \sim \frac{n^{1+\alpha}}{\alpha + 1},$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\alpha}} \sum_{i=1}^n \frac{\nu^2}{h_i} f^{\widehat{\theta}_{i-1}}(x) = \frac{\nu^2}{\alpha + 1} f(x) \quad \text{a.s..}$$

Now (7.8) leads to

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\alpha}} \sum_{i=1}^n \mathbb{E} [W_i^2(x)|\mathcal{F}_{i-1}] = \frac{\nu^2}{\alpha + 1} f(x) \quad \text{a.s.}, \tag{7.9}$$

while (7.5), together with (7.6) and Cesaro’s Theorem, implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}^2 [W_i(x)|\mathcal{F}_{i-1}] = f^2(x). \tag{7.10}$$

Then, as $\alpha > 0$, we can conclude from (7.9) and (7.10) that

$$\lim_{n \rightarrow \infty} \frac{\langle N(x) \rangle_n}{n^{1+\alpha}} = \frac{\nu^2}{\alpha + 1} f(x) \quad \text{a.s..}$$

We obtain from the strong law of large numbers for martingales given e.g. by Theorem 1.3.15 of Duflo (1997) that for any $\gamma > 0$, $(N_n(x))^2 = o\left(n^{1+\alpha} (\log(n))^{1+\gamma}\right)$ a.s., which ensures that for all $x \in I_1$,

$$\frac{1}{n} N_n(x) \xrightarrow{n \rightarrow \infty} 0 \quad \text{a.s..} \tag{7.11}$$

Combining (7.7) and (7.11), one obtains that for all $x \in I_1$,

$$\widehat{f}_n(x) \xrightarrow{n \rightarrow \infty} f(x) \quad \text{a.s.} \quad (7.12)$$

ending the proof of Theorem 3.

7.2. Proof of Theorem 4

We show that, for all $x \in I_1$,

$$\mathbb{E} \left[\left| \widehat{f}_n(x) - f(x) \right|^2 \right] \xrightarrow{n \rightarrow \infty} 0.$$

It follows from the bias-variance decomposition that

$$\mathbb{E} \left[\left| \widehat{f}_n(x) - f(x) \right|^2 \right] = B_n(x) + V_n(x), \quad (7.13)$$

where

$$B_n(x) = \left| \mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) \right|^2, \quad (7.14)$$

$$V_n(x) = \mathbb{E} \left[\left| \widehat{f}_n(x) - \mathbb{E} \left[\widehat{f}_n(x) \right] \right|^2 \right]. \quad (7.15)$$

We can write

$$\begin{aligned} \mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} [W_i(x) - f(x)] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\mathbb{E} [W_i(x) | \mathcal{F}_{i-1}] - f(x)]. \end{aligned}$$

In addition, (7.5) implies that

$$\mathbb{E} \left[\left| \mathbb{E} [W_n(x) | \mathcal{F}_{n-1}] - f^{\widehat{\theta}_{n-1}}(x) \right| \right] \xrightarrow{n \rightarrow \infty} 0. \quad (7.16)$$

It follows from the boundedness of $f^{\widehat{\theta}_{n-1}}(x)$ and (7.6), together with the Dominated Convergence Theorem, that

$$\mathbb{E} \left[\left| f^{\widehat{\theta}_{n-1}}(x) - f(x) \right| \right] \xrightarrow{n \rightarrow \infty} 0. \quad (7.17)$$

Hence, we deduce from (7.16) and (7.17) that

$$\mathbb{E} [\mathbb{E} [W_i(x) | \mathcal{F}_{n-1}] - f(x)] \xrightarrow{n \rightarrow \infty} 0,$$

which implies by Cesaro's Theorem that

$$\left| \mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) \right| \xrightarrow{n \rightarrow \infty} 0$$

leading to

$$B_n(x) \xrightarrow{n \rightarrow \infty} 0. \tag{7.18}$$

Now we focus on the variance term $V_n(x)$. For all $1 \leq i \leq n$ and all $x \in I_1$, let

$$U_i(x) = W_i(x) - \mathbb{E}[W_i(x)]. \tag{7.19}$$

We have the decomposition

$$V_n(x) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}[U_i(x)^2] + \frac{2}{n^2} \sum_{i=1, i < j}^n \mathbb{E}[U_i(x)U_j(x)]. \tag{7.20}$$

