Theory of Stochastic Processes
Vol. 25 (41), no. 2, 2020, pp. 61–73

FATMA BEN KHADHER AND YOUSRI SLAOUI

STRONG CONSISTENCY OF THE MODE OF MULTIVARIATE RECURSIVE KERNEL DENSITY ESTIMATOR UNDER STRONG MIXING HYPOTHESIS

In this research paper, we define a kernel estimator of the mode based on the recursive kernel density estimator developed by [23]. In addition, we establish its almost sure convergence under strong mixing hypothesis. Finally, we corroborate these theoretical results through numerical simulations.

1. Introduction

The estimation of mode function stands for a classical problem in statistics which has whetted considerable interest in various fields of applications. Indeed, it is widely used in machine learning applications and, in particular, in clustering methods (see [5]; [36]; [17]), computer vision (see [46]; [41]), power systems (see [45]; [35]), control (see [15]) and bioinformatics (see [13]). Multiple research works related to this topic within the frame work of nonparametric estimation have been elaborated. Among the most prominent ones, we mention [27], [34] and [42]. Recently, there has been a spate of interest in recursive estimation which has drawn the attention of multiple researchers. The basic merit of the recursive estimator lies in the fact that it can not only be updated with each additional new observation especially in large sample sizes but it can also be much better in terms of computational costs. In this work, our central focus is upon a recursive kernel estimator of the mode function defined by stochastic approximation method.
Let $X_1, \cdots, X_n$ be identically distributed $\mathbb{R}^d$-valued random vectors and let f denote the probability density of X_i, $i = 1, \cdots, n$. We consider a compact set Ω such that $\Omega \subset \mathbb{R}^d$, and we define the mode as follows

$$\theta := \arg\max_{y \in \Omega} f(y).$$

We assume that θ is unique.
In order to define our estimator of the mode, we first begin by constructing a stochastic algorithm for the estimation of the function f at a point x. We present an algorithm to search for the zero of the function $g : y \longmapsto f(x) - y$. Following Robbins-Monro's procedure, this algorithm is defined below as

(i) $f_0(x)$ is an arbitrary choice belonging to $\mathbb{R}$,
(ii) $\forall n \geqslant 1$, we set $f_n(x) = f_{n-1}(x) + \gamma_n W_n(x)$, where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero and $W_n(x)$ is an observation of the function g at the point $f_{n-1}(x)$.

To construct $W_n(x)$, we follow the approach of [28, 29] and [42] which are based on the classical property of stochastic algorithms (which is $\mathbb{E}\left[W_n(x)|\mathcal{F}_1^{n-1}\right] = 0$, where $\mathcal{F}_1^{n-1}$ stands for the σ-field of events generated by $\{X_1, \ldots, X_{n-1}\}$). In addition, we introduce a kernel K (which is a function satisfying $\int_{\mathbb{R}^d} K(z)dz = 1$), and a bandwidth (h_n) (which is a sequence of positive real numbers that goes to zero when $n \longrightarrow \infty$), and

2000 *Mathematics Subject Classification.* 62G05; 62G07; 62L20; 62E20; 60F05; 60E05.
Key words and phrases. Nonparametric estimation; Density estimation; Stochastic approximation; Mode; Strong mixing; Strong consistency.

we set $W_n(x) = K_{h_n}(x - X_n) - f_{n-1}(x)$, with $K_h(x) := h^{-d}K(h^{-1}x)$. Therefore, the recursive estimator of the density function f at the point x can be written as

$$f_n(x) = \pi_n f_0 + \pi_n \sum_{k=1}^{n} \pi_k^{-1}\gamma_k h_k^{-d} K\left(\frac{x - X_k}{h_k}\right)$$

with $\pi_n = \prod_{k=1}^{n}(1-\gamma_k)$. Our estimator of mode θ is defined as the random variable θ_n maximizing the recursive estimator f_n of f, which is expressed as

$$\theta_n := \arg\max_{t\in\Omega} f_n(t). \tag{1}$$

In the following, we assume that $X_1, \cdots, X_n$ satisfy the α-mixing dependency property, as defined below.

Definition 1.1. Let $X = (X_i)_{i\geqslant 1}$ be a sequence of random variables. Given a positive integer n, set

$$\alpha(n) = \sup_k \sup\left\{ \mid \mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B) \mid,\ A \in \mathcal{F}_1^k(x) \text{ and } B \in \mathcal{F}_{k+n}^{\infty}(x) \right\},$$

where $\mathcal{F}_i^k(x)$ is the σ-field of events generated by $\{X_i, \ldots, X_k\}$.
The sequence is α-mixing if the mixing coefficient $\alpha(n) \longrightarrow 0$ as $n \longrightarrow \infty$.

The α-mixing condition was introduced by [32]. It is also called the strong mixing property. There exist various examples of stochastic processes satisfying the α-mixing condition, such as ARMA models, GARCH models, the EXPAR models and the bilinear Markovian models. There are several types of α-mixing condition with different forms. If $\alpha(n) = O\left(n^{-k}\right)$ for some $k > 0$, the process is polynomially strong mixing. The process is exponentially strong mixing if $\alpha(n) = O\left(\exp^{-an}\right)$ for some mixing rate $a > 0$. Then, if there exists $\rho \in]0,1[$ such that $\alpha(n) = O(\rho^n)$, the process is geometrically strong mixing. Many practical applications of the α-mixing are illustrated in literature (see [9], [2], [4] and [8] for more details). There exist different mixing conditions, such as β-mixing condition (see [18]), ρ-mixing condition (see [14]), ϕ-mixing condition (see [16]) and ψ-mixing condition (see [1]).

