On the consistency of mode estimate for spatially dependent data

Ahmad Younso

Present address: of F. Author

Metrika manuscript No.
(will be inserted by the editor)

On the consistency of mode estimate for spatially dependent data

Ahmad Younso

Received: date / Accepted: date

Abstract This paper is concerned with estimating the density mode for random field by kernel method under some α-mixing condition. The almost sure uniform convergence of the density estimator is proved. The rate of almost sure uniform convergence of the density gradient estimator is given under mild conditions. The unknown density is supposed unimodal and its mode is estimated by a kernel estimate. The strong consistency of the mode estimate is investigated and the rate of convergence is given. An optimal bandwidth selection procedure is proposed and a simulation study is used to obtain empirical results.

Keywords Random field · Density · Mode · Kernel estimate · Bandwidth · Consistency.

1 Introduction

The literature dealing with nonparametric estimation (density and regression functions) and classification using kernel method with spatially dependent data is extensive, see for example: [7,6], [22], [4] and [23]. The almost sure uniform consistency of estimators on a compact in $\mathbb{R}^d$ is studied by may authors, see for example [7,6] for the density estimator and [4] for the regression estimator. The almost sure uniform consistency of these estimators on the whole $\mathbb{R}^d$ is established in the independent case (see for example [16]) and in the temporally dependent case (see for example [5]) but to the best of our knowledge, it is still unexplored in the spatially dependent case. We may face many problems

A. Younso
(a) MISTEA, Universit Montpellier, INRAE, Institut Agro, Montpellier, France.
(b) Department of Mathematical Statistics, Faculty of Sciences, Damascus University, Damascus, Syria
Tel.: +33-760666251
E-mail: ahmad.younso@inrae.fr

in the nonparametric spatial statistics in which we need to extend uniform consistency of density estimators to whole $\mathbb{R}^d$, especially those related to the estimate of density modes. The mode estimates of density function remain unexplored in the spatially dependent case despite the existence of many fields where the knowledge of density modes is of great interest. For example, in unsupervised problems where modes are used as measure of typicality of a set of data. In particular, in modern applications, mode estimation is often used in clustering, with the modes representing cluster centers. There is an extensive literature on mode estimation in the independent (or temporally dependent) case, see the key references: [15], [9], [18], [2,1], [20] and [11]. Most of the existing works are concerned with the consistency of the estimators and rates achievable by various approaches. The common approaches consist of estimating the mode of an unknown unimodal density by maximizing an estimate of the density on $\mathbb{R}^d$ using kernel estimate. Our aim is to extend some consistency results related to the kernel estimates of both the density and the mode from the independent (or temporally dependent) case to the spatially dependent case under some mild conditions. A very important problem in kernel estimation problem is to choose the smoothing parameter that is called bandwidth or window. We propose an optimal bandwidth selection procedure.

2 Kernel estimates for the density and the mode by random field

Let $\{X_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{Z}^N}$ (with $N \geq 1$) be a random field (spatial process) defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in $\mathbb{R}^d$ (with $d \geq 1$). Assume that the random field is strictly stationary and that for each $\mathbf{i} \in \mathbb{Z}^N$, $X_{\mathbf{i}}$ has the same distribution as a variable X. A point $\mathbf{i} = (i_1, ..., i_N) \in \mathbb{Z}^N$ will be referred to as a site. For $\mathbf{n} = (n_1, ..., n_N) \in (\mathbb{N}^*)^N$, we denote by $\mathcal{I}_{\mathbf{n}}$ the rectangular region $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} \in \mathbb{Z}^N : \ 1 \leq i_k \leq n_k, \ \forall k = 1, ..., N\}$ on which we observe the above spatial process. We denote $\hat{\mathbf{n}} = n_1 \times ... \times n_N = \text{card}(\mathcal{I}_{\mathbf{n}})$ and we write $\mathbf{n} \to \infty$ if $\min_{1\leq k\leq N} n_k \to \infty$ and $\max_{1\leq i,j\leq N} |n_i/n_j| \leq C$ for some generic constant C such that $0 < C < \infty$. Suppose X has an unknown density f. We first consider the problem of estimating the density f based on a set of observations $\{X_{\mathbf{i}}, \ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}\}$. We define a kernel estimator $f_{\mathbf{n}}$ of f at the point $x = (x_1, ..., x_d) \in \mathbb{R}^d$ by

$$f_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} K\left(\frac{x - X_{\mathbf{i}}}{b_{\mathbf{n}}}\right), \tag{1}$$