If $i < j$, $\mathbb{E}[U_i(x)U_j(x)|\mathcal{F}_{j-1}] = U_i(x)\mathbb{E}[U_j(x)|\mathcal{F}_{j-1}]$. In addition, (7.5) implies that

$$\left| \mathbb{E}[U_j(x)|\mathcal{F}_{j-1}] - f^{\widehat{\theta}_{j-1}}(x) + \mathbb{E}[f^{\widehat{\theta}_{j-1}}(x)] \right| \leq 2C_2h_j^2.$$

Hence,

$$\begin{aligned} -2C_2h_j^2|U_i(x)| &\leq \mathbb{E}[U_i(x)U_j(x)|\mathcal{F}_{j-1}] - U_i(x)f^{\widehat{\theta}_{j-1}}(x) + U_i(x)\mathbb{E}[f^{\widehat{\theta}_{j-1}}(x)] \\ &\leq 2C_2h_j^2|U_i(x)|. \end{aligned}$$

Thus, taking expectation in the previous inequality leads to

$$\begin{aligned} -2C_2h_j^2\mathbb{E}[|U_i(x)|] &\leq \mathbb{E}[U_i(x)U_j(x)] - \mathbb{E}[U_i(x)f^{\widehat{\theta}_{j-1}}(x)] + \mathbb{E}[U_i(x)]\mathbb{E}[f^{\widehat{\theta}_{j-1}}(x)] \\ &\leq 2C_2h_j^2\mathbb{E}[|U_i(x)|]. \end{aligned}$$

Finally, we obtain that

$$\left| \mathbb{E}[U_i(x)U_j(x)] \right| \leq \left| \mathbb{E}[U_i(x)f^{\widehat{\theta}_{j-1}}(x)] - \mathbb{E}[U_i(x)]\mathbb{E}[f^{\widehat{\theta}_{j-1}}(x)] \right| + 2C_2h_j^2\mathbb{E}[|U_i(x)|]. \tag{7.21}$$

Since

$$\begin{aligned} &\mathbb{E}[U_i(x)f^{\widehat{\theta}_{j-1}}(x)] - \mathbb{E}[U_i(x)]\mathbb{E}[f^{\widehat{\theta}_{j-1}}(x)] \\ &= \mathbb{E}[U_i(x)(f^{\widehat{\theta}_{j-1}}(x) - f(x))] + (f(x) - \mathbb{E}[f^{\widehat{\theta}_{j-1}}(x)])\mathbb{E}[U_i(x)], \end{aligned} \tag{7.22}$$

(7.21), (7.22), together with the Cauchy-Schwartz inequality implies that

$$\mathbb{E}[|U_i(x)U_j(x)|] \leq 2\sqrt{\mathbb{E}[U_i(x)^2]} \left(\sqrt{\mathbb{E}\left[\left(f^{\widehat{\theta}_{j-1}}(x) - f(x)\right)^2\right]} + C_2h_j^2 \right). \tag{7.23}$$

The definition (7.19) of $U_i(x)$ also leads to $\mathbb{E}[U_i^2(x)] \leq \mathbb{E}[W_i^2(x)]$ which implies, by (7.8), that

$$\mathbb{E}[U_i^2(x)] \leq \frac{\nu^2}{h_i} \mathbb{E}[f^{\widehat{\theta}_{i-1}}(x)] + C_3h_i. \tag{7.24}$$

Let C be a constant which does not depend on n . Now, (3.8) implies that, for all $n \geq 0$,

$$\mathbb{E} \left[\left(\widehat{\theta}_n - \theta \right)^2 \right] \leq \frac{C}{n}. \tag{7.25}$$

Then, using the regularity of f , we obtain that, for all $x \in I_1$,

$$|f^t(x) - f(x)| \leq \sup_{t \in \Theta} |\partial f^t(x)| |t - \theta|.$$

Hence, (7.1) and (7.25) lead to

$$\sqrt{\mathbb{E} \left[\left| f^{\widehat{\theta}_{n-1}}(x) - f(x) \right|^2 \right]} \leq \frac{C}{\sqrt{n}}. \tag{7.26}$$

In all, (7.23), (7.24), and (7.26) imply that

$$\mathbb{E} [|U_i(x)U_j(x)|] \leq 2 \left(\sqrt{\frac{\nu^2}{h_i} \mathbb{E} [f^{\widehat{\theta}_{i-1}}(x)]} + C_3 h_i \right) \left(\frac{C}{\sqrt{j}} + C_2 h_j^2 \right). \tag{7.27}$$

Using the boundedness of $f^t(x)$, we obtain that

$$\mathbb{E} [|U_i(x)U_j(x)|] \leq C \left(\frac{1}{\sqrt{j h_i}} + \frac{h_j^2}{\sqrt{h_i}} \right). \tag{7.28}$$

