Journal of Stochastic Analysis

Volume 1 | Number 3 Article 4

August 2020

Nonparametric Recursive Method for Kernel-Type Function Estimators for Censored Data

Salim Bouzebda
Alliance Sorbonne Universités, Université de Technologie de Compiègne, L.M.A.C., Compiègne, France, salim.bouzebda@utc.fr

Yousri Slaoui
Univ. Poitiers, Lab. Math. et Appl., Futuroscope Chasseneuil, France, yousri.slaoui@math.univ-poitiers.fr

Journal of Stochastic Analysis
Vol. 1, No. 3 (2020) Article 4 (19 pages)
DOI: 10.31390/josa.1.3.04

NONPARAMETRIC RECURSIVE METHOD FOR KERNEL-TYPE FUNCTION ESTIMATORS FOR CENSORED DATA

SALIM BOUZEBDA* AND YOUSRI SLAOUI

Abstract. In the present paper, we study general kernel type estimators for censored data defined by the stochastic approximation algorithm. We establish a central limit theorem for the proposed estimators. We characterize the strong pointwise convergence rate for the nonparametric recursive general kernel-type estimators under some mild conditions.

1. Introduction

Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. Over years ago, [37] studied some properties of kernel density estimators introduced by [1] and [38]. Nonparametric density and regression function estimation has been the subject of intense investigation by both statisticians and probabilists for many years and this has led to the development of a large variety of methods. Kernel nonparametric function estimation methods have long attracted a great deal of attention, for good sources of references to research literature in this area along with statistical applications consult [45], [49], [18], [17], [40], [36], [28], [39], [47], [22], [19] and the references therein. There are basically no restrictions on the choice of the kernel $K(\cdot)$ in our setup, apart from satisfying classical conditions. The selection of the bandwidth, however, is more problematic. The choice of the bandwidth is crucial to obtain a good rate of consistency for of the kernel-type estimators. It has a big influence on the size of the bias. One has to find an appropriate bandwidth that produces an estimator which has a good balance between the bias and the variance of the kernel-type estimator, for more discussion refer to [31]. It is worth noticing that the bandwidth selection methods studied in the literature can be divided into three broad classes: the cross-validation techniques, the plug-in ideas and the bootstrap. Recently, some general methods based upon empirical process techniques are developed in order to prove uniform in bandwidth consistency of a class of kernel-type function estimators (density, regression, entropy and copula), we may refer to [23, 24], [5, 6], [4], [7] and [8]. Further, recursive kernel density estimators defined by stochastic approximation method have been proposed by [41], recursive kernel distribution estimators have been done by [42], recursive regression estimators

Received 2020-3-7; Accepted 2020-7-6; Communicated by the editors.
2010 *Mathematics Subject Classification.* Primary 60F10; 62G07; 62G05; 62L12.
Key words and phrases. Density estimation, stochastic approximation algorithm, asymptotic normality, regression estimation, censored data, recursive estimation.
* Corresponding author.

have been done by [43, 44] and recursive kernel-type estimators for spatial data was proposed by [9, 10, 11, 12].

This work concerns a nonparametric estimation of the recursive general kernel-type estimators for censored data defined by the stochastic approximation algorithm. To the best of our knowledge, the results presented here, respond to a problem that has not been studied systematically up to the present, which was the basic motivation of the present paper.

The problem of censoring, is frequently encountered in certain area of statistical applications. The didactique example of censoring is arguably the study of the survival times of patients to a given chronic disease in a medical follow-up study lasting up to a fixed time t. If a patient is diagnosed with the disease at time s, then the survival time will be known if and only if the patient dies before time t. If this is not the case, then the only information available is that the survival time is not less than the censoring time $t-s$. In mathematical terms, the information available to the practitioner is the triple $(T, C, \mathbf{X})$ defined in $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^d$. Here T is the variable of interest, C a censoring variable and $\mathbf{X}$ a concomitant variable. Throughout, we work with a sample $\{(T_i, C_i, \mathbf{X}_i)_{1\leq i\leq n}\}$ of independent and identically distributed replicæ of $(T, C, \mathbf{X})$, $n \geq 1$. Actually, in the right censorship model, the pairs (T_i, C_i), $1 \leq i \leq n$, are not directly observed and the corresponding information is given by

$$Z_i := \min\{T_I, C_I\} \text{ and } \delta_i := \mathbb{1}\{T_i \leq C_i\}, \quad 1 \leq i \leq n,$$

with $\mathbb{1}\{A\}$ standing for the indicator function of A. Accordingly, the observed sample is

$$\mathcal{D}_n = \{(Z_i, \delta_i, \mathbf{X}_i), i = 1, \ldots, n\}.$$

Survival data in clinical trials or failure time data in reliability studies, for example, are often subject to such censoring. To be more specific, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life which result in death. In the sequel, we impose the following assumptions upon the distribution of $(\mathbf{X}, T)$. Denote by I a given compact set in $\mathbb{R}^d$ with nonempty interior and set, for any $\alpha > 0$,

$$I_\alpha = \{\mathbf{x} : \inf_{\mathbf{u}\in I} |\mathbf{x}-\mathbf{u}|_{\mathbb{R}^d} \leq \alpha\},$$

with $|\cdot|_{\mathbb{R}^d}$ standing for the usual Euclidean norm on $\mathbb{R}^d$. We will assume that, for a given $\alpha > 0$, $(\mathbf{X}, T)$ [resp. $\mathbf{X}$] has a density function $g_{\mathbf{X},T}$ [resp. $g_{\mathbf{X}}$] with respect to the Lebesgue measure on $I_\alpha \times \mathbb{R}$ [resp. I_α]. For $-\infty < t < \infty$, set

$$F(t) = \mathbb{P}(T \leq t),\ G(t) = \mathbb{P}(C \leq t),\ \text{and } H(t) = \mathbb{P}(Z \leq t),$$

the right-continuous distribution functions of T, C and Z respectively. For any right-continuous distribution function L defined on $\mathbb{R}$, denote by

$$T_L = \sup\{t \in \mathbb{R} : L(t) < 1\}$$

the upper point of the corresponding distribution. In this paper, we will mostly focus on the regression function of $\psi(Z)$ evaluated at $\mathbf{X} = \mathbf{x}$, for $\mathbf{x} \in I_\alpha$, given by

$$m_\psi(\mathbf{x}) = \mathbb{E}(\psi(Z) \mid \mathbf{X} = \mathbf{x}),$$

whenever it exists, is an unknown function, with real values. We will deal with the following family of estimators

$$\begin{aligned}
&\Psi_{n,h_n}(\mathbf{x}, f, K) \\
&= (1-\gamma_n)\,\Psi_{n-1,h_{n-1}}(\mathbf{x}, f, K) \\
&\quad +\gamma_n h_n^{-d}\delta_n G(Z_n)^{-1}\left\{(c_f(\mathbf{x})f(Z_n)+d_f(\mathbf{x}))K\left(h_n^{-1}\left(\mathbf{x}-\mathbf{X}_n\right)\right)\right\}, \quad (1.1)
\end{aligned}$$

where (γ_n) is a nonrandom positive sequence tending to zero as $n \to \infty$, (h_n) is a nonrandom positive sequence tending to zero as $n \to \infty$, called bandwidth, $f(\cdot), c_f(\cdot)$ and $d_f(\cdot)$ are some specific functions. For more explanation about the motivation of considering the proposed family of estimators (1.1), first, by considering, $c_f(\mathbf{x}) = 1/g_{\mathbf{X}}(\mathbf{x})$ and $d_f(\mathbf{x}) = -\mathbb{E}(f(Y) \mid \mathbf{X} = \mathbf{x})/g_{\mathbf{X}}(\mathbf{x})$ this corresponds to regression setting, see equation (3.1) in [23]. An other choice $c_f(\mathbf{x}) = 0$ and $d_f(\mathbf{x}) = 1$, which correspond to the kernel density estimator. Moreover, the introduction of the function $f(\cdot)$ in (1.1), is motivated by the following choices. Further, the choice $f(y) = y$ (or $f(y) = y^k$, where k is a strictly positive integer) into (1.1) let to the recursive Nadaraya-Watson kernel regression function estimator of

$$m(\mathbf{x}) := \mathbb{E}(Y \mid \mathbf{X} = \mathbf{x}).$$

Finally, the choice $f(y) = f_t(y) = \mathbb{1}\{y \le t\}$ can be used to study the recursive kernel estimator of the conditional distribution function

$$F(t|\mathbf{x}) := \mathbb{P}(Y \le t|X = \mathbf{x}).$$

More motivation on the use of the function $f(\cdot)$ in (1.1), can be found in [15]. The function $G(\cdot)$ is generally unknown and has to be estimated. We will denote by $G_n(\cdot)$ the Kaplan-Meier estimator of the function $G(\cdot)$, see [30]. Namely, adopting the conventions $\prod_\emptyset = 1$ and $0^0 = 1$ and setting

$$N_n(u) = \sum_{i=1}^{n} \mathbb{1}\{Z_i \ge u\},$$

we have

$$G_n(u) = 1 - \prod_{i:Z_i \le u} \left\{\frac{N_n(Z_i) - 1}{N_n(Z_i)}\right\}^{(1-\delta_i)}, \quad \text{for} \quad u \in \mathbb{R}.$$