where $K : \mathbb{R}^d \to \mathbb{R}$, the kernel, is a symmetric bounded density function and $b_{\mathbf{n}}$, the bandwidth, is a strictly positive number depending on $\mathbf{n}$ and such that $b_{\mathbf{n}} \to 0$ as $\mathbf{n} \to \infty$. Some asymptotic properties of the estimator (1) are studied by [22] and [7] under mixing condition defined later on. Assume that f is unimodal density and denote by $\boldsymbol{m}$ the mode of f. Let $\hat{\boldsymbol{m}}_{\mathbf{n}}$ be the estimator of $\boldsymbol{m}$ using the set of observations $\{X_{\mathbf{i}}, \ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}\}$. The estimator $\hat{\boldsymbol{m}}_{\mathbf{n}}$ is called kernel mode estimator (or empirical mode). The empirical mode $\hat{\boldsymbol{m}}_{\mathbf{n}}$ estimates $\boldsymbol{m}$ by maximizing $f_{\mathbf{n}}(x)$ on $\mathbb{R}^d$, *i.e.*,

$$\hat{\boldsymbol{m}}_{\mathbf{n}} \in \arg\max_{\mathbb{R}^d} f_{\mathbf{n}}. \tag{2}$$

In the independent (or temporally dependent) case, various work such as [15] and [18], [20] and [11] establish consistency results of the approach under regularity assumptions. In this paper, we will extend the strong consistency results to the spatially dependent case. For this aim, we first need to extend the almost sure uniform consistency of the density estimator established by [16] in the independent case to the spatially dependent case. To establish the consistency results, we propose a spatial dependence condition that is widely used in nonparametric functional estimation.

3 General assumptions

We will assume throughout the paper that the spatial process $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^N}$ satisfies the following mixing condition: there exists a function $\varphi : \mathbb{R} \to \mathbb{R}^+$ with $\varphi(t) \searrow 0$ as $t \to \infty$, such that whenever $E, E' \subset \mathbb{Z}^N$ with finite cardinality,

$$\begin{aligned}\alpha\big(\mathcal{B}(E), \mathcal{B}(E')\big) &:= \sup\{|\mathbb{P}(A\cap C) - \mathbb{P}(A)\mathbb{P}(C)|, A \in \mathcal{B}(E), C \in \mathcal{B}(E')\}\\ &\leq h(\mathrm{Card}(E), \mathrm{Card}(E'))\varphi(\mathrm{dist}(E, E')),\end{aligned}$$

where $\mathcal{B}(E)$ (resp. $\mathcal{B}(E')$) denotes the Borel σ-field generated by $(X_{\mathbf{i}})_{\mathbf{i}\in E}$ (resp. $(X_{\mathbf{i}})_{\mathbf{i}\in E'}$), $\mathrm{Card}(E)$ (resp. $\mathrm{Card}(E')$) the cardinality of E (E'), $\mathrm{dist}(E, E')$ the Euclidean distance between E and E', and $h : \mathbb{N}^2 \to \mathbb{R}_+$ is a symmetric positive function which is nondecreasing in each variable. It will be also assumed for simplicity that h satisfies

$$h(m_1, m_2) \leq L(m_1 + m_2)^{\xi}, \quad \forall m_1, m_2 \in \mathbb{N}^*, \tag{3}$$

For some $L > 0$ and $\xi \geq 0$. If $h \equiv 1$, the random field called strongly mixing. and $\|.\|$ denotes the Euclidean norm. They are satisfied by many spatial models. Examples can be found in [14], [19] and [10].
We suppose also that $\varphi(t)$ tends to zero at a polynomial rate, *i.e*,

$$\varphi(t) = O(t^{-\theta}), \tag{4}$$

for some $\theta > 0$. We suppose the following assumptions hold.

H1. $f(x) \to 0$ if $\|x\| \to \infty$, with $\|.\|$ denotes the Euclidean norm.
H2. For each $\mathbf{i} \neq \mathbf{j}$, $(X_{\mathbf{i}}, X_{\mathbf{j}})$ has a density $f_{\mathbf{i},\mathbf{j}}$ such that

$$\sup_{u,v\in\mathbb{R}^d} |f_{\mathbf{i},\mathbf{j}}(u, v) - f(u)f(v)| \leq C, \quad \text{for some } \; C > 0.$$

H3. $\int_{\mathbb{R}^d} \|x\|^2 f(x)dx < \infty$.
H4. K satisfies a Lipschitz condition, *i.e*, there exists $R > 0$ such that for all $x, y \in \mathbb{R}^d$, $|K(x) - K(y)| < R\|x - y\|$.

H5. $\sup_{x\in\mathbb{R}^d} \|x\|^{d+1}K(x) < \infty$.
H6. The density f satisfies a Lipschitz condition.