The mode estimator has been investigated by several authors. Based on independent and identically distributed (iid) random data, the weak consistency and the asymptotic normality of the kernel sample mode was addressed by [27]. This result was extended in several directions by [6], [10, 11] and [44]. Strong consistency was explored by [24] and [43]. Asymptotic normality of kernel estimate of the mode was elaborated by [31]. The multidimensional study of the mode was carried out by [34] and [19].
Based on dependent random data, some studies have been performed for mode estimation. In ϕ-mixing condition as well as the conditional case, the strong consistency was enacted by [7]. In alpha mixing case, the strong consistency was established by [25] and the asymptotic normality was set forward by [21]. Numerous works were conducted, under censored and truncated data, to explore the property of nonparametric mode estimators (see [20], [26] and [12]).

The majority of properties of mode estimators are related to those of density estimators. We need always to handle the density case before that of the mode. This paper investigates the estimation of the mode, which is based on nonparametric recursive kernel density estimator developed by [23], under strong mixing conditions. The rest of the paper is organized as follows. In Section 2, the assumptions and main results are displayed. Section 3 is devoted to simulation study. Finally, a conclusion is presented in Section 4. The details of proofs are exhibited in Section 5 along with some auxiliary results.

2. Assumptions and main results

We consider stepsizes and bandwidths, which belong to the following class of regularly varying sequences.

Definition 2.1. Let $\gamma \in \mathbb{R}$ and $(\gamma_n)_{n\geqslant 1}$ be a nonrandom positive sequence. We state that $\gamma_n \in \mathcal{GS}(\gamma)$ if $\lim_{n\to\infty} n[1 - \frac{\gamma_{n-1}}{\gamma_n}] = \gamma$.

The assumptions to which we shall refer are the following:

(A1) The kernel function K:$\mathbb{R}^d \longrightarrow \mathbb{R}$ is a bounded probability density, lipschitz and satisfies for all $j \in \{1, \dots, d\}$, $\int_{\mathbb{R}} z_j K(z) dz_j = 0$ and $\int_{\mathbb{R}^d} z_j^2 K(z) dz < \infty$.

(A2)

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

(ii) $h_n \in \mathcal{GS}(-a)$ with $a \in]0, \alpha/d[$.

(iii) $\lim_{n\to\infty} n\gamma_n \in]\min\{2a, (1 - ad)/2\}, \infty]$.

(A3) f is bounded, twice differentiable on Ω, and, for all $i, j \in \{1, \cdots, d\}$, $\partial^2 f/\partial x_i \partial x_j$ is bounded.

(A4) The joint density $f_{(i,j)}$ of (X_i, X_j) exists for all (i, j), and there exists a constant $M > 0$ such that

$$\sup_{|i-j|\geqslant 1} \sup_{t_1,t_2\in\Omega} \left| f_{(i,j)}(t_1, t_2) - f(t_1)f(t_2) \right| < M.$$

(A5) The mixing coefficient of the X_i's satisfies $\alpha(n) = O(n^{-\nu})$ for some $\nu \geqslant 3$.

(A6) The mode θ satisfies the following property: for any $\varepsilon > 0$ and x, there exists $\eta \neq 0$ such that $|\theta - x| > \varepsilon$ implies that $|f(\theta) - f(x)| > \eta$.

(A7)

(i) $n^{1/\nu} \gamma_n^{1-1/\nu} \underset{n\longrightarrow\infty}{\longrightarrow} 0$.

(ii)

$$\begin{cases} a(d\nu - 2) - \alpha(d + 2) > 6 & \text{if } a \geq \alpha/(d + 4) \\ a(d - 2\nu - 6) - \alpha > 6 & \text{if } a < \alpha/(d + 4). \end{cases}$$

Remark 2.1. Assumption **(A1)** on the kernel is widely used in the recursive and non-recursive framework for the functional estimation. Assumptions **(A2)** on the stepsize and the bandwidth are used in the recursive framework for the estimation of the density function ([23]; [37, 38, 39]). Hypothesis **(A2)**(i) and **(A2)**(ii) ensure that the bandwidth (h_n) and the stepsize (γ_n) go to zero as n goes to infinity. Moreover, the stepsize (γ_n) goes to zero more rapidly than the bandwidth (h_n). Assumption **(A2)**(iii) on the limit as n goes to infinity of $(n\gamma_n)$ is usual in the framework of stochastic approximation algorithms. It implies that the limit of $(n\gamma_n)^{-1}$ is finite. Assumption **(A3)** on the function f allows us to calculate the properties of our estimator. Condition **(A4)** is needed to calculate the covariance. **(A5)** states a condition on the mixing coefficient. Assumption **(A6)** is classical in mode estimation. Finally, hypothesis **(A7)** provides a condition for the bandwidth allowing the estimation of the covariance term.

Throughout this paper, we shall use the following notation:

$$\varepsilon = \lim_{n\to\infty} (n\gamma_n)^{-1}, \tag{2}$$

$$\mu_j^2 = \int_{\mathbb{R}^d} z_j^2 K(z) dz, \quad \forall j \in \{1, \cdots, d\}, \tag{3}$$

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

$$Z_n(x) = \frac{1}{h_n^d} K\left(\frac{x - X_n}{h_n}\right).$$

The almost sure convergence is denoted by *a.s.*.

Now, we shall prove the consistency of our estimator (1) and give the rate of convergence.

Proposition 2.1. *Let Assumptions **(A1)-(A7)** hold.*

$$\sup_{t\in\Omega}|f_n(t)-f(t)| = \begin{cases} O\left(\sqrt{\gamma_n h_n^{-d}\log n}\right) & \text{if } a>\alpha/(d+4) \\ O\left(\max\left(\sqrt{\gamma_n h_n^{-d}\log n}, h_n^2\right)\right) & \text{if } a=\alpha/(d+4) \\ O\left(h_n^2\sqrt{\log n}\right) & \text{if } a<\alpha/(d+4) \end{cases}$$

a.s. as $n\to\infty$.