Then, we have

$$\begin{aligned}
&\widehat{\Psi}_{n,h_n}(\mathbf{x}, f, K) \\
&= (1-\gamma_n)\,\widehat{\Psi}_{n-1,h_{n-1}}(\mathbf{x}, f, K) \\
&\quad +\gamma_n h_n^{-d}\delta_n G_n(Z_n)^{-1}\left\{(c_f(\mathbf{x})f(Z_n)+d_f(\mathbf{x}))K\left(h_n^{-1}\left(\mathbf{x}-\mathbf{X}_n\right)\right)\right\} \quad (1.2)
\end{aligned}$$

Moreover, we set $\Psi_{0,h_0}(\mathbf{x}, f, K) = 0$ and

$$\Pi_n = \prod_{j=1}^{n}\left(1-\gamma_j\right),$$

then, we will investigate the following family of estimators

$$\begin{aligned}
&\widehat{\Psi}_{n,h_n}(\mathbf{x}, f, K) \\
&= \Pi_n \sum_{i=1}^{n} \frac{\gamma_i}{\Pi_i h_i^d}\delta_i G_n(Z_i)^{-1}\left\{(c_f(\mathbf{x})f(Z_i)+d_f(\mathbf{x}))K(h_i^{-1}(\mathbf{x}-\mathbf{X}_i))\right\} \quad (1.3)
\end{aligned}$$

In particular, when $\mathbb{P}(C=\infty)=1$, we have $G(t)=0$, for all $t<\infty$, and obtain the uncensored case, $G_n(\cdot)$ then being equal to the usual empirical distribution function based upon $Y_1,\ldots,Y_n$. The recursive property in (1.2) is particularly useful when the number of the observed data increase since $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$ can be easily updated with each additional observation. In fact, if $(T_n,C_n,\mathbf{X}_n)$ is a new observation, the estimators $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$ can be updated recursively by the relation (1.2). From a practical point of view, this arrangement provides important savings in computational time and storage memory which a consequence of the fact that the estimate updating is independent of the history of the data. The main drawback of the classical kernel kernel estimator is the use of all data at each step of estimation. From a theoretical point of view, the main advantage of the investigation of such general family of estimators is that we can prove almost sure consistency with exact rate for several kernel-type estimators simultaneously. It is worth noting that the quantity $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$ includes as particular cases : the kernel type density estimator, the Nadaraya-Watson ([35] and [48]) estimator and the kernel type estimator of the conditional distribution, we may refer to [23, 24] for more details. In this sense, the present paper extends, in non trivial, way some previous results by considering a general kernel-type estimators given in (1.3).

The remainder of this paper is organized as follows. In the forthcoming section we give the assumption and the main results. More precisely, we provide the bias and the asymptotic variance. We establish the asymptotic normality of $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$ in Theorem 2.3. The Strong pointwise convergence rate is characterized in Theorem 2.5. We calculate the Mean Squared Error (MSE) and provide the optimal bandwidth in Proposition 2.6. Some concluding remarks and possible future developments are mentioned in Section 3. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to the Section 4.

2. Assumptions and Main Results

We define the following class of regularly varying sequences.

Definition 2.1. Let $\gamma\in\mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n)\in\mathcal{GS}(\gamma)$ if

$$\lim_{n\to+\infty} n\left[1-\frac{v_{n-1}}{v_n}\right]=\gamma. \tag{2.1}$$

Condition (2.1) was introduced by [26] to define regularly varying sequences (see also [3]). Noting that the acronym $\mathcal{GS}$ stand for (Galambos and Seneta). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b\in\mathbb{R}$, $n^{\gamma}(\log n)^{b}$, $n^{\gamma}(\log\log n)^{b}$, and so on.
In this section, we investigate asymptotic properties of the proposed estimators (1.3). The assumptions to which we shall refer are the following:

(A1): $K:\mathbb{R}^d\to\mathbb{R}$ is a continuous, bounded function satisfying

$$\int_{\mathbb{R}^d} K(\mathbf{z})\,d\mathbf{z}=1,$$

and, for all $j\in\{1,\ldots,d\}$, $\int_{\mathbb{R}^d} z_j K(\mathbf{z})\,d\mathbf{z}=0$ and

$$\int_{\mathbb{R}^d} z_j^2\,|K(\mathbf{z})|\,d\mathbf{z}<\infty.$$

(A2): (i)

$$(\gamma_n) \in \mathcal{GS}(-\alpha) \text{ with } \alpha \in (1/2, 1].$$

(ii)

$$(h_n) \in \mathcal{GS}(-a) \quad \text{with} \quad a \in (0, \alpha/d).$$

(iii)

$$\lim_{n\to\infty} (n\gamma_n) \in (\min\{2a, (\alpha - ad)/2\}, \infty].$$

(A3): (i) $g_{\mathbf{X},T}(\mathbf{s}, t)$ is twice continuously differentiable with respect to $\mathbf{s}$.
(ii) $s \mapsto \int_{\mathbb{R}} f(t)\, g_{\mathbf{X},T}(\mathbf{s}, t)\, dt$ is a bounded function continuous at $\mathbf{s} = \mathbf{x}$.
$s \mapsto \int_{\mathbb{R}} |f(t)|\, g_{\mathbf{X},T}(\mathbf{s}, t)\, dt$ is a bounded function.
(iii)

$$\int_{\mathbb{R}} |f(t)| \left| \frac{\partial g_{\mathbf{X},T}}{\partial x_i}(\mathbf{x}, t) \right| dt < \infty, i = 1, \ldots, d,$$

and

$$\mathbf{s} \mapsto \int_{\mathbb{R}} f(t) \frac{\partial^2 g_{\mathbf{X},T}}{\partial s_i \partial s_j}(\mathbf{s}, t)\, dt,$$

for $i, j = 1, \ldots, d$, is a bounded function continuous at $\mathbf{s} = \mathbf{x}$.

It is interesting to underline that the intuition behind the use of such bandwidth (h_n) belonging to $\mathcal{GS}(-a)$ is that the ratio h_{n-1}/h_n is equal to $1 + a/n + o(1/n)$, then using such bandwidth and using the assumption $(A2)$ on the bandwidth and on the stepsize, Lemma 4.1 ensures that the bias and the variance will depend only on h_n and not on $h_1, \ldots, h_n$, then the $MISE$ will depend also only on h_n, which will be helpful to deduce an optimal bandwidth. Assumption $(\mathbf{A2})(iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity is standard in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\gamma_n]^{-1}\right)$ is finite. To unburden our notation a bit and for simplicity, we introduce the following quantities

$$\xi = \lim_{n\to\infty} (n\gamma_n)^{-1}, \tag{2.2}$$

$$L(\mathbf{x}, f) = \mathbb{E}[f(Z) \mid \mathbf{X} = \mathbf{x}]\, g_{\mathbf{X}}(\mathbf{x}),$$

$$\Psi(\mathbf{x}, f) = c_f(\mathbf{x})\, L(\mathbf{x}, f) + d_f(\mathbf{x}) g_{\mathbf{X}}(\mathbf{x}),$$

$$\widetilde{g}_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}\left[G(Z)^{-1} \mid \mathbf{X} = \mathbf{x}\right] g_{\mathbf{X}}(\mathbf{x}),$$

$$LG(\mathbf{x}, f) = \mathbb{E}\left[\frac{f(T)}{G(T)} \mid \mathbf{X} = \mathbf{x}\right] g_{\mathbf{X}}(\mathbf{x}),$$