Note that Assumption **H1** holds for example if f is uniformly continuous. It is useless as soon as the support of f is bounded. Assumption **H2**, used by [7], controls the dependency through the distance between $f_{\mathbf{i},\mathbf{j}}(u,v)$ and $f(u)f(v)$ and can be linked with the mixing condition. Assumptions **H3-H4** are classical in nonparametric functional estimation. Assumption **H5** is a particular case of Parzen-Rosenblatt condition that will be defined later. Assumption **H6** is used by [7] to establish the uniform convergence rate of the kernel density estimator on a compact set in $\mathbb{R}^d$.

4 Main results

4.1 Almost sure uniform convergence of the density estimator

We denote by $\boldsymbol{B}(x_0, r') = \{x \in \mathbb{R}^d : \|x - x_0\| \le r'\}$ the closed ball centered at x with radius $r' > 0$. Let $g(\mathbf{n}) = \prod_{i=1}^N (\log n_i)(\log\log n_i)^{1+\epsilon}$ for some $\epsilon > 0$. It is well known that

$$\sum_{\mathbf{n}\in\mathbb{N}^{*N}} 1/(\hat{\mathbf{n}}g(\mathbf{n})) < \infty. \tag{5}$$

The kernel K is called Parzen-Rosenblatt kernel if:

$$\lim_{\|x\|\to\infty} \|x\|^d K(x) = 0. \tag{6}$$

Condition (6) is satisfied by many kernels such as: naive kernel, Gaussian kernel and Epanechnikov kernel. It is used to obtain the uniform convergence of $\mathbb{E}f_{\mathbf{n}}$ toward f on a compact set D of $\mathbb{R}^d$. Note that (6) may be satisfied if **H5** holds. Define

$$\theta_1^* = \frac{d(\theta + 4N(d+1))}{\theta - 2N(\xi+1)} \quad \text{and} \quad \theta_2^* = -\frac{\theta}{\theta - 2N(\xi+1)}.$$

Theorem 1 *Suppose that* **H1-H5**, *(3) and (4) hold with $\theta > 2N$ and that f is bounded. If as $\mathbf{n}\to\infty$,*

$$b_{\mathbf{n}} \to 0, \;\; \hat{\mathbf{n}}b_{\mathbf{n}}^d/(\log\hat{\mathbf{n}}) \longrightarrow \infty \tag{7}$$

and

$$\hat{\mathbf{n}}b_{\mathbf{n}}^{\theta_1^*}(\log\hat{\mathbf{n}})^{\theta_2^*}\, g(\mathbf{n})^{-2N/(\theta-2N(\xi+1))} \longrightarrow \infty, \tag{8}$$

then, as $\mathbf{n}\to\infty$,

$$\sup_{x\in\mathbb{R}^d} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)| \longrightarrow 0 \;\; a.s.$$

[16] proves the result of Theorem 1 in the independent case using techniques of empirical processes which are not developed for spatial processes. Theorem 1 extends the uniform consistency on a compact set established by [7] to whole $\mathbb{R}^d$. Condition (7) is a spatial version of the *i.i.d.* one (see [17]) that can be obtained from (8) as $\theta \to \infty$.
The proofs of the following theorems are immediate consequences of Bochner's Lemma.

Theorem 2 *Let $D \subset \mathbb{R}^d$ be a compact set. Suppose that (6) is satisfied and that f is continuous. Then as $\mathbf{n} \to \infty$,*

$$\sup_{x \in D} |\mathbb{E} f_{\mathbf{n}}(x) - f(x)| \longrightarrow 0.$$

Theorem 3 *Suppose that f is uniformly continuous on $\mathbb{R}^d$. Then, as $\mathbf{n} \to \infty$,*

$$\sup_{x \in \mathbb{R}^d} |\mathbb{E} f_{\mathbf{n}}(x) - f(x)| \longrightarrow 0.$$

The proof of the following corollary is immediate from Theorem 1 and Theorem 3.

Corollary 1 *Suppose that* **H1-H5***, (6), (8), (3) and (4) hold with $\theta > 2N$ and that f is bounded and uniformly continuous. Then, as $\mathbf{n} \to \infty$,*

$$\sup_{x \in \mathbb{R}^d} |f_{\mathbf{n}}(x) - f(x)| \longrightarrow 0 \;\; a.s.$$

The proof of the following corollary is immediate from Theorem 1 and Theorem 2.