Proposition 2.2. *Under the assumption of Proposition* 2.1*, we have*

$$\theta_n-\theta = \begin{cases} O\left(\left(\gamma_n h_n^{-d}\log n\right)^{1/4}\right) & \text{if } a>\alpha/(d+4) \\ O\left(\max\left(\left(\gamma_n h_n^{-d}\log n\right)^{1/4}, h_n\right)\right) & \text{if } a=\alpha/(d+4) \\ O\left(h_n\left(\log n\right)^{1/4}\right) & \text{if } a<\alpha/(d+4) \end{cases}$$

a.s. as $n\to\infty$.

3. Simulation study

In this section, we aim to compare our proposed recursive kernel estimator of mode, defined by (1), with the mode estimator based on the well-known non recursive kernel density estimator introduced by [33],

$$\tilde{\theta}_n := \arg\max_{t\in\Omega}\tilde{f}_n(t), \tag{4}$$

where $\tilde{f}_n(t)=\frac{1}{nh_n^d}\sum_{k=1}^n K\left(\frac{x-X_k}{h_n}\right)$.

3.1. **The study design.** Let us consider the following simulation design, we simulate $N=500$ samples of sizes, $n=50$, $n=100$, $n=150$ and a sequence of m-dependent variables

$$X_i=\sum_{i}^{i+m}\sqrt{|Y_i|},$$

where $(Y_i)_i$ are generated from the following mixture distributions:

- $Y\sim\frac{1}{2}\mathcal{N}(2.5,6)+\frac{1}{2}\mathcal{N}(9,1)$.
- $Y\sim\frac{1}{2}\mathcal{N}(2,6)+\frac{1}{2}\mathcal{N}(8,1)$.

Next, we calculate the $\overline{ISE}$ (Integrated Squared Error) and the $\overline{IAE}$ (Integrated Absolute Error) of the two estimators;

$$\overline{ISE}=\frac{1}{N}\sum_{i=1}^N\int_{\mathbb{R}}\left(\widehat{\theta}^{[i]}(x)-\theta(x)\right)^2dx \quad \text{and} \quad \overline{IAE}=\frac{1}{N}\sum_{i=1}^N\int_{\mathbb{R}}\left|\widehat{\theta}^{[i]}(x)-\theta(x)\right|dx,$$

where $\widehat{\theta}^{[i]}$ corresponds to the estimator computed from the ith sample. In order to calculate the $\overline{ISE}$ and the $\overline{IAE}$ of the two mode estimators, we need to use the following quantities:

- The normal kernel function K.
- The stepsize $\gamma_n=n^{-1}$.
- The bandwidth (h_n) is chosen with plug-in method, given in [38].

		$Y \sim \frac{1}{2}\mathcal{N}(2.5,6) + \frac{1}{2}\mathcal{N}(9,1)$		
		$n = 50$	$n = 100$	$n = 150$
		$\overline{ISE}$	$\overline{ISE}$	$\overline{ISE}$
$m = 2$				
	Non recursive	0.315226	0.148972	0.103131
	Recursive	**0.158550**	**0.029883**	**0.021263**
$m = 4$				
	Non recursive	0.194686	0.133191	0.100986
	Recursive	**0.113563**	**0.125054**	**0.020975**

TABLE 1. $\overline{ISE}$ for $N = 500$ trials of the non recursive estimator (4) and the recursive estimator (1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{ISE}$.

		$Y \sim \frac{1}{2}\mathcal{N}(2.5,6) + \frac{1}{2}\mathcal{N}(9,1)$		
		$n = 50$	$n = 100$	$n = 150$
		$\overline{IAE}$	$\overline{IAE}$	$\overline{IAE}$
$m = 2$				
	Non recursive	0.561450	0.385970	0.321140
	Recursive	**0.398184**	**0.172867**	**0.145820**
$m = 4$				
	Non recursive	0.441232	0.275033	0.115408
	Recursive	**0.336991**	**0.155993**	**0.111827**

TABLE 2. $\overline{IAE}$ for $N = 500$ trials of the non recursive estimator (4) and the recursive estimator (1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{IAE}$.

		$Y \sim \frac{1}{2}\mathcal{N}(2,6) + \frac{1}{2}\mathcal{N}(8,1)$		
		$n = 50$	$n = 100$	$n = 150$
		$\overline{ISE}$	$\overline{ISE}$	$\overline{ISE}$
$m = 3$				
	Non recursive	0.337509	0.136912	0.079324
	Recursive	**0.305999**	**0.115895**	**0.058715**
$m = 5$				
	Non recursive	0.154778	0.142782	0.071498
	Recursive	**0.303375**	**0.99815**	**0.045986**

TABLE 3. $\overline{ISE}$ for $N = 500$ trials of the non recursive estimator (4) and the recursive estimator (1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{ISE}$.

3.2. **Results.** For each configuration of the simulation design parameters, we calculate the $\overline{ISE}$ and the $\overline{IAE}$ of the non recursive estimator (4) and the recursive estimator (1). From Table 1, Table 2, Table 3 and Table 4, it is clear that, the proposed recursive estimator (1) outperformed the non recursive estimator (4) in all the considered situations. We can observe that the $\overline{ISE}$ decreases as m increases. We can observe also that the $\overline{ISE}$ decreases as the sample size n increases. This simulation study shows the good performance of the recursive estimator with an appropriate choice of stepsize and bandwidth parameters.