$$V(\mathbf{x}, f) = c_f^2(\mathbf{x})\, LG\left(\mathbf{x}, f^2\right) + d_f^2 \widetilde{g}_{\mathbf{x}}(\mathbf{x}) + 2c_f(\mathbf{x})\, d_f(\mathbf{x})\, LG(\mathbf{x}, f),$$

$$g_{ij}^{(2)}(\mathbf{x}) = \frac{\partial^2 g_{\mathbf{X}}}{\partial x_i \partial x_j}(\mathbf{x}), \tag{2.3}$$

$$L_{ij}^{(2)}(\mathbf{x}, f) = \frac{\partial^2 L}{\partial x_i \partial x_j}(\mathbf{x}, f),$$

$$B(\mathbf{x}, f) = \sum_{i_1, i_2 = 1}^{d} \int_{\mathbb{R}^d} z_{i_1} z_{i_2} K(\mathbf{z})\, d\mathbf{z} \left(c_f(\mathbf{x})\, L_{i_1 i_2}^{(2)}(\mathbf{x}, f) + d_f(\mathbf{x})\, g_{i_1 i_2}^{(2)}(\mathbf{x})\right),$$

$$R(K) = \int_{\mathbb{R}^d} K(\mathbf{z})^2\, d\mathbf{z}.$$

Some explanation about this notation can be found in [9, 10, 12].

2.1. Bias and variance. Our first result is the following proposition, which gives the bias and variance of $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$.

Proposition 2.2 (Bias and variance of $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$)**.** *Let Assumptions* (**A1**)-(**A3**) *hold, and assume that, for all* $i,j\in\{1,\ldots,d\}$ $g_{ij}^{(2)}(\cdot)$, $L_{ij}^{(2)}(\cdot,f)$, $\widetilde{g}_{\mathbf{X}}(\cdot)$ *and* $LG(\cdot,f)$ *are continuous at* $\mathbf{x}$*.*

(1) If $a\in(0,\alpha/(d+4)]$*, then*

$$\mathbb{E}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]-\Psi(\mathbf{x},f) = \frac{1}{2(1-2a\xi)}B(\mathbf{x},f)\,h_n^2+o\left(h_n^2\right). \tag{2.4}$$

If $a\in(\alpha/(d+4),1)$*, then*

$$\mathbb{E}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]-\Psi(\mathbf{x},f)=o\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{2.5}$$

(2) If $a\in[\alpha/(d+4),1)$*, then*

$$\operatorname{Var}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right] = \frac{R(K)}{2-(\alpha-ad)\xi}V(\mathbf{x},f)\frac{\gamma_n}{h_n^d} +o\left(\frac{\gamma_n}{h_n^d}\right). \tag{2.6}$$

If $a\in(0,\alpha/(d+4))$*, then*

$$\operatorname{Var}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]=o\left(h_n^4\right). \tag{2.7}$$

(3) If $\lim_{n\to\infty}(n\gamma_n)>\max\{2a,(\alpha-ad)/2\}$*, then* (2.4) *and* (2.6) *hold simultaneously.*

2.2. Central limit theorem and strong convergence. Let us now state the following theorem, which gives the weak convergence rate of the family of estimators Ψ_{n,h_n} defined in (1.1). Below, we write $Z\overset{\mathcal{D}}{=}\mathcal{N}(\mu,\sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2, $\overset{\mathcal{D}}{\rightarrow}$ denotes the convergence in distribution and $\overset{\mathbb{P}}{\rightarrow}$ the convergence in probability.

Theorem 2.3 (Weak pointwise convergence rate)**.** *Let Assumptions* (**A1**)-(**A3**) *hold, and assume that, for all* $i,j\in\{1,\ldots,d\}$ $g_{ij}^{(2)}(\cdot)$, $L_{ij}^{(2)}(\cdot,f)$, $\widetilde{g}_{\mathbf{X}}(\cdot)$ *and* $LG(\cdot,f)$ *are continuous at* $\mathbf{x}$*.*

(1) If there exists $c\geq 0$ *such that* $\gamma_n^{-1}h_n^{d+4}\to c$*, then*

$$\sqrt{\gamma_n^{-1}h_n^d}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)\right) \overset{\mathcal{D}}{\rightarrow}\ \mathcal{N}\left(\frac{\sqrt{c}}{2(1-2a\xi)}B(\mathbf{x},f),\frac{R(K)}{(2-(\alpha-ad)\xi)}V(\mathbf{x},f)\right).$$

(2) If $\gamma_n^{-1}h_n^{d+4}\to\infty$*, then*

$$\frac{1}{h_n^2}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)\right)\overset{\mathbb{P}}{\rightarrow}\frac{1}{2(1-2a\xi)}B(\mathbf{x},f).$$

Remark 2.4. When the bandwidth (h_n) is chosen such that

$$\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+4}=0,$$

(which corresponds to under smoothing), it follows from Theorem 2.3 that

$$\sqrt{\gamma_n^{-1}h_n^d}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)\right) \overset{\mathcal{D}}{\to} \mathcal{N}\left(0,\frac{R(K)}{(2-(\alpha-ad)\xi)}V(\mathbf{x},f)\right),$$

and then, in order to minimize the asymptotic variance the stepsize (γ_n) should be equal to $\left([1-ad]\,n^{-1}\right)$.

In the following theorem, we give the strong pointwise convergence rate of $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$.

Theorem 2.5 (Strong pointwise convergence rate)**.** *Let Assumptions* (**A1**)-(**A3**) *hold, and assume that, for all* $i,j\in\{1,\dots,d\}$ $g_{ij}^{(2)}(\cdot)$, $L_{ij}^{(2)}(\cdot,f)$, $\widetilde{g}_{\mathbf{X}}(\cdot)$ *and* $LG(\cdot,f)$ *are continuous at* $\mathbf{x}$.

(1) If there exists $c_1\geq 0$ *such that*

$$\gamma_n^{-1}h_n^{d+4}/\left(\ln\left[\sum_{i=1}^n\gamma_i\right]\right)\to c_1,$$

then, with probability one, the sequence

$$\left(\sqrt{\frac{\gamma_n^{-1}h_n^d}{2\ln\left[\sum_{i=1}^n\gamma_i\right]}}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)\right)\right)$$

is relatively compact[1] *and its limit set is the interval*

$$\left[\frac{1}{2(1-2a\xi)}\sqrt{\frac{c_1}{2}}B(\mathbf{x},f)-\sqrt{\frac{V(\mathbf{x},f)R(K)}{2-(\alpha-ad)\xi}},\right.$$
$$\left.\frac{1}{2(1-2a\xi)}\sqrt{\frac{c_1}{2}}B(\mathbf{x},f)+\sqrt{\frac{V(\mathbf{x},f)R(K)}{2-(\alpha-ad)\xi}},\right].$$

(2) If

$$\gamma_n^{-1}h_n^{d+4}/\left(\ln\left[\sum_{i=1}^n\gamma_i\right]\right)\to \infty,$$

then, with probability one,

$$\lim_{n\to\infty}\frac{1}{h_n^2}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)\right)=\frac{1}{2(1-2a\xi)}B(\mathbf{x},f).$$

2.3. MSE and choice of the optimal bandwidth. In the following proposition we give the MSE of the family of estimators $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$.

Proposition 2.6 (MSE of $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$)**.** *Let Assumptions* (**A1**)-(**A3**) *hold, and assume that, for all* $i,j\in\{1,\dots,d\}$ $g_{ij}^{(2)}(\cdot)$, $L_{ij}^{(2)}(\cdot,f)$, $\widetilde{g}_{\mathbf{X}}(\cdot)$ *and* $LG(\cdot,f)$ *are continuous at* $\mathbf{x}$.

(1) If $a\in(0,\alpha/(d+4))$, *then*

$$\mathrm{MSE}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]=\frac{1}{4(1-2a\xi)^2}h_n^4B(\mathbf{x},f)+o\left(h_n^4\right).$$

[1] Recall that a relatively compact subset is a subset whose closure is compact

(2) If $a=\alpha/\left(d+4\right)$, then

$$\begin{aligned}\mathrm{MSE}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right] &= \frac{1}{2-\left(\alpha-ad\right)\xi}\frac{\gamma_n}{h_n^d}R\left(K\right)V\left(\mathbf{x},f\right)\\ &\quad+\frac{1}{4\left(1-2a\xi\right)^2}h_n^4B\left(\mathbf{x},f\right)+o\left(h_n^4\right).\end{aligned}$$

(3) If $a\in\left(\alpha/\left(d+4\right),1\right)$, then

$$\mathrm{MSE}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right] = \frac{1}{2-\left(\alpha-ad\right)\xi}\frac{\gamma_n}{h_n^d}R\left(K\right)V\left(\mathbf{x},f\right)+o\left(\frac{\gamma_n}{h_n^d}\right).$$

The following corollary ensures that the bandwidth which minimize the MSE depend on the stepsize (γ_n) and then the corresponding MSE depend also on the stepsize (γ_n).