Corollary 2 *Let $D \subset \mathbb{R}^d$ be a compact set. Suppose that (6) is satisfied and that f is continuous. Under assumptions of Theorem 1, we have, as $\mathbf{n} \to \infty$,*

$$\sup_{x \in D} |f_{\mathbf{n}}(x) - f(x)| \longrightarrow 0 \;\; a.s.$$

4.2 **Strong consistency of the kernel mode estimator**

Our aim in this section is to establish some consistency results related to the empirical mode $\hat{\boldsymbol{m}}_{\mathbf{n}}$ defined in (2). We first prove the almost sure convergence of $\hat{\boldsymbol{m}}_{\mathbf{n}}$ towards the exact mode $\boldsymbol{m}$. Then, we study the rate of almost sure convergence for this estimator under smoothness conditions on f. The strong consistency of $\hat{\boldsymbol{m}}_{\mathbf{n}}$ in the independent univariate case is investigated by [18]. [13] extend the result of [18] to the multivariate case. More recently, the strong consistency of the kernel mode estimator is established by [11] in the ψ-weakly dependent case. In the following theorem, we investigate the strong consistency of the empirical mode in the spatially dependent case.

Theorem 4 *Suppose that the assumptions of Theorem 1 are verified and that the density f is continuous in a neighborhood of the mode $\boldsymbol{m}$. If for any $\delta > 0$,*

$$\sup_{x \in \boldsymbol{B}(\boldsymbol{m},\delta)^c} f(x) < f(\boldsymbol{m}). \tag{9}$$

then, as $\mathbf{n} \to \infty$,

$$\hat{\boldsymbol{m}}_{\mathbf{n}} \longrightarrow \boldsymbol{m} \;\; a.s.$$

Note that condition (9) is in line with the assumption that the density f is unimodal. Now, we study the strong convergence rate of $\boldsymbol{m}_{\mathbf{n}}$. We will prove that the convergence rate is

$$\psi_{\mathbf{n}} = \left(\hat{\mathbf{n}} b_{\mathbf{n}}^{d+2} / \log \hat{\mathbf{n}}\right)^{-1/2}. \tag{10}$$

The rate (10) is the same as [11] and [20] in the univariate case ($d = 1$). We assume that f is unimodal and has its mode $\boldsymbol{m}$ in the compact $D \subset \mathbb{R}^d$, *i.e.*,

$$\boldsymbol{m} \in \arg\max_D f. \tag{11}$$

In this case the kernel mode estimator is given by

$$\hat{\boldsymbol{m}}_{\mathbf{n}} \in \arg\max_D f_{\mathbf{n}}. \tag{12}$$

Condition (11) is used by [11] in the univariate case. We denote ∇f and $\nabla^2 f$ the gradient and Hessian of the density f, respectively. Then, we have $\nabla f(\boldsymbol{m}) = 0$. We assume the following assumptions hold.

- **H′1**. f is differentiable function of order 3 with all partial derivatives bounded.
- **H′2**. $\frac{\partial f(x)}{\partial x_k}$ is Lipschizian (for each $k = 1, ..., d$).
- **H′3**. $\nabla^2 f(\boldsymbol{m})$ is negative definite.
- **H′4**. K is differentiable such that $\frac{\partial K(x)}{\partial x_k}$ is Lipschizian (for each $k = 1, ..., d$).
- **H′5**. $\frac{\partial K(x)}{\partial x_k}$ is bounded and integrable (for each $k = 1, ..., d$).
- **H′6**. $\lim_{\|x\| \to \infty} K(x) = 0$.
- **H′7**. $\int_{\mathbb{R}^d} \|x\|^2 K(x) dx < \infty$.

Hypotheses **H′1**-**H′6** are classical to establish different asymptotic properties of the kernel mode estimator. For these types of hypotheses, see for example: [13], [11] and [8]. For instance, **H′3** is supposed to consider only the interior mode. **H′6** is weaker than Parzen-Rosenblatt condition. Assumption **H′7** is used by [5] in the temporally dependent case.
Before we state the rate of convergence for the mode estimator (12), we first need to investigate the rate of uniform convergence of $\nabla f_{\mathbf{n}}$, the gradient of $f_{\mathbf{n}}$, on a compact set. Define

$$\theta_3^* = \frac{(d+2)\theta + Nd(d+1)}{\theta - N(d+3+2\xi)} \quad \text{and} \quad \theta_4^* = \frac{N(d+1) - \theta}{\theta - N(d+3+2\xi)}.$$

Lemma 1 *Let $D \subset \mathbb{R}^d$ be a compact set. Suppose that* $\mathbf{H'1}$-$\mathbf{H'2}$*,* $\mathbf{H'4}$*,* $\mathbf{H'6}$-$\mathbf{H'7}$ *and (6) are satisfied. If as $\mathbf{n} \to \infty$,*