		$Y \sim \frac{1}{2}\mathcal{N}(2,6) + \frac{1}{2}\mathcal{N}(8,1)$		
		$n = 50$	$n = 100$	$n = 150$
		$\overline{IAE}$	$\overline{IAE}$	$\overline{IAE}$
$m = 3$				
	Non recursive	0.580955	0.281645	0.192126
	Recursive	**0.553173**	**0.242312**	**0.126076**
$m = 5$				
	Non recursive	0.393418	0.224771	0.99556
	Recursive	**0.550795**	**0.224771**	**0.106438**

Table 4. $\overline{IAE}$ for $N = 500$ trials of the non recursive estimator (4) and the recursive estimator (1), for $n = 50$, $n = 100$ and $n = 150$. The bold values indicates the smallest values of $\overline{IAE}$.

4. Conclusion

In this paper, we attempted to elaborate a recursive kernel mode estimator based on stochastic approximation algorithm. We established the strong consistency of this estimator under α-mixing condition. Investing the same selected parameters in [23], which minimize the mean squared error of recursive density estimator, the proposed recursive mode estimator maintains the same convergence rate with non-recursive mode estimator defined by (4). The two previous estimators are asymptotically equivalent. In addition, the main merit of our estimator resides in its update, when a new sample information becomes available. Tackling this area is extremely interesting as it offers new perspectives for future works to consider multiple directions within this framework. This involves the elaboration of recursive mode estimation for dependent strong mixing functional data like in [40]. Furthermore, our proposed recursive kernel mode estimator is promising and can be extended in such a way as addressing recursive nonparametric estimation in the Bayesian work (see [3]).

5. Proofs

Before setting the outlines of the proofs, we introduce the following technical lemma, which is proved in [23], and which will be used throughout the demonstrations.

Lemma 5.1.
Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$ and $m > 0$ such that $m - v^*\varepsilon > 0$ where ε is defined in (2). Then,

$$\lim_{n\to\infty} v_n \pi_n^m \sum_{k=1}^{n} \pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\varepsilon}.$$

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

$$\lim_{n\to\infty} v_n \pi_n^m \left[\sum_{k=1}^{n} \pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + C\right] = 0.$$

Proof of Proposition 2.1. The proof rests on the following decomposition

$$|f_n(t) - f(t)| \quad \leq \quad |f_n(t) - \mathbb{E}[f_n(t)]| + |\mathbb{E}[f_n(t)] - f(t)|$$

and is based on the proofs of the following three lemmas.

Lemma 5.2. *Under Assumptions **(A1)-(A3)**, we have*

$$\sup_{t\in\Omega} |\mathbb{E}[f_n(t)] - f(t)| = \begin{cases} O\left(h_n^2\right) & \text{if } a \leq \alpha/(d+4) \\ o\left(\sqrt{\gamma_n h_n^{-d}}\right) & \text{if } a > \alpha/(d+4) \end{cases}$$

as $n \to \infty$.

The proof of Lemma 5.2 is presented in [23].

Lemma 5.3. *(Fuk-Nagaev) Let* $(W_i)_{i\in\mathbb{N}}$ *be a sequence of centered real random variables, with a strong mixing coefficient* $\alpha(n) = O\left(n^{-\nu}\right)$, $\nu > 1$, *such that* $\forall n \in \mathbb{N}$, $1 \leqslant i \leqslant n$, $|W_i| < +\infty$. *Hence, for all* $\varepsilon > 0$ *and* $r > 1$, *there exists a constant* c *such that*

$$\mathbb{P}\left\{|\sum_{k=1}^{n} W_i| > \varepsilon\right\} \leqslant c\left(1 + \frac{\varepsilon^2}{16rS_n^2}\right)^{-r/2} + ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{\nu+1}$$

where $S_n^2 = \sum_{i,j=1}^{n} | \mathbb{C}ov\left(W_i, W_j\right) |$.

For more details about previous Lemma 5.3, we refer to [30], p. 87, 6.19b.

Lemma 5.4. *Under Assumptions **(A1)-(A7)**, we have*

$$\sup_{t\in\Omega} |f_n(t) - \mathbb{E}[f_n(t)]| = \begin{cases} O\left(\sqrt{\gamma_n h_n^{-d}\log n}\right) & \text{if } a \geq \alpha/(d+4) \qquad (5)\\ O\left(h_n^2\sqrt{\log n}\right) & \text{if } a < \alpha/(d+4) \qquad (6)\end{cases}$$

a.s. as $n \to \infty$.

Proof of Lemma 5.4. The proof relies upon the following assertion: the compact set Ω can be covered by a finite number λ_n of balls $\mathcal{B}_k(t_k^*, b_n)$ centered at t_k^*, $1 \leqslant k \leqslant \lambda_n$ where b_n satisfies

$$b_n = \gamma_n^{1/2} h_n^{1+d/2}. \tag{7}$$

Since Ω is bounded, one can find $l > 0$ such that $\lambda_n \leq l b_n^{-1}$. For any $t \in \Omega$, there exists k such that

$$|t - t_k^*| \leqslant b_n. \tag{8}$$

Now, we set for $t \in \Omega$

$$T_i(t) = \pi_i^{-1}\gamma_i h_i^{-d}\left\{K\left(\frac{t - X_i}{h_i}\right) - \mathbb{E}\left(K\left(\frac{t - X_i}{h_i}\right)\right)\right\}. \tag{9}$$