Corollary 2.7. *Let Assumptions* (**A1**)-(**A3**) *hold. To minimize the* MSE *of* $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$*, the stepsize (γ_n) must be chosen in $\mathcal{GS}\left(-1\right)$, the bandwidth (h_n) must equal*

$$\left(d^{1/(d+4)}\frac{\left(1-2a\xi\right)^{2/(d+4)}}{\left(2-\left(\alpha-ad\right)\xi\right)^{1/(d+4)}}\left\{R\left(K\right)\frac{V\left(\mathbf{x},f\right)}{B\left(\mathbf{x},f\right)}\right\}^{1/(d+4)}\gamma_n^{1/(d+4)}\right).$$

Then, we have

$$\begin{aligned}&\mathrm{MSE}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]\\ &= \left(1+\frac{d^{4/(d+4)}}{4}\right)\left(1-2a\xi\right)^{-2d/(d+4)}\left(2-\left(\alpha-ad\right)\xi\right)^{-4/(d+4)}\\ &\quad\times R\left(K\right)^{4/(d+4)}V\left(\mathbf{x},f\right)^{4/(d+4)}B\left(\mathbf{x},f\right)^{d/(d+4)}\gamma_n^{4/(d+4)}\\ &\quad+o\left(\gamma_n^{4/(d+4)}\right).\end{aligned}$$

The following corollary shows that, for a special choice of the stepsize

$$\left(\gamma_n\right)=\left(\gamma_0n^{-1}\right),$$

which fulfilled that

$$\lim_{n\to\infty}n\gamma_n=\gamma_0$$

and that $(\gamma_n)\in\mathcal{GS}\left(-1\right)$, the optimal value for h_n depend on γ_0 and then the corresponding MSE depend on γ_0.

Corollary 2.8. *Let Assumptions* (**A1**)-(**A3**) *hold. To minimize the* MSE *of* $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)$*, the stepsize (γ_n) must be chosen in $\mathcal{GS}\left(-1\right)$, $\lim_{n\to\infty}n\gamma_n=\gamma_0$, the bandwidth (h_n) must equal*

$$\left(\left(\gamma_0-2/\left(d+4\right)\right)^{2/(d+4)}\left(\frac{d}{4}\right)^{1/(d+4)}\left\{R\left(K\right)\frac{V\left(\mathbf{x},f\right)}{B\left(\mathbf{x},f\right)}\right\}^{1/(d+4)}n^{-1/(d+4)}\right),\tag{2.8}$$

from which we obtain

$$\begin{aligned}
&\mathrm{MSE}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right] \\
&= \left(1+d^{4/(d+4)}/4\right)\gamma_0^2\left(\gamma_0-2/\left(d+4\right)\right)^{-2(d+2)/(d+4)} \\
&\quad \times R\left(K\right)^{4/(d+4)}V\left(\mathbf{x},f\right)^{4/(d+4)}B\left(\mathbf{x},f\right)^{d/(d+4)}n^{-4/(d+4)} \\
&\quad +o\left(n^{-4/(d+4)}\right).
\end{aligned}$$

Moreover, the minimum of

$$\gamma_0^2\left(\gamma_0-2/\left(d+4\right)\right)^{-2(d+2)/(d+4)}$$

is reached at $\gamma_0=1$, then the bandwidth (h_n) must be equal to

$$\left(\left(\frac{d+2}{d+4}\right)^{2/(d+4)}\left(\frac{d}{4}\right)^{1/(d+4)}\left\{R\left(K\right)\frac{V\left(\mathbf{x},f\right)}{B\left(\mathbf{x},f\right)}\right\}^{1/(d+4)}n^{-1/(d+4)}\right), \qquad (2.9)$$

from which we obtain

$$\begin{aligned}
&\mathrm{MSE}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right] \\
&= \left(1+d^{4/(d+4)}/4\right)\left(\left(d+2\right)/\left(d+4\right)\right)^{-2(d+2)/(d+4)} \\
&\quad \times R\left(K\right)^{4/(d+4)}V\left(\mathbf{x},f\right)^{4/(d+4)}B\left(\mathbf{x},f\right)^{d/(d+4)}n^{-4/(d+4)} \\
&\quad +o\left(n^{-4/(d+4)}\right).
\end{aligned}$$

Remark 2.9. For notational convenience, we have chosen the same bandwidth sequence for each margins. This assumption can be dropped easily. If one wants to make use of the vector bandwidths (see, in particular, Chapter 12 of [19]). With obvious changes of notation, our results and their proofs remain true when h_n is replaced by a vector bandwidth $\mathbf{h}_n=(h_n^{(1)},\ldots,h_n^{(d)})$, where $\min h_n^{(i)}>0$. In this situation we set

$$h_n=\prod_{i=1}^{d}h_n^{(i)},$$

and for any vector $\mathbf{v}=(v_1,\ldots,v_d)$ we replace $\mathbf{v}/h_n$ by $(v_1/h_n^{(1)},\ldots,v_1/h_n^{(d)})$. For ease of presentation we chose to use real-valued bandwidths throughout.

3. Concluding Remarks

This paper proposes a general recursive kernel type estimators for spatial data defined by the stochastic approximation algorithm (1.3). The asymptotic laws of the proposed estimators are established under general conditions. In particular, we have obtained the central limit theorem and the strong pointwise convergence rate. We have discussed the MSE that is used the specify the optimal bandwidth in some sense. It would be of interest to extend the present work the case of functional data, that is the covariate $\mathbf{X}$ is $\mathcal{X}$-valued function, where $\mathcal{X}$ is an abstract space. A future research direction would be to extend our findings to the setting of serially dependent observations.

4. Mathematical Developments

This section is devoted to the proofs of our results. The previously presented notation continues to be used in the following.
For any distribution function (df) $L(\cdot)$ recall that

$$\tau_L = \sup\{t : L(t) < 1\}$$

be its support's right endpoint. Further, we will denote by τ_F (resp. τ_G) the upper endpoints of $F(\cdot)$ (resp. of $G(\cdot)$). In the following we assume that $\tau_F < \infty$, $G(\tau_F) > 0$, $\tau_H < \min(\tau_F, \tau_G)$ and C is independent to $(\mathbf{X}, T)$.
Now, we define the sequence (m_n) by setting

$$(m_n) \quad = \quad \begin{cases} \frac{\log\log n}{\sqrt{\gamma_n^{-1} h_n^{d+1}}} & \text{if } \dfrac{\log\log n}{\sqrt{\gamma_n^{-1} h_n^{d+5}}} = \infty, \\ h_n^2 & \text{otherwise.} \end{cases} \tag{4.1}$$

Further, we consider the following notation throughout this section

$$\mathcal{T}_n(\mathbf{x}, f) = h_n^{-d} \delta_n G(Z_n)^{-1} \{c_f(\mathbf{x}) f(Z_n) + d_f(\mathbf{x})\} K\left(h_n^{-1}(\mathbf{x} - \mathbf{X}_n)\right).$$

and we use the fact that,

$$\mathbb{1}_{\{T_1 \le C_1\}} \varphi(Z_1) = \mathbb{1}_{\{T_1 \le C_1\}} \varphi(T_1)$$

for all measurable function $\varphi(\cdot)$. Then, we readily obtain that

$$\begin{aligned} &\mathcal{T}_n(\mathbf{x}, f) \\ &\quad = \quad h_n^{-d} \mathbb{1}_{\{T_n < C_n\}} G(T_n)^{-1} \{c_f(\mathbf{x}) f(T_n) + d_f(\mathbf{x})\} K\left(h_n^{-1}(\mathbf{x} - \mathbf{X}_n)\right) \end{aligned} \tag{4.2}$$