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{d+6} / \log \hat{\mathbf{n}} \longrightarrow 0,$$

then, for each $k = 1, ..., d$,

$$\sup_{x \in D} \left| \mathbb{E} \frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right| = O\left(\psi_{\mathbf{n}}\right),$$

with

$$\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} = \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^{d+1}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \frac{\partial}{\partial x_k} K\left(\frac{x - X_{\mathbf{i}}}{b_{\mathbf{n}}}\right).$$

Theorem 5 *Let $D \subset \mathbb{R}^d$ be a compact set. Suppose that* $\mathbf{H2}$*,* $\mathbf{H'1}$-$\mathbf{H'2}$*,* $\mathbf{H'4}$-$\mathbf{H'5}$*, (3) and (4) hold with $\theta > 2N$ and that f is bounded. If as $\mathbf{n} \to \infty$,*

$$b_{\mathbf{n}} \to 0, \quad \hat{\mathbf{n}} b_{\mathbf{n}}^{d+2} / (\log \hat{\mathbf{n}}) \longrightarrow \infty, \tag{13}$$

and

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_3^*} \left(\log \hat{\mathbf{n}}\right)^{\theta_4^*} g(\mathbf{n})^{-2N/(\theta - N(\xi+2))} \longrightarrow \infty, \tag{14}$$

then, for each $k = 1, ..., d$,

$$\sup_{x \in D} \left| \frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \mathbb{E} \frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} \right| = O\left(\psi_{\mathbf{n}}\right) \quad a.s.$$

The proof of the following corollary is an immediate consequence of Lemma 1 and Theorem 5.

Corollary 3 *Let $D \subset \mathbb{R}^d$ be a compact set. Suppose that assumptions of Theorem 5 is verified. If in addition,* $\mathbf{H'5}$-$\mathbf{H'7}$ *are verified and as $\mathbf{n} \to \infty$,*

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{d+6} / \log \hat{\mathbf{n}} \longrightarrow 0,$$

then, for each $k = 1, ..., d$,

$$\sup_{x \in D} \left| \frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right| = O\left(\psi_{\mathbf{n}}\right) \quad a.s.$$

Note that our estimator of ∇f achieves the same rate as that of Theorem 1 in [11] in the temporal univariate case.

Lemma 2 *If f is twice differentiable and* $\mathbf{H'3}$ *is fulfilled, then, there exists $\epsilon > 0$ such that, for each $x \in \boldsymbol{B}(\boldsymbol{m}, \epsilon)$,*

$$C_1 \|x - \boldsymbol{m}\|^2 \leq f(\boldsymbol{m}) - f(x) \leq C_2 \|x - \boldsymbol{m}\|^2, \tag{15}$$

for some $0 < C_1 \leq C_2$.

The convergence rate of the empirical mode $\hat{\boldsymbol{m}}_{\mathbf{n}}$ towards the exact mode $\boldsymbol{m}$ is stated in the following theorem.

Theorem 6 *Suppose that* **H2**, **H′1**-**H′2**, **H′4**, *(3) and (4) hold with* $\theta > 2N$. *If as* $\mathbf{n} \to \infty$,

$$b_{\mathbf{n}} \to 0, \ \ \hat{\mathbf{n}} b_{\mathbf{n}}^{d+2}/(\log \hat{\mathbf{n}}) \longrightarrow \infty, \ \ \hat{\mathbf{n}} b_{\mathbf{n}}^{d+6}/\log \hat{\mathbf{n}} \longrightarrow 0,$$

and

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{\theta}} \left(\log \hat{\mathbf{n}}\right)^{\tilde{\theta}_2} g(\mathbf{n})^{-2N/(\theta-2N(\xi+1))} \longrightarrow \infty, \tag{16}$$

with $\tilde{\theta}_1 = \max\{\theta_1^*, \theta_3^*\}$ *and* $\tilde{\theta}_2 = \min\{\theta_2^*, \theta_4^*\}$, *then,*

$$|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}| = O\left(\psi_{\mathbf{n}}\right) \quad a.s.$$

Note that if (16) is satisfied, then (8) and (14) are immediate.

4.3 Numerical studies

First, a simulation study is conducted, then an application is image analysis is given.