Evidently, we get

$$\begin{aligned} \pi_n \sum_{i=1}^{n} T_i(t) &= f_n(t) - \mathbb{E}\left(f_n(t)\right) \\ &= \{(f_n(t) - f_n(t_k^*)) - (\mathbb{E}(f_n(t) - \mathbb{E}(f_n(t_k^*))\} + \{f_n(t_k^*) - \mathbb{E}(f_n(t_k^*)\} \\ &:= \pi_n \sum_{i=1}^{n} \tilde{T}_i(t) + \pi_n \sum_{i=1}^{n} T_i(t_k^*) \end{aligned}$$

with

$$\begin{aligned} \tilde{T}_i(t) = \pi_i^{-1}\gamma_i h_i^{-d}&\left\{K\left(\frac{t - X_i}{h_i}\right) - K\left(\frac{t_k^* - X_i}{h_i}\right)\right\} \\ &- \pi_i^{-1}\gamma_i h_i^{-d}\left\{\mathbb{E}\left(K\left(\frac{t - X_i}{h_i}\right)\right) - \mathbb{E}\left(K\left(\frac{t_k^* - X_i}{h_i}\right)\right)\right\}. \end{aligned}$$

As a matter of fact, we have

$$\begin{aligned} \sup_{t\in\Omega}\left|\pi_n \sum_{i=1}^{n} T_i(t)\right| &\leqslant \max_{k\leqslant\lambda_n}\sup_{t\in\mathcal{B}_k}\left|\pi_n \sum_{i=1}^{n} \tilde{T}_i(t)\right| + \max_{k\leqslant\lambda_n}\left|\pi_n \sum_{i=1}^{n} T_i(t_k^*)\right| \\ &:= U_1 + U_2. \end{aligned}$$

In order to investigate U_1, we observe that

$$\begin{aligned}\left|\pi_n \sum_{i=1}^{n} \tilde{T}_i(t)\right| &\leqslant \pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-d}\left|K\left(\frac{t-X_i}{h_i}\right)-K\left(\frac{t_k^*-X_i}{h_i}\right)\right| \\ &\quad +\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-d}\mathbb{E}\left[\left|K\left(\frac{t-X_i}{h_i}\right)-K\left(\frac{t_k^*-X_i}{h_i}\right)\right|\right] \\ &:= V_1(t)+V_2(t).\end{aligned}$$

Assumptions **(A1)**, (7) and (8) and the application of Lemma 5.1 provide

$$\begin{aligned}V_1(t) &\leqslant c\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-d}\left|\frac{t-t_k^*}{h_i}\right| \\ &\leqslant c\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-(d+1)}\left|t-t_k^*\right| \\ &\leqslant c b_n h_n^{-(d+1)}\frac{1}{1+a(d+1)\varepsilon} \\ &\leqslant c\gamma_n^{1/2}h_n^{-d/2}\frac{1}{1+a(d+1)\varepsilon} \\ &= O\left(\sqrt{\gamma_n h_n^{-d}}\right),\end{aligned}$$

and

$$\begin{aligned}V_2(t) &\leqslant c\pi_n \sum_{i=1}^{n} \pi_i^{-1}\gamma_i h_i^{-(d+1)}\mathbb{E}\left[\left|t-t_k^*\right|\right] \\ &= O\left(\sqrt{\gamma_n h_n^{-d}}\right).\end{aligned}$$

Thus, we get

$$U_1 = O\left(\sqrt{\gamma_n h_n^{-d}}\right) a.s \quad \text{as} \quad n\to\infty.$$

Now, in order to study U_2, we use Lemma 5.3. For that, let

$$W_i = \pi_n T_i(t_k^*) = \pi_n\pi_i^{-1}\gamma_i h_i^{-d}\left\{K\left(\frac{t_k^*-X_i}{h_n}\right)-\mathbb{E}\left(K\left(\frac{t_k^*-X_i}{h_n}\right)\right)\right\}. \tag{10}$$

Then, we have to calculate

$$\begin{aligned}S_n^2 &= \sum_{i,j=1}^{n} \mid \mathbb{C}ov\left(W_i,W_j\right)\mid \\ &= \sum_{i\neq j}\left|\mathbb{C}ov\left(W_i,W_j\right)\right| + \sum_{i=1}^{n}\mathbb{V}ar\left(W_i\right) \\ &:= S_n^{2*} + \sum_{i=1}^{n}\mathbb{V}ar\left(W_i\right).\end{aligned}$$

On the one hand, under **(A1)-(A3)**, we obtain

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{V}ar\left(W_i\right) &= \pi_n^2\sum_{i=1}^{n}\pi_i^{-2}\gamma_i^2\mathbb{V}ar\left(Z_i(t_k^*)\right) \\ &= \begin{cases} O\left(\gamma_n h_n^{-d}\right) & \text{if } a\geq \alpha/(d+4) \\ o\left(h_n^4\right) & \text{if } a<\alpha/(d+4),\end{cases}\end{aligned}$$

see Proposition 1 in [23] for more details about computation of the variance. Now, from (10) as well as under assumptions **(A1)** and **(A4)**, we have

$$
\begin{aligned}
|\mathbb{C}ov\,(W_i,W_j)| &= \Bigg|\mathbb{E}\left[\pi_n^2\pi_i^{-1}\pi_j^{-1}\gamma_i\gamma_j h_i^{-d}h_j^{-d}K\left(\frac{t_k^*-X_i}{h_n}\right)K\left(\frac{t_k^*-X_j}{h_n}\right)\right] \\
&\quad - \mathbb{E}\left[\pi_n\pi_i^{-1}\gamma_i h_i^{-d}K\left(\frac{t_k^*-X_i}{h_n}\right)\right]\mathbb{E}\left[\pi_n\pi_j^{-1}\gamma_j h_j^{-d}K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\Bigg| \\
&= \Bigg|\pi_n^2\pi_i^{-1}\pi_j^{-1}\gamma_i\gamma_j h_i^{-d}h_j^{-d}\left(\mathbb{E}\left[K\left(\frac{t_k^*-X_i}{h_n}\right)K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\right. \\
&\quad \left.- \mathbb{E}\left[K\left(\frac{t_k^*-X_i}{h_n}\right)\right]\mathbb{E}\left[K\left(\frac{t_k^*-X_j}{h_n}\right)\right]\right)\Bigg| \\
&= \pi_n^2\pi_i^{-1}\pi_j^{-1}\gamma_i\gamma_j\int_{\mathbb{R}^{2d}}K(t_1)K(t_2)\,\big|f_{(i,j)}(t_k^*-t_1h_i,t_k^*-t_2h_j) \\
&\qquad - f(t_k^*-t_1h_i)f(t_k^*-t_2h_j)\big|\,dt_1dt_2 \\
&\leqslant M\pi_n^2\pi_i^{-1}\gamma_i\pi_j^{1}\gamma_j \\
&= O(\pi_n^2\pi_i^{-1}\gamma_i\pi_j^{1}\gamma_j). \qquad (11)
\end{aligned}
$$