Moreover, we have

$$\begin{aligned} &\left|\widehat{\Psi}_{n,h_n}(\mathbf{x}, f, K) - \Psi_{n,h_n}(\mathbf{x}, f, K)\right| \\ &\quad = \quad \Pi_n \left|\sum_{i=1}^n \Pi_i^{-1} \gamma_i h_i^{-d} \delta_i \left[\frac{1}{G_n(Z_i)} - \frac{1}{G(Z_i)}\right] \{c_f(\mathbf{x}) f(Z_i) + d_f(\mathbf{x})\} \right. \\ &\qquad\quad \left. \times K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_i}\right)\right| \\ &\quad = \quad \Pi_n \left|\sum_{i=1}^n \Pi_i^{-1} \gamma_i h_i^{-d} \mathbb{1}_{\{T_i < C_i\}} \left[\frac{1}{G_n(T_i)} - \frac{1}{G(T_i)}\right] \{c_f(\mathbf{x}) f(T_i) + d_f(\mathbf{x})\} \right. \\ &\qquad\quad \left. \times K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_i}\right)\right| \\ &\quad \le \quad \Pi_n \left|\sum_{i=1}^n \Pi_i^{-1} \gamma_i h_i^{-d} \left[\frac{G_n(T_i) - G(T_i)}{G_n(T_i) G(T_i)}\right] \{c_f(\mathbf{x}) f(T_i) + d_f(\mathbf{x})\} \right. \\ &\qquad\quad \left. \times K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_i}\right)\right| \\ &\quad \le \quad \frac{\sup_{t \le \tau_H}(|G_n(t) - G(t)|)}{G_n(\tau_H) G(\tau_H)} \Pi_n \left|\sum_{i=1}^n \Pi_i^{-1} \gamma_i h_i^{-d} \{c_f(\mathbf{x}) f(T_i) + d_f(\mathbf{x})\} \right. \\ &\qquad\quad \left. \times K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_i}\right)\right|. \end{aligned}$$

Then by using the strong law of large numbers (SLLN) and the law of iterated logarithm (LIL) on the censoring law (see formula (4.28) in [16], see also [25]), we have

$$\begin{aligned}\sup_{x\in S}\left|\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi_{n,h_n}(\mathbf{x},f,K)\right| &= O\left(\sqrt{\frac{\log\log n}{nh_n^{d+1}}}\right)\\ &= o\left(m_n\right).\end{aligned} \tag{4.3}$$

The following simple lemma will play an instrumental role in the sequel.

Lemma 4.1. *Let* $(v_n)\in\mathcal{GS}(v^*)$*,* $(\gamma_n)\in\mathcal{GS}(-\alpha)$*, and* $m>0$ *such that* $m-v^*\xi>0$ *where* ξ *is defined in* (2.2)*. We have*

$$\lim_{n\to+\infty} v_n\Pi_n^m\sum_{i=1}^n\Pi_i^{-m}\frac{\gamma_i}{v_i}=\frac{1}{m-v^*\xi}. \tag{4.4}$$

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

$$\lim_{n\to+\infty} v_n\Pi_n^m\left[\sum_{i=1}^n\Pi_i^{-m}\frac{\gamma_i}{v_i}\alpha_i+\delta\right]=0. \tag{4.5}$$

The proof is given in [33]. Lemma 4.1 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption $(\mathbf{A2})(iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity.

Proof of Proposition 2.2. We first note that we have

$$\begin{aligned}&\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)\\ &= \widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi_{n,h_n}(\mathbf{x},f,K)+\Psi_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f).\end{aligned}$$

Then, it follows from (4.3), that the asymptotic behavior of $\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)$ can be deduced from the one of $\Psi_{n,h_n}(\mathbf{x},f,K)-\Psi(\mathbf{x},f)$. Moreover, in view of (1.1) and (4.2), we can write that

$$\mathbb{E}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right]-\Psi(\mathbf{x},f)=\Pi_n\sum_{i=1}^n\Pi_i^{-1}\gamma_i\left\{\mathbb{E}\left[\mathcal{T}_i\left(\mathbf{x},f\right)\right]-\Psi(\mathbf{x},f)\right\}.$$

Since, we have

$$\begin{aligned}&\mathbb{E}\left[\mathcal{T}_i\left(\mathbf{x},f\right)\right]\\ &= \mathbb{E}\left[h_i^{-d}\mathbb{E}\left[\mathbb{1}_{\{T_i<C_i\}}|T_i,\mathbf{X}_i\right]G(T_i)^{-1}\left\{c_f\left(\mathbf{x}\right)f\left(T_i\right)+d_f\left(\mathbf{x}\right)\right\}\right.\\ &\qquad\left.\times K\left(h_i^{-1}\left(\mathbf{x}-\mathbf{X}_i\right)\right)\right]\\ &= \mathbb{E}\left[h_i^{-d}\left\{c_f\left(\mathbf{x}\right)f\left(T_i\right)+d_f\left(\mathbf{x}\right)\right\}K\left(h_i^{-1}\left(\mathbf{x}-\mathbf{X}_i\right)\right)\right].\end{aligned}$$

Taylor's expansion with integral remainder ensures that

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{T}_i\left(\mathbf{x}, f\right)\right] - \Psi(\mathbf{x}, f) \\
&\quad = c_f\left(\mathbf{x}\right)\int_{\mathbb{R}^d} K\left(\mathbf{z}\right)\left[L(\mathbf{x}-\mathbf{z}h_i, f) - L(\mathbf{x}, f)\right] d\mathbf{z} \\
&\qquad + d_f\left(\mathbf{x}\right)\int_{\mathbb{R}^d} K\left(\mathbf{z}\right)\left[g_{\mathbf{x}}(\mathbf{x}-\mathbf{z}h_i) - g_{\mathbf{x}}(\mathbf{x})\right] d\mathbf{z} \\
&\quad = \frac{h_i^2}{2}\left\{\int_{\mathbb{R}^d} K\left(\mathbf{z}\right)\sum_{i_1,i_2=1}^{d} z_{i_1}z_{i_2}d\mathbf{z}\left(c_f\left(\mathbf{x}\right)L_{i_1i_2}^{(2)}(\mathbf{x}, f) + d_f\left(\mathbf{x}\right)g_{i_1i_2}^{(2)}(\mathbf{x})\right)\right\} \\
&\qquad + h_i^2\left(c_f\left(\mathbf{x}\right)\delta_i\left(\mathbf{x}\right) + d_f\left(\mathbf{x}\right)\widetilde{\delta}_i\left(\mathbf{x}\right)\right),
\end{aligned}$$

where

$$\begin{aligned}
&\delta_i\left(\mathbf{x}, f\right) \\
&\quad = \sum_{i_1,i_2=1}^{d}\int_{\mathbb{R}^d}\int_0^1\left(1-\eta\right)z_{i_1}z_{i_2}K\left(\mathbf{z}\right)\left[L_{i_1i_2}^{(2)}\left(\mathbf{x}-\mathbf{z}h_i\eta, f\right) - L_{i_1i_2}^{(2)}\left(\mathbf{x}, f\right)\right]d\eta d\mathbf{z}, \\
&\widetilde{\delta}_i\left(\mathbf{x}, f\right) \\
&\quad = \sum_{i_1,i_2}^{d}\int_{\mathbb{R}^d}\int_0^1\left(1-\eta\right)z_{i_1}z_{i_2}K\left(\mathbf{z}\right)\left[g_{i_1i_2}^{(2)}\left(\mathbf{x}-\mathbf{z}h_i\eta, f\right) - g_{i_1i_2}^{(2)}\left(\mathbf{x}, f\right)\right]d\eta d\mathbf{z}.
\end{aligned}$$

By the fact that $L_{i_1i_2}^{(2)}(\cdot)$ and $g_{i_1i_2}^{(2)}(\cdot)$ are bounded and continuous at $\mathbf{x}$ for all $i_1, i_2 \in \{1, \ldots, d\}$, we readily infer that

$$\lim_{i\to\infty}\delta_i\left(\mathbf{x}, f\right) = 0 \quad\text{and}\quad \lim_{i\to\infty}\widetilde{\delta}_i\left(\mathbf{x}, f\right) = 0.$$