4.3.1 Bandwidth selection and simulation study

In practice, the choice of a bandwidth $b_{\mathbf{n}}$ is a crucial problem to the kernel density. An unfavorable choice of $b_{\mathbf{n}}$ may lead to catastrophic error rates. Various techniques for the bandwidth selection have been developed for nonparametric kernel smoothing method. Among the different selection techniques to select the parameter $b_{\mathbf{n}}$, one can propose the asymptotic mean integrated squared error ($AMISE$) criterion. [3] shows that the kernel density estimator in the spatial case has exactly the same asymptotic mean integrated squared error as in the *i.i.d.* case, *i.e.*,

$$AMISE(b_{\mathbf{n}}) = \mathbb{E} \int_{\mathbb{R}^d} (f_{\mathbf{n}}(x) - f(x))^2 dx = \frac{b_{\mathbf{n}}^2}{4} \Gamma_d^2 + \frac{\int_{\mathbb{R}^d} K^2(u) du}{\hat{\mathbf{n}} b_{\mathbf{n}}^d},$$

where

$$\Gamma_d^2 = \int_{\mathbb{R}^d} \left(\sum_{1 \le i,j \le d} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \int_{\mathbb{R}^d} u_i u_j K(u) du \right)^2 dx.$$

Consequently, the asymptotically optimal bandwidth is given by

$$b_{\mathbf{n}opt} = \frac{d \int_{\mathbb{R}^d} K^2(u) du}{\Gamma_d^2} \hat{\mathbf{n}}^{-1/(d+4)}.$$

The solution is found by taking the derivative of the $AMISE(b)$ with respect to b and setting it equal to zero. By substituting $b_{\mathbf{n}opt}$ into the $AMISE$ expression, the optimal $AMISE$ rate is given by $O(\hat{\mathbf{n}}^{-4/(4+d)})$. Unfortunately, the optimal bandwidth depends on the unknown quantity Γ_d^2 since the partial derivatives of f of order two are unknown. [21] proposes to try the bandwidth computed by replacing f in the formula of Γ_d^2 by a normal density function with mean vector μ and variance-covariance matrix Σ. This normal density is called a reference density. We estimate μ and Σ from data and get the reference density. In the following simulation study, we consider the univariate case and we suppose K is the normal kernel. In this case, one can easily verify that the asymptotically optimal bandwidth b_{opt} can be estimated by $\widehat{b}_{opt} = 1.06\hat{\sigma}\hat{\mathbf{n}}^{-1/5}$, with $\hat{\sigma}$ is the standard deviation estimated from data. Consider $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^N}$ is a real Gaussian random field with $\mathbb{E}X_{\mathbf{i}} = 0$ and $\text{var}X_{\mathbf{i}} = 2$ for each $\mathbf{i} \in \mathbb{Z}^N$. Suppose the covariance function is given by $c(\|\mathbf{i}-\mathbf{j}\|) = \text{cov}(X_{\mathbf{i}}, X_{\mathbf{j}}) = \|\mathbf{i}-\mathbf{j}\|^{-5}$, for any $\mathbf{i} \neq \mathbf{j}$. Set $\mathbf{n} = (n, n)$. For the density estimate, we let n take the three values 10, 20 and 30 to show how the estimation improves when n increases (see Figure 1). For the mode estimate, we let n varies from 1 to 30. To show how the estimated values vary around the exact mode, the sites are ordered and enumerated according to lexicographic order from the site $(1,1)$ to site $(30,30)$. Recall that the origin here is $(1,1)$ not $(0,0)$. Hence, for each site (s,t) with $1 \leq s,t \leq 30$, in the one hand, a number $k \in \{1, ..., 900\}$ is assigned and in the other hand, the estimation of the mode is determined based on the kernel estimate of the density constructed using the set of observations $\{X_{(i,j)}, 1 \leq i \leq s, 1 \leq j \leq t\}$. Figure 2 shows the variation of $\hat{\boldsymbol{m}}_{\mathbf{n}}$ as a function of k. We observe that the larger the value of k, the closer the estimated value is to the exact mode.

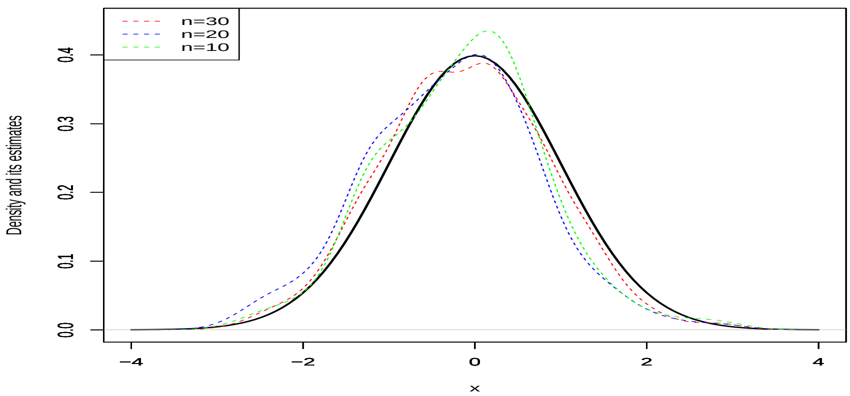

Fig. 1 The best estimated curve corresponds to the largest n where the black solid curve represents the exact density.