Next, to asses the term S_n^{2*}, we use a technique developed by [22]. We define the sets

$$F_1=\{(i,j)\text{ such that }1\leqslant|i-j|\leqslant\beta_n\}$$

and

$$F_2=\{(i,j)\text{ such that }\beta_n+1\leqslant|i-j|\leqslant n-1\}$$

where $\beta_n=o(n)$. Let

$$\mathcal{F}_{1,n}=\sum_{i,j\in F_1}|\mathbb{C}ov\,(W_i,W_j)|\ \text{ and }\ \mathcal{F}_{2,n}=\sum_{i,j\in F_2}|\mathbb{C}ov\,(W_i,W_j)|\,.$$

Applying the upper bound in (11), we have

$$
\begin{aligned}
\mathcal{F}_{1,n} &\leqslant M\pi_n^2\sum_{i,j\in F_1}\pi_i^{-1}\gamma_i\pi_j^{-1}\gamma_j \\
&\leqslant M\pi_n^2\sum_{j=1}^{n}\sum_{k=1}^{\beta_n}\pi_{k+j}^{-1}\gamma_{k+j}\pi_j^{-1}\gamma_j \\
&\leqslant M\pi_n^2\sum_{j=1}^{n}\sum_{k=1}^{\beta_n}\pi_j^{-2}\gamma_j^2\frac{1}{(1-\gamma_{j+1})\cdots(1-\gamma_{j+k})} \\
&\leqslant M\beta_n\pi_n^2\sum_{j=1}^{n}\pi_j^{-2}\gamma_j^2,
\end{aligned}
$$

and applying Lemma 5.1, we get

$$
\begin{aligned}
\mathcal{F}_{1,n} &\leqslant M\beta_n\gamma_n\frac{1}{2-\alpha\varepsilon} \\
&= O\left(\beta_n\gamma_n\right).
\end{aligned}
$$

For F_2, we use the Davydov inequality for mixing processes (see Rio 2000, p. 10, Formula 1.12a). This leads us to get, for all $i\neq j$

$$|\mathbb{C}ov\,(W_i,W_j)|\leqslant c\alpha\left(|i-j|\right).$$

Therefore, using **(A5)**, we obtain

$$\begin{aligned}
\mathcal{F}_{2,n} &\leqslant c\sum_{j=1}^{n}\sum_{\beta_n+1\leqslant k\leqslant n-1}\alpha\left(k\right)\\
&< cn\int_{\beta_n+1}^{n-1}k^{-v}dk\\
&= O\left(n\beta_n^{1-v}\right).
\end{aligned}$$

Choosing $\beta_n=\left(n\gamma_n^{-1}\right)^{1/\nu}$ and under **(A7)**(i), we obtain

$$S_n^{2*}=\mathcal{F}_{1,n}+\mathcal{F}_{2,n}=O\left(n^{1/\nu}\gamma_n^{1-1/\nu}\right)=o\left(1\right).$$

Finally, we get

$$S_n^2=\begin{cases}O\left(\gamma_n h_n^{-d}\right) & \text{if } a\geq\alpha/(d+4) \qquad (12)\\ o\left(h_n^4\right) & \text{if } a<\alpha/(d+4). \qquad (13)\end{cases}$$

As a matter of fact, we apply Lemma 5.3 in the case $a\geq\alpha/(d+4)$. We obtain, for any k

$$\begin{aligned}
\mathbb{P}\left\{\left|\pi_n\sum_{k=1}^{n}T_i(t_k^*)\right|>\varepsilon\right\} &\leqslant c\left(1+\frac{\varepsilon^2}{16rS_n^2}\right)^{-r/2}+ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{\nu+1}\\
&:= c\left(\Gamma_{1,n}+\Gamma_{2,n}\right).
\end{aligned}$$

By taking

$$\varepsilon=\varepsilon_0\left(\sqrt{\gamma_n h_n^{-d}\log n}\right)\quad \textit{and}\quad r=c\log n(\log_2 n)^{1/\nu} \qquad (14)$$

and using Taylor series expansion of $\log(1+x)$ as well as (12)-(14), we infer

$$\Gamma_{1,n}\leqslant cn^{-\varepsilon_0^2/2}$$

and

$$\Gamma_{2,n}\leqslant c\varepsilon_0^{-(\nu+1)}n\gamma_n^{-(\nu+1)/2}h_n^{d(\nu+1)/2}(\log n)^{(\nu-1)/2}\log_2 n$$