In the case $a \leq \alpha/(d+4)$, we have

$$\lim_{n\to\infty}\left(n\gamma_n\right) > 2a,$$

an application of Lemma 4.1 then yields to

$$\begin{aligned}
&\mathbb{E}\left[\Psi_{n,h_n}\left(\mathbf{x}, f, K\right)\right] - \Psi\left(x, f\right) \\
&\quad = \frac{1}{2}\int_{\mathbb{R}^d} K\left(\mathbf{z}\right)\sum_{i_1,i_2=1}^{d} z_{i_1}z_{i_2}d\mathbf{z}\left(c_f\left(\mathbf{x}\right)L_{i_1i_2}^{(2)}(\mathbf{x}, f) + d_f\left(\mathbf{x}\right)g_{i_1i_2}^{(2)}(\mathbf{x})\right) \\
&\qquad \times\Pi_n\sum_{i=1}^{n}\Pi_i^{-1}\gamma_i h_i^2\left[1+o\left(1\right)\right] \\
&\quad = \frac{h_n^2}{2(1-2a\xi)}\int_{\mathbb{R}^d} K\left(\mathbf{z}\right)\sum_{i_1,i_2=1}^{d} z_{i_1}z_{i_2}d\mathbf{z}\left(c_f\left(\mathbf{x}\right)L_{i_1i_2}^{(2)}(\mathbf{x}, f) + d_f\left(\mathbf{x}\right)g_{i_1i_2}^{(2)}(\mathbf{x})\right) \\
&\qquad \times\left[1+o\left(1\right)\right].
\end{aligned}$$

We so obtain (2.4), as sought. In the case $a > \alpha/(d+4)$, we have

$$h_n^2 = o\left(\sqrt{\gamma_n h_n^{-d}}\right),$$

since

$$\lim_{n\to\infty}\left(n\gamma_n\right) > \left(\alpha - ad\right)/2,$$

we will make use of Lemma 4.1 to infer that

$$\begin{aligned}\mathbb{E}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right]-\Psi(\mathbf{x},f) &= \Pi_n\sum_{i=1}^{n}\Pi_i^{-1}\gamma_i o\left(\sqrt{\gamma_i h_i^{-d}}\right)\\ &= o\left(\sqrt{\gamma_n h_n^{-d}}\right).\end{aligned}$$

This when combined with (4.3) implies that (2.5) holds. Further, for computing the variance, we have

$$\operatorname{Var}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right] = \Pi_n^2\sum_{i=1}^{n}\Pi_i^{-2}\gamma_i^2\operatorname{Var}\left[\mathcal{T}_i\left(\mathbf{x},f\right)\right].$$

Recall the trivial relation

$$\operatorname{Var}\left[\mathcal{T}_i\left(\mathbf{x},f\right)\right] = \mathbb{E}\left[\mathcal{T}_i^2\left(\mathbf{x},f\right)\right]-\left(\mathbb{E}\left[\mathcal{T}_i\left(\mathbf{x},f\right)\right]\right)^2.$$

First, we have

$$\begin{aligned}&\mathbb{E}\left[\mathcal{T}_i^2\left(\mathbf{x},f\right)\right]\\ &= \mathbb{E}\Big[h_i^{-2d}\mathbb{E}\left[\mathbb{1}_{\{T_i<C_i\}}|T_i,\mathbf{X}_i\right]G(T_i)^{-2}\left\{c_f\left(\mathbf{x}\right)f\left(T_i\right)+d_f\left(\mathbf{x}\right)\right\}^2\\ &\quad\times K^2\left(h_i^{-1}\left(\mathbf{x}-\mathbf{X}_i\right)\right)\Big]\\ &= \mathbb{E}\left[h_i^{-2d}G(T_i)^{-1}\left\{c_f\left(\mathbf{x}\right)f\left(T_i\right)+d_f\left(\mathbf{x}\right)\right\}^2K^2\left(h_i^{-1}\left(\mathbf{x}-\mathbf{X}_i\right)\right)\right].\end{aligned}$$

Then, we obtain

$$\begin{aligned}&\operatorname{Var}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right]\\ &= \Pi_n^2\sum_{i=1}^{n}\frac{\Pi_i^{-2}\gamma_i^2}{h_i^d}\left[V\left(\mathbf{x},f\right)\int_{\mathbb{R}^d}K^2\left(\mathbf{z}\right)d\mathbf{z}+\nu_i\left(\mathbf{x},f\right)-h_i^d\widetilde{\nu}_i\left(\mathbf{x},f\right)\right],\end{aligned}$$

where

$$\begin{aligned}\nu_i\left(\mathbf{x},f\right) &= c_f^2\left(\mathbf{x}\right)\int_{\mathbb{R}}K^2\left(\mathbf{z}\right)\left[LG\left(\mathbf{x}-\mathbf{z}h_i,f^2\right)-LG\left(\mathbf{x},f^2\right)\right]d\mathbf{z}\\ &\quad+d_f^2\left(\mathbf{x}\right)\int_{\mathbb{R}}K^2\left(\mathbf{z}\right)\left[\widetilde{g}_{\mathbf{x}}\left(\mathbf{x}-\mathbf{z}h_i\right)-\widetilde{g}_{\mathbf{x}}\left(\mathbf{x}\right)\right]d\mathbf{z}\\ &\quad+2c_f\left(\mathbf{x}\right)d_f\left(\mathbf{x}\right)\int_{\mathbb{R}}K^2\left(\mathbf{z}\right)\left[LG\left(\mathbf{x}-\mathbf{z}h_i,f\right)-LG\left(\mathbf{x},f\right)\right]d\mathbf{z},\\ \widetilde{\nu}_i\left(\mathbf{x},f\right) &= \left(\int_{\mathbb{R}^d}K\left(\mathbf{z}\right)\left(c_f\left(\mathbf{x}\right)L\left(\mathbf{x}-\mathbf{z}h_i,f\right)+c_f\left(\mathbf{x}\right)g_{\mathbf{x}}\left(\mathbf{x}-\mathbf{z}h_i\right)\right)dz\right)^2.\end{aligned}$$

In view of the condition (**A3**), we have

$$\lim_{i\to\infty}\nu_i\left(\mathbf{x},f\right)=0 \text{ and } \lim_{i\to\infty}h_i^d\widetilde{\nu}_i\left(\mathbf{x},f\right)=0.$$

In the case $a\geq\alpha/\left(d+4\right)$, we have

$$\lim_{n\to\infty}\left(n\gamma_n\right)>\left(\alpha-ad\right)/2,$$

and by applying Lemma 4.1, we readily obtain

$$\begin{aligned}\operatorname{Var}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right] &= \Pi_n^2\sum_{i=1}^n\frac{\Pi_i^{-2}\gamma_i^2}{h_i^d}\left[V(\mathbf{x},f)\int_{\mathbb{R}^d}K^2(\mathbf{z})\,d\mathbf{z}+o(1)\right]\\ &= \frac{1}{2-(\alpha-ad)\xi}\frac{\gamma_n}{h_n^d}\left[R(K)V(\mathbf{x},f)+o(1)\right],\end{aligned}$$

then (2.6) follows from (4.3). In the case $a<\alpha/(d+4)$, we have

$$\gamma_n h_n^{-d}=o\left(h_n^4\right).$$

Lemma 4.1 ensures that

$$\begin{aligned}\operatorname{Var}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right] &= \Pi_n^2\sum_{i=1}^n\Pi_i^{-2}\gamma_i o\left(h_i^4\right)\\ &= o\left(h_n^4\right),\end{aligned}$$

then (2.7) follows from (4.3). Hence the proof is complete. □

Proof of Theorem 2.3. Let us at first assume that, if $a\geq\alpha/(d+4)$, then

$$\begin{aligned}&\sqrt{\gamma_n^{-1}h_n^d}\left(\Psi_{n,h_n}(\mathbf{x},f,K)-\mathbb{E}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right]\right)\\ &\qquad\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(0,\frac{R(K)}{(2-(\alpha-ad)\xi)}V(\mathbf{x},f)\right).\end{aligned}\tag{4.6}$$

In the case when $a>\alpha/(d+4)$, Part 1 of Theorem 2.3 follows from the combination of (2.5) and (4.6). In the case when $a=\alpha/(d+4)$, Parts 1 and 2 of Theorem 2.3 follow from the combination of (2.4), (4.3) and (4.6). In the case $a<\alpha/(d+4)$, (2.7) implies that

$$h_n^{-2}\left(\Psi_{n,h_n}(\mathbf{x},f,K)-\mathbb{E}\left(\Psi_{n,h_n}(\mathbf{x},f,K)\right)\right)\stackrel{\mathbb{P}}{\to}0,$$

and a combination of (2.4) with (4.3) gives Part 2 of Theorem 2.3. Let us prove (4.6). In view of (1.1), we have

$$\begin{aligned}&\Psi_{n,h_n}(\mathbf{x},f,K)-\mathbb{E}\left[\Psi_{n,h_n}(\mathbf{x},f,K)\right]\\ &\quad=\Pi_n\sum_{i=1}^n\Pi_i^{-1}\gamma_i\left(\mathcal{T}_i(\mathbf{x},f)-\mathbb{E}\left[\mathcal{T}_i(\mathbf{x},f)\right]\right).\end{aligned}$$