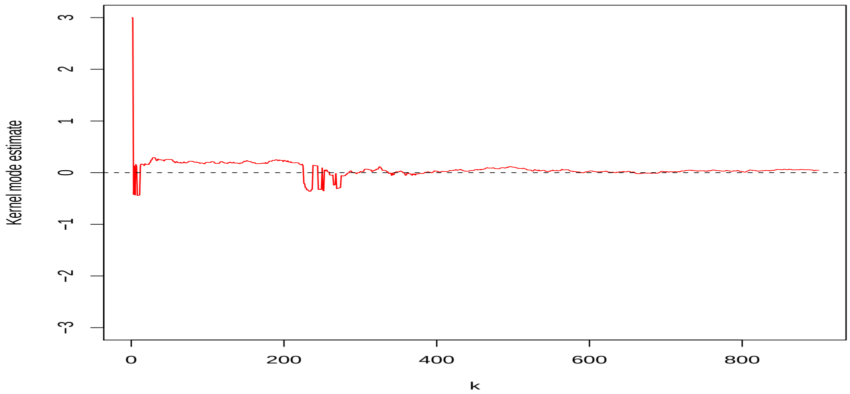

Fig. 2 The larger k, the closer the estimated value is to the exact mode ($\boldsymbol{m} = 0$).

4.3.2 Numerical application

One of the features that better describes a density estimate is the list of its modes and the intervals of values around which data concentrate. For example, the density estimate of intensities of an image made of different zones shall exhibit different peaks, each one of them ideally corresponding to a different region in the image. In this case, a proper segmentation of the image can be obtained by computing the appropriate thresholds that separate the modes in the density estimate. Most threshold selection algorithms assume that the intensity density is multi-modal; typically bimodal. However, some types of images are essentially unimodal since a much larger proportion of just one class of pixels (e.g. the background) is present in the image, and dominates the density. In this numerical study, we will see how to segment objects from a background. We use the coins image from R package *imager* (see Figure 3).

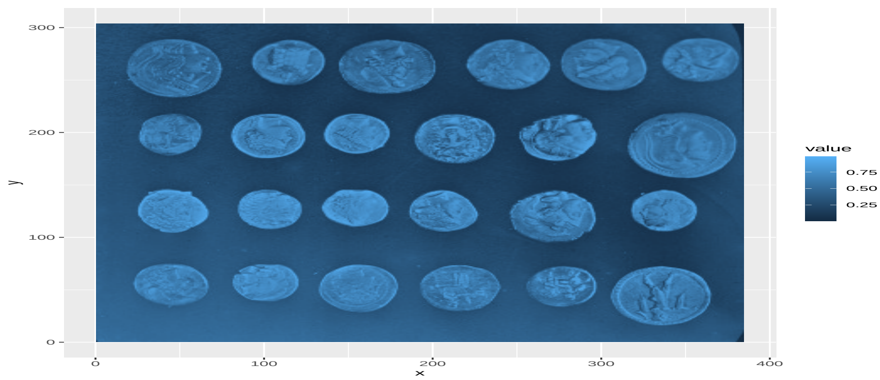

Fig. 3 Coins image with width $n_1 = 384$ and height $n_2 = 303$.

This image shows several coins outlined against a darker background. The number of pixels along the x-axis is called the width ($n_1 = 384$), along the y-axis its height ($n_2 = 303$). However, the image is composed of $n_1 \times n_2 = 116352$ pixels (sites) and in each pixel the intensity (gray level) is measured. Observe that the image data has class "cimg" which is converted into "data.frame". Here are the first six lines of the data table of the coins image.

x	y	Intensity
1	1	0.1843137
2	1	0.4823529
3	1	0.5215686
4	1	0.5058824
5	1	0.5372549
6	1	0.5176471

This data table contains the intensity value in each pixel with the pixels coordinates. The density function of the intensity is estimated by the kernel estimate which is maximized to get the estimated mode ($\hat{\boldsymbol{m}}_{\mathbf{n}} = 0.17$). The bandwidth is determined according to the method defined above ($b_{\mathbf{n}} = 0.02133$). Figure 4 shows the estimated density curve.