where $\log_2 n=\log(\log n)$ for $n>2$. Consequently,

$$\begin{aligned}
&\mathbb{P}\quad\left\{\max_{k=1,\cdots,\lambda_n}\left|\pi_n\sum_{k=1}^{n}T_i(t_k^*)\right|>\varepsilon_0\left(\sqrt{\gamma_n h_n^{-d}\log n}\right)\right\}\\
&\leqslant \sum_{i=1}^{\lambda_n}\mathbb{P}\left\{\left|\pi_n\sum_{k=1}^{n}T_i(t_k^*)\right|>\varepsilon_0\left(\sqrt{\gamma_n h_n^{-d}\log n}\right)\right\}\\
&\leqslant \lambda_n c\left\{\Gamma_{1,n}+\Gamma_{2,n}\right\}\\
&\leqslant lb_n^{-1}c\left\{\Gamma_{1,n}+\Gamma_{2,n}\right\}\\
&\leqslant lc\left\{n^{(\alpha-\varepsilon_0^2)/2}h_n^{-(2+d)/2}+\varepsilon_0^{-(\nu+1)}n^{\alpha(\nu+2)/2+1}h_n^{(d\nu-2)/2}(\log n)^{(\nu-1)/2}\log_2 n\right\}\\
&:= lc\left\{\tilde{\Gamma}_{1,n}+\tilde{\Gamma}_{2,n}\right\},
\end{aligned}$$

with

$$\tilde{\Gamma}_{1,n}:=b_n^{-1}\Gamma_{1,n}\quad \textit{and}\quad \tilde{\Gamma}_{2,n}:=b_n^{-1}\Gamma_{2,n}.$$

Now, referring to **(A7)**(ii), we have

$$h_n^{(d\nu-2)/2}=o\left(n^{-\alpha(\nu+2)/2-2}(\log n)^{-(\nu+1)/2}(\log_2 n)^{-3}\right),$$

which yields

$$\tilde{\Gamma}_{2,n}=o\left(\frac{1}{n\log n(\log_2 n)^2}\right),$$

corresponding to the general term of the convergent Bertrand series. For $\tilde{\Gamma}_{1,n}$, an appropriate choice of ε_0 can be made $O\left(n^{-3/2}\right)$, which corresponds to the general term of convergent series. Hence, $\sum_{n\geqslant 1}\left\{\tilde{\Gamma}_{1,n}+\tilde{\Gamma}_{2,n}\right\}<\infty$, and therefore (5) follows by applying Borel Cantelli Lemma. The same steps shall be used in the second case if $a<\alpha/(d+4)$. The result (6) is a consequence of Borel Cantelli Lemma after applying Lemma 5.3 and choosing

$$\varepsilon=\varepsilon_0 h_n^2\sqrt{\log n} \quad \textit{and} \quad r=c\log n(\log_2 n)^{1/\nu}.$$

Proof of Proposition 2.2. Standard argument yields

$$\begin{aligned} |f(\theta_n)-f(\theta)| &\leqslant |f(\theta_n)-f_n(\theta_n)|+|f_n(\theta_n)-f(\theta)| \\ &\leqslant \sup_{t\in\Omega}|f_n(t)-f(t)|+|f_n(\theta_n)-f(\theta)|. \end{aligned} \tag{15}$$

Since

$$|f_n(\theta_n)-f(\theta)|=\left|\sup_{t\in\Omega}f_n(t)-\sup_{t\in\Omega}f(t)\right|\leqslant \sup_{t\in\Omega}|f_n(t)-f(t)|,$$

then we have

$$|f(\theta_n)-f(\theta)|\leqslant 2\sup_{t\in\Omega}|f_n(t)-f(t)|. \tag{16}$$

The a.s. consistency of θ_n follows then immediately from (2.1) and **(A6)**. Now a Taylor expansion provides

$$\begin{aligned} f(\theta_n)-f(\theta) &= (\theta_n-\theta)f'(\theta)+\frac{1}{2}(\theta_n-\theta)^2 f^{(2)}(\theta_n^*) \\ &= \frac{1}{2}(\theta_n-\theta)^2 f^{(2)}(\theta_n^*), \end{aligned}$$

where θ_n^* is between θ and θ_n. Therefore, based on (16) and **(A3)**, we get

$$\begin{aligned} |\theta_n-\theta| &\leqslant \sqrt{\frac{2\,|f(\theta_n)-f(\theta)|}{\left|f^{(2)}(\theta_n^*)\right|}} \\ &\leqslant 2\sqrt{\frac{\sup_{t\in\Omega}|f_n(t)-f(t)|}{|\,f^{(2)}(\theta_n^*)\,|}}. \end{aligned}$$

Thus, by (2.1) the proof holds.

References

1. J. R. Blum, D. L. Hanson and L. H. Koopmans, *On the strong law of large numbers for a class of stochastic processes*, Z. Wahrsch. Ver. Geb. **2** (1963), 1–11.
2. D. Bosq, *Nonparametric statistics for stochastic processes: estimation and prediction*, Springer, New York, 1999.
3. S. Boukabour and A. Masmoudi, *Semiparametric Bayesian networks for continuous data*, Comm. Statist. Theory Methods (2020), doi.org/10.1080/03610926.2020.1738486.
4. R. Bradley, *Introduction to strong mixing conditions*, Kendrick Press, Heber City, 2007.
5. Y. Cheng, *Mean shift, mode seeking, and clustering*, IEEE Trans. Pattern Anal. Mach. Intell. **17** (1995), 790–799.
6. H. Chernoff, *Estimation of the mode*, Ann. Inst. Statist. Math. **16** (1964), 31–41.
7. G. Collomb, W. Hardle and S. Hassani, *A note on prediction via estimation of the conditional mode function*, J. Statist. Plann. Inference **15** (1987), 227–236.
8. J. Dedecker, P. Doukhan, G. Lang, J . R. Leon, S. Louhichi and C. Prieur, *Weak dependence with examples and applications*, Springer, Berlin, 2007.
9. P. Doukhan, *Mixing: properties and examples*, Springer, New York, 1994.
10. W. F. Eddy, *Optimum kernel estimators of the mode*, Ann. Statist. **8** (1980), 870–882.
11. W. F. Eddy, *The asymptotic distributions of kernel estimators of the mode*, Probab. Theory Related Fields **59** (1982), 279–290.