Now, we set

$$\mathcal{Y}_i(\mathbf{x})=\Pi_i^{-1}\gamma_i\left\{\mathcal{T}_i(\mathbf{x},f)-\mathbb{E}\left[\mathcal{T}_i(\mathbf{x},f)\right]\right\}.\tag{4.7}$$

An application of Lemma 4.1 ensures that

$$\begin{aligned}v_n^2 &= \sum_{i=1}^n\operatorname{Var}\left(\mathcal{Y}_i(\mathbf{x})\right)\\ &= \sum_{i=1}^n\Pi_i^{-2}\gamma_i^2\operatorname{Var}\left(\mathcal{T}_i(\mathbf{x},f)\right)\\ &= \sum_{i=1}^n\frac{\Pi_i^{-2}\gamma_i^2}{h_i^d}\left[R(K)V(\mathbf{x},f)+o(1)\right]\\ &= \frac{1}{\Pi_n^2}\frac{\gamma_n}{h_n^d}\left[\frac{R(K)}{2-(\alpha-ad)\xi}V(\mathbf{x},f)+o(1)\right].\end{aligned}\tag{4.8}$$

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

$$\mathbb{E}\left[\left|\mathcal{T}_i\left(\mathbf{x},f\right)\right|^{2+p}\right] = O\left(\frac{1}{h_i^{d(1+p)}}\right). \tag{4.9}$$

By using the fact that

$$\lim_{n\to\infty}\left(n\gamma_n\right) > \left(\alpha - ad\right)/2,$$

implies that there exists $p>0$, such that

$$\lim_{n\to\infty}\left(n\gamma_n\right) > \frac{1+p}{2+p}\left(\alpha - ad\right).$$

Applying Lemma 4.1, it follows that

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[\left|\mathcal{Y}_i\left((\mathbf{x}\right)\right|^{2+p}\right] &= O\left(\sum_{i=1}^{n}\Pi_i^{-2-p}\gamma_i^{2+p}\mathbb{E}\left[\left|\mathcal{T}_i\left(\mathbf{x},f\right)\right|^{2+p}\right]\right)\\ &= O\left(\sum_{i=1}^{n}\frac{\Pi_i^{-2-p}\gamma_i^{2+p}}{h_i^{d(1+p)}}\right)\\ &= O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}h_n^{d(1+p)}}\right),\end{aligned}$$

from which we conclude that

$$\begin{aligned}\frac{1}{v_n^{2+p}}\sum_{i=1}^{n}\mathbb{E}\left[\left|\mathcal{Y}_i\left((\mathbf{x}\right)\right|^{2+p}\right] &= O\left(\left[\gamma_n h_n^{-d}\right]^{p/2}\right)\\ &= o\left(1\right).\end{aligned}$$

The convergence in (4.6) then follows from the application of Lyapounov's Theorem. □

Proof of Theorem 2.5. We start by setting that

$$\mathcal{L}_n\left(\mathbf{x}\right) = \sum_{i=1}^{n}\mathcal{Y}_i\left(\mathbf{x}\right),$$

and

$$s_n = \sum_{i=1}^{n}\gamma_i,$$

where $\mathcal{Y}_i$ is defined in (4.7), and set $\gamma_0 = h_0 = 1$.

- *Let us first consider the case $a \geq \alpha/\left(d+4\right)$.*

We set $H_n^2 = \Pi_n^2\gamma_n^{-1}h_n^d$, we then have

$$\begin{aligned}\ln\left(H_n^{-2}\right) &= -2\ln\left(\Pi_n\right) + \ln\left(\prod_{i=1}^{n}\frac{\gamma_{i-1}^{-1}h_{i-1}^{d}}{\gamma_i^{-1}h_i^{d}}\right)\\ &= \left(2-\xi\left(\alpha-ad\right)\right)s_n + o\left(s_n\right).\end{aligned} \tag{4.10}$$

Since $2-\xi\left(\alpha-ad\right)>0$, it follows, in particular, that

$$\lim_{n\to+\infty}H_n^{-2} = \infty.$$

Moreover, since we have

$$\lim_{n\to+\infty}\frac{H_n^2}{H_{n-1}^2} = 1,$$

it follows from (4.8) that

$$\lim_{n\to+\infty} H_n^2 \sum_{i=1}^{n} \operatorname{Var}\left[\mathcal{Y}_i\left(\mathbf{x}\right)\right] = \frac{V\left(\mathbf{x},f\right)R\left(K\right)}{2-\left(\alpha-ad\right)\xi}.$$

Now, in view of (4.9), we readily infer that

$$\mathbb{E}\left[\left|\mathcal{Y}_i\left(\mathbf{x}\right)\right|^3\right] = O\left(\Pi_i^{-3}\gamma_i^3 h_i^{-2d}\right).$$

Keeping in mind equation (4.10), we have from Lemma 4.1 the following

$$\begin{aligned} n^{-3/2}\sum_{i=1}^{n}\mathbb{E}\left(\left|H_n\mathcal{Y}_i\left(\mathbf{x}\right)\right|^3\right) &= O\left(n^{-3/2}H_n^3\sum_{i=1}^{n}\Pi_i^{-3}\gamma_i^3 h_i^{-2d}\right) \\ &= o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right). \end{aligned}$$

Then, an application of Theorem 1 of [32] ensures, with probability one, that the sequence

$$\left(\frac{H_n\mathcal{L}_n\left(\mathbf{x}\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right) = \left(\frac{\sqrt{\gamma_n^{-1}h_n^d}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K) - \mathbb{E}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\frac{V\left(\mathbf{x},f\right)R\left(K\right)}{2-\left(\alpha-ad\right)\xi}},\sqrt{\frac{V\left(\mathbf{x},f\right)R\left(K\right)}{2-\left(\alpha-ad\right)\xi}}\right]. \tag{4.11}$$

Moreover, it follows from (4.10), that

$$\lim_{n\to\infty}\frac{\ln\ln\left(H_n^{-2}\right)}{\ln s_n} = 1;$$

and then, with probablility one, the sequence

$$\left(\sqrt{\gamma_n^{-1}h_n^d}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K) - \mathbb{E}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]\right)/\sqrt{2\ln s_n}\right)$$

is relatively compact, and its limit set is the interval given in (4.11). Then, by combining (2.4) with (2.5) we conclude the proof of Theorem 2.5 in the case $a \geq \alpha/\left(d+4\right)$.

- *Let us now consider the case $a < \alpha/\left(d+4\right)$.*

We set

$$H_n^{-2} = \Pi_n^{-2}h_n^4\left(\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right)^{-1}.$$

We infer that

$$\begin{aligned} \ln\left(H_n^{-2}h_n^4\right) &= -2\ln\left(\Pi_n\right) + \ln\left(\prod_{i=1}^{n}\frac{h_{i-1}^{-4}}{h_i^{-4}}\right) \\ &= \left(2-4a\xi\right)s_n + o\left(s_n\right). \end{aligned} \tag{4.12}$$

Since $2-4a\xi > 0$, we readily infer that

$$\lim_{n\to+\infty} H_n^{-2}h_n^4 = \infty.$$

Moreover, since we have

$$\lim_{n\to+\infty}\frac{H_n^2}{H_{n-1}^2} = 1,$$

in view of (4.8) and applying Lemma 4.1, we conclude that

$$\lim_{n\to+\infty} H_n^2 \sum_{i=1}^{n} \mathrm{Var}\left[\mathcal{Y}_i\left(\mathbf{x}\right)\right] = o\left(1\right).$$

Now, in view of (4.9), it follows from (4.12) and Lemma 4.1 that

$$\begin{aligned} & n^{-3/2}\sum_{i=1}^{n}\mathbb{E}\left(\left|H_n\mathcal{Y}_i\left(\mathbf{x}\right)\right|^3\right) \\ &= O\left(n^{-3/2}H_n^3h_n^{-6}\left[\ln\ln\left(\Pi_n^{-2}h_n^4\right)\right]^{3/2}\sum_{i=1}^{n}\Pi_i^{-3}\gamma_i^3h_i^{-2d}\right) \\ &= o\left(\left[\ln\left(H_n^{-2}\right)\right]^{-1}\right). \end{aligned}$$