Moreover, if $h_n = 1/n^\alpha$,

$$\begin{aligned} \sum_{i=1, i < j}^n \frac{1}{\sqrt{j h_i}} &= \sum_{j=2}^n \frac{1}{j^{1/2}} \sum_{i=1}^{j-1} i^{\alpha/2} \leq \sum_{j=2}^n \frac{j^{\alpha/2+1}}{j^{1/2}} \leq n^{(3+\alpha)/2}, \\ \sum_{i=1, i < j}^n \frac{h_j^2}{\sqrt{h_i}} &= \sum_{j=2}^n h_j^2 \sum_{i=1}^{j-1} i^{\alpha/2} \leq \sum_{j=2}^n \frac{j^{\alpha/2+1}}{j^{2\alpha}} \leq n^{2-3\alpha/4}. \end{aligned}$$

From these calculations and from (7.28),

$$\frac{1}{n^2} \sum_{i=1, i < j}^n \mathbb{E} [|U_i(x)U_j(x)|] \leq C \left(n^{(-1+\alpha)/2} + n^{-3\alpha/4} \right), \tag{7.29}$$

which tends to 0 as n goes to infinity, as $0 < \alpha < 1$. With (7.24) and the boundness of f^t , we have

$$\frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [U_i^2(x)] \leq C \frac{1}{n^2} \sum_{i=1}^n \frac{1}{h_i} \leq C \frac{n^{\alpha+1}}{n^2} \leq C n^{-1+\alpha}, \tag{7.30}$$

which tends to 0 as n goes to infinity, as $\alpha < 1$. Hence, (7.20), (7.29), and (7.30) let us conclude that, for all $x \in I_1$,

$$V_n(x) \xrightarrow{n \rightarrow \infty} 0. \quad (7.31)$$

With (7.13), (7.18), and (7.31), we finish the proof of Theorem 4.

References

- Agullo, M., Cuesta, J. A. Lescornel, H. and Loubes, J.-M. (2014). A Parametric Registration model for warped distributions with Wasserstein's distance. *J. Multivariate Anal.*, to appear.
- Bercu, B. and Fraysse, P. (2012). A robbins-monro procedure for estimation in semiparametric regression models. *Ann. Statist.* **40**, 666-693.
- Bolstad, B. M., Irizarry, R. A., Åstrand, M. and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. (with discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 211-252.
- Chen, H. F., Lei, G. and Gao, A. J. (1988). Convergence and robustness of the robbins-monro algorithm truncated at randomly varying bounds. *Stoch. Process. Appl.* **27**, 217-231.
- Cuesta, J. A. and Matrán, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* **17**, 1264-1276.
- Devroye, L. (1988). The kernel estimate is relatively stable. *Probab. Theory Related Fields* **77**, 521-536.
- Duflo, M. (1997). *Random iterative models*, volume 34 of *Applications of Mathematics*. Springer-Verlag, Berlin.
- Gallón, S., Loubes, J.-M. and Maza, E. (2013). Statistical properties of the quantile normalization method for density curve alignment. *Math. Biosci.* **242**, 129-142.
- Gamboa, F., Loubes, J.-M. and Maza, E. (2007). Semi-parametric estimation of shifts. *Electronic J. Statist.* **1**, 616-640.
- Hall, P. (1982). Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Process. Appl.* **13**, 11-25.
- Hall, P. (1984). On the influence of extremes on the rate of convergence in the central limit theorem. *Ann. Probab.* **12**, 154-172.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press Inc., New York.
- Kushner, H. J. and Yin, G. G. (2003). Stochastic approximation and recursive algorithms and applications, volume 35 of *Applications of Mathematics*. Springer-Verlag, New York.
- Lelong, J. (2008). Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett.* **78**, 2632-2636.
- Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Ann. Statist.* **36**, 686-718.
- Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**, 223-241.

- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22**, 400-407.
- Robbins, H. and Siegmund, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. *Optimizing Methods Stat.*, 233-257.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 832-837.
- Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6**, 177-184.
- Van der Vaart, A. (2000). *Asymptotic Statistics*. Number 3. Cambridge Univ. Pr.

Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251 F-33400 Talence, France.

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

Inria Saclay - Île-de-France, Bâtiment Alan Turing, 1 rue Honoré d'Estienne d'Orves, Campus de l'École Polytechnique, 91120 Palaiseau, France.

E-mail: helene.lescornel@inria.fr

Université Paul Sabatier, Institut de Mathématiques de Toulouse, UMR CNRS 5251 F-31062 Toulouse, France.

E-mail: jean-michel.loubes@math.univ-toulouse.fr

(Received January 2013; accepted April 2014)