12. A. Gannoun and D. Louani, *Asymptotic properties of kernel estimators of the mode under censoring*, Technical Report University of Montpellier 11 **9** (1996), 96–04.
13. S. B. Hedges and P. Shah, *Comparison of mode estimation methods and application in molecular clock analysis*, BMC bioinformatics **4** (2003), 31.
14. H. O. Hirschfeld, *A connection between correlation and contingency*, Math. Proceed. of the Cambridge Philosophical Soc **31** (1935), 520–524.
15. M. W. Hofbaur and B. C. Williams, *Mode estimation of probabilistic hybrid systems*, In International Workshop on Hybrid Systems: Computation and Control (HSCC) (2002), 253–266.
16. I. A. Ibragimov, *Some limit theorems from stationary processes*, Theory Proc. Appl. **7** (1962), 349–382.
17. H. Jiang and S. Kpotufe, *Modal-set estimation with an application to clustering*, arXiv preprint arXiv:1606.04166, 2016.
18. A. N. Kolmogorov, *Uber die analytischen methoden inder wahrschein lichkeitsrechnung*, Math. Ann. **104** (1931), 415–458.
19. V. D. Konakov, *On the asymptotic normality of the mode of multidimensional distributions*, Theory Probab. Appl. **19** (1974), 794–799.
20. D. Louani, *On the asymptotic normality of the kernel estimators of the density function and its derivatives under censoring*, Comm. Statist. Theory Methods **27** (1998), 2909–2924.
21. D. Louani and E. Ould Saïd, *Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis*, J. Nonparametr. Stat. **11** (1999), 413–442.
22. E. Masry, *Recursive probability density estimation for weakly dependent process*, IEEE Trans. Inform. Theory **32** (1986), 254–267.
23. A. Mokkadem, M. Pelletier and Y. Slaoui, *The stochastic approximation method for the estimation of a multivariate probability density*, J. Statist. Plann. Inference **139** (2009), 2459–2478.
24. E. N. Nadaraya, *On nonparametric estimates of density functions and regression curves*, Theory Probab. Appl. **10** (1965), 186–190.
25. E. Ould Saïd, *Estimation non paramtrique du mode conditionnel application la prvision*, C. R. Math. Acad. Sci. Paris **316** (1993), 943–947.
26. E. Ould Saïd, *Strong uniform consistency of nonparametric estimation of the censored conditional mode function*, J. Nonparametr. Stat. **17** (2005), 797–806.
27. E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. **33** (1962), 1065–1076.
28. P. Révész, *Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I*, Studia Sci. Math. Hung. **8** (1973), 391–398.
29. P. Révész, *How to apply the method of stochastic approximation in the non-parametric estimation of a regression function*, Math. Operationsforsch. Statist., Ser. Statistics. **8** (1977), 119–126.
30. E. Rio, *Thorie asymptotique des processus alatoires faiblement dpendants. Mathmatiques et applications*, Springer, New York, 2000.
31. J. P. Romano, *On weak convergence and optimality of kernel density estimates of the mode*, Ann. Statist. **16** (1988), 629–647.
32. M. Rosenblatt, *A central limit theorem and a strong mixing condition*, Proc. Natl. Acad. Sci. USA **42** (1956), 43–47.
33. M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. **27** (1956), 832–837.
34. M. Samanta, *Nonparametric estimation of the mode of a multivariate density*, South African Statist. J. **7** (1973), 109–117.
35. S. A. Nezam Sarmadi and V. Venkatasubramanian, *Electromechanical mode estimation using recursive adaptive stochastic subspace identification*, IEEE Trans. Power Syst. **29** (2014), 349–358.
36. Y. A. Sheikh, E. A. Khan and T. Kanade, *Mode-seeking by medoidshifts*, In Proc. 11th IEEE Internat. Conf. on Computer Vision (2007), 1–8.
37. Y. Slaoui, *Large and moderate principles for recursive kernel density estimators defined by stochastic approximation method*, Serdica Math. J. **39** (2013), 53–82.
38. Y. Slaoui, *Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method*, J. Probab. Stat (2014), doi:10.1155/2014/739640.
39. Y. Slaoui, *Bias reduction in kernel density estimation*, J. Nonparametr. Stat. **30** (2018), 505–522.
40. Y. Slaoui, *Recursive nonparametric regression estimation for independent functional data*, Statist. Sinica **30** (2020), 417–437.
41. W. Tao, H. Jin and Y. Zhang, *Colour image segmentation based on mean shift and normalized cuts*, IEEE Trans. Syst., Man, Cybern. B, Cybern. **37** (2007), 1382–1389.

42. A. B. Tsybakov, *Recurrent estimation of the mode of a multidimensional distribution*, Probl. Inf. Transm. **8** (1990), 119–126.
43. J. Van Ryzin, *On strong consistency of density estimates*, South African Statist. J. **40** (1969), 1765–1772.
44. P. Vieu, *A note on density mode function*, Statist. Probab. Lett. **26** (1996), 297–307.
45. B. Williams, S. Chung and V. Gupta, *Mode estimation of model-based programs: monitoring systems with complex behavior*, In Proc. Int. Joint Conf. Artif. Intell. (IJCAI) (2001), 579–590.
46. P. Yin, H. -Y. Tourapis, A. Tourapis and J. Boyce, *Fast mode decision and motion estimation for JVT/H.264*, Proceedings of IEEE International Conference on Image Processing (2003), 853–856.

Univ. Monastir, Laboratoire analyse, géométrie et applications, FSM, Tunisie
E-mail address: fatma.b.khadher@gmail.com

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