Then, an application of Theorem 1 of [32] ensures that, with probability one,

$$\begin{aligned} & \lim_{n\to\infty}\frac{H_n\mathcal{L}_n\left(\mathbf{x}\right)}{\sqrt{2\ln\ln\left(H_n^{-2}\right)}} \\ &= \lim_{n\to\infty} h_n^{-2}\sqrt{\frac{\ln\ln\left(\Pi_n^{-2}h_n^4\right)}{2\ln\ln\left(H_n^{-2}\right)}}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\mathbb{E}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]\right) \\ &= 0. \end{aligned}$$

Moreover, equation (4.12) ensures that

$$\lim_{n\to\infty}\frac{\ln\ln\left(H_n^{-2}\right)}{\ln\ln\left(\Pi_n^{-2}h_n^4\right)} = 1.$$

From which we infer that

$$\lim_{n\to\infty} h_n^{-2}\left(\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)-\mathbb{E}\left[\widehat{\Psi}_{n,h_n}(\mathbf{x},f,K)\right]\right) = 0 \quad \text{a.s.},$$

whence Theorem 2.5, in the case $a < \alpha/\left(d+4\right)$, follows from (2.4). □

Acknowledgements

The authors would like to thank the referees for their very helpful comments, which led to a considerable improvement of the original version of this paper.

References

1. Akaike, H.: An approximation to the density function. *Ann. Inst. Statist. Math., Tokyo*, **6**, (1954), 127–132.
2. Blanke, D. and Bosq, D.: *Inference and Prediction in Large Dimensions*, Wiley Series in Probability and Statistics. *Dunod*. 2007.
3. Bojanic, R. and Seneta, E.: A unified theory of regularly varying sequences. *Math. Z.*, **134** (1973), 91–106.
4. Bouzebda, S.: On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, **21** (2012), no. 3, 153–188.
5. Bouzebda, S. and Elhattab, I.: A strong consistency of a nonparametric estimate of entropy under random censorship. *C. R. Math. Acad. Sci. Paris*, **347** (2009), no. 13-14, 821–826.
6. Bouzebda, S. and Elhattab, I.: Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy. *Electron. J. Stat.*, **5** (2011), 440–459.
7. Bouzebda, S., Elhattab, I., and Seck, C. T.: Uniform in bandwidth consistency of nonparametric regression based on copula representation. *Statist. Probab. Lett.*, **137** (2018), 173–182.

8. Bouzebda, S. and Nemouchi, B.: Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U-statistics involving functional data. *J. Nonparametr. Stat.*, **32** (2020), no. 2, 452–509.
9. Bouzebda, S. and Slaoui, Y.: Nonparametric recursive method for kernel-type function estimators for spatial data. *Statist. Probab. Lett.*, **139** (2018), 103–114.
10. Bouzebda, S. and Slaoui, Y.: Large and moderate deviation principles for recursive kernel estimators of a regression function for spatial data defined by stochastic approximation method. *Statist. Probab. Lett.*, **151** (2019), 17–28.
11. Bouzebda, S. and Slaoui, Y.: Bandwidth selector for nonparametric recursive density estimation for spatial data defined by stochastic approximation method. *Comm. Statist. Theory Methods*, **49** (2020a), no. 12, 2942–2963.
12. Bouzebda, S. and Slaoui, Y.: Large and moderate deviation principles for recursive kernel estimators for spatial data. *J. Stoch. Anal.*, **1** (2020b), no. 1, 1–27.
13. Clarkson, J. A. and Adams, C. R.: On definitions of bounded variation for functions of two variables. *Trans. Amer. Math. Soc.*, **35** (1933), no. 4, 824–854.
14. Collomb, G.: Estimation non-paramétrique de la régression: revue bibliographique. *Internat. Statist. Rev.*, **49** (1981), no, 1, 75–93.
15. Deheuvels, P.: One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague)*, **47** (2011), no. 6, 855–865.
16. Deheuvels, P. and Einmahl, J. H. J.: Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. *Ann. Probab.*, **28** (2000), no. 3, 1301–1335.
17. Devroye, L.: *A course in density estimation*, volume 14 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA. 1987.
18. Devroye, L. and Györfi, L.: *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The L_1 view. 1985.
19. Devroye, L. and Lugosi, G.: *Combinatorial methods in density estimation*. Springer Series in Statistics. Springer-Verlag, New York. 2001.
20. Efron, B.: Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7** (1979), no. 1, 1–26.
21. Efron, B. and Tibshirani, R. J.: *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York. 1993.
22. Eggermont, P. P. B. and LaRiccia, V. N.: *Maximum penalized likelihood estimation. Vol. I*. Springer Series in Statistics. Springer-Verlag, New York. Density estimation. 2001.
23. Einmahl, U. and Mason, D. M.: An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, **13** (2000) no. 1, 1–37.
24. Einmahl, U. and Mason, D. M.: Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33** (2005) no. 3, 1380–1403.
25. Földes, A. and Rejtő, L.: A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56** (1981), no. 1, 75–86.
26. Galambos, J. and Seneta, E.: Regularly varying sequences. *Proc. Amer. Math. Soc.*, **41** (1973), 110–116.
27. Györfi, L., Kohler, M., Krzyżak, A., and Walk, H.: *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York. 2002.
28. Härdle, W.: *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge. 1990.
29. Hobson, E. W.: *The theory of functions of a real variable and the theory of Fourier's series. Vol. I*. Dover Publications Inc., New York, N.Y. 1958.
30. Kaplan, E. L. and Meier, P.: Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53** (1958), 457–481.
31. Mason, D. M.: Proving consistency of non-standard kernel estimators. *Stat. Inference Stoch. Process.*, **15** (2012), no. 2, 151–176.
32. Mokkadem, A. and Pelletier, M.: Compact law of the iterated logarithm for matrix-normalized sums of random vectors. *Theory Probab. Appl.*, **52** (2007), no. 4, 2459–2478.
33. Mokkadem, A. Pelletier, M. and Slaoui, Y.: The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference*, **139** (2009), 2459–2478.

34. Müller, H.-G.: *Nonparametric regression analysis of longitudinal data*, volume 46 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin. 1988.
35. Nadaraya, È. A.: On a regression estimate. *Teor. Verojatnost. i Primenen.*, **9** (1964), 157–159.
36. Nadaraya, È. A.: *Nonparametric estimation of probability densities and regression curves*, volume 20 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht. Translated from the Russian by Samuel Kotz. 1989.
37. Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33** (1962), 1065–1076.
38. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27** (1956), 832–837.
39. Scott, D. W.: *Multivariate density estimation.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. Theory, practice, and visualization, A Wiley-Interscience Publication. 1992.
40. Silverman, B. W.: *Density estimation for statistics and data analysis.* Monographs on Statistics and Applied Probability. Chapman & Hall, London. 1986.
41. Slaoui, Y.: Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *Journal of Probability and Statistics*, **2014a**, ID 739640, doi:10.1155/2014/739640.
42. Slaoui, Y.: The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.* **23** (2014b), 306–325.
43. Slaoui, Y.: Plug-In Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.* **69** (2015), 483–509.
44. Slaoui, Y.: Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface.* **9** (2016), 375–388.
45. Tapia, R. A. and Thompson, J. R.: *Nonparametric probability density estimation*, volume 1 of *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins University Press, Baltimore, Md. 1978.
46. Vituškin, A. G.: *O mnogomernyh variaciyah.* Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow. 1955.
47. Wand, M. P. and Jones, M. C.: *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London. 1995.
48. Watson, G. S.: Smooth regression analysis. *Sankhyā Ser. A*, **26** (1964), 359–372.
49. Wertz, W.: *Statistical density estimation: a survey*, volume 13 of *Angewandte Statistik und Ökonometrie [Applied Statistics and Econometrics]*. Vandenhoeck & Ruprecht, Göttingen. With German and French summaries. 1978.

Salim Bouzebda : Alliance Sorbonne Universités, Université de Technologie de Compiègne,, L.M.A.C., Compiègne, France
E-mail address: salim.bouzebda@utc.fr

Yousri Slaoui : Univ. Poitiers, Lab. Math. et Appl., Futuroscope Chasseneuil, France
E-mail address: Yousri.Slaoui@math.univ-poitiers.fr