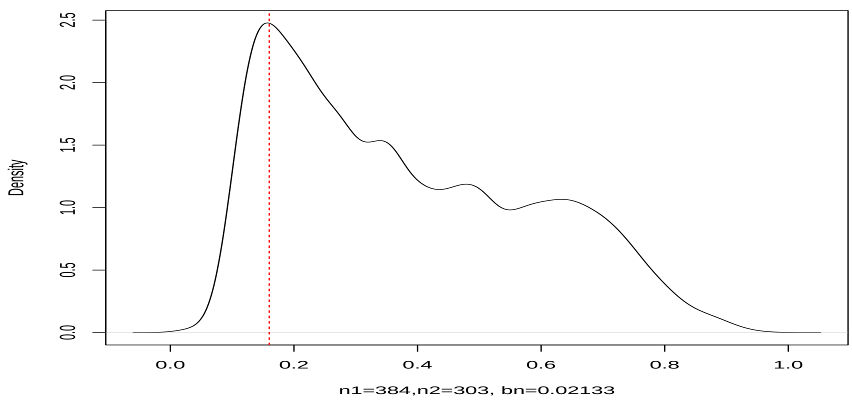

Fig. 4 Estimated density of the intensity (solid line) with the location of the estimated mode (point of intersection of the dashed line with the x-axis, $x = 0.17$).

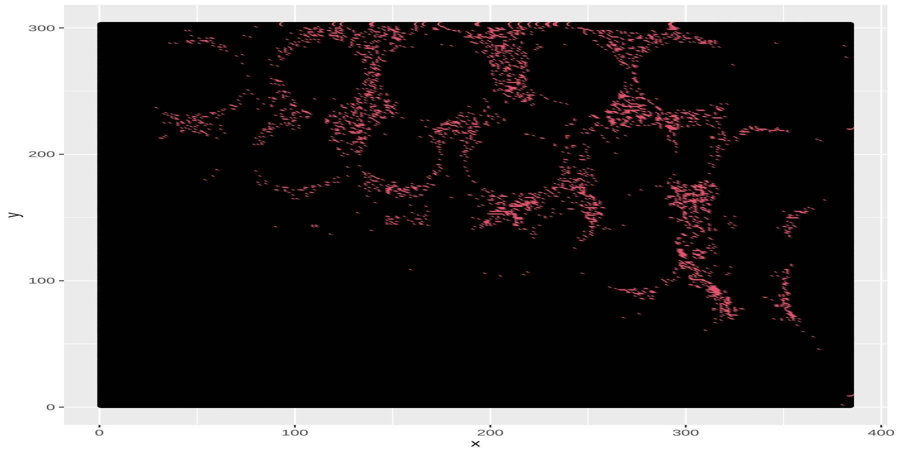

Fig. 5 Locations of the estimated mode on the image pixels (pink color).

Figure 5 dispays the pixels where the estimated mode is observed. It is clear that the majority of mode pixels are located in the background area. Since the background area is not totally covered by the pixels of the mode, it means that it is not uniform. This implies that to segment the image into objects and a background, we will need to determine a threshold based on the density curve. As shown on Figure 4, the estimated density has two significant beaks that correspond to background and objects of interest, a peak around the mode $x = 0.17$, and a second, smaller peak near to $x = 0.65$. We choose as a threshold $x = 0.55$, this is the value that minimize $f_{\mathbf{n}}(x)$ on the interval $[0.17, 0.65]$ (see Figure 6).

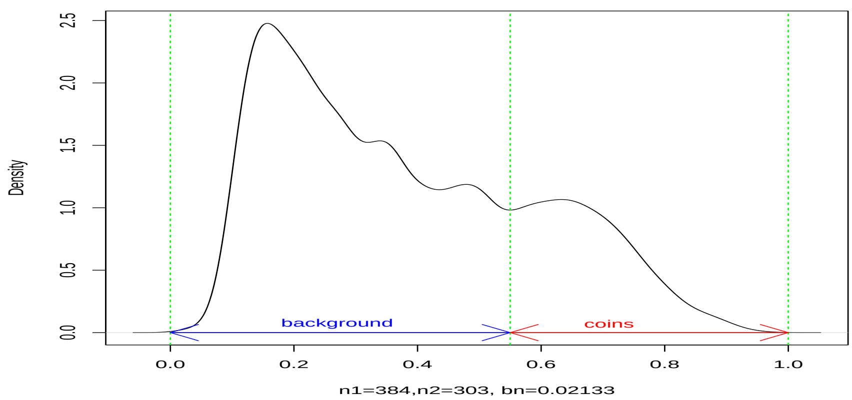

Fig. 6 Threshold of the intensity values based on local maximas of the density estimation.

According to this choice of threshold, we have Figure 7 that displays the segmentation of image into two zones, a background zone in pink color and objects zone in black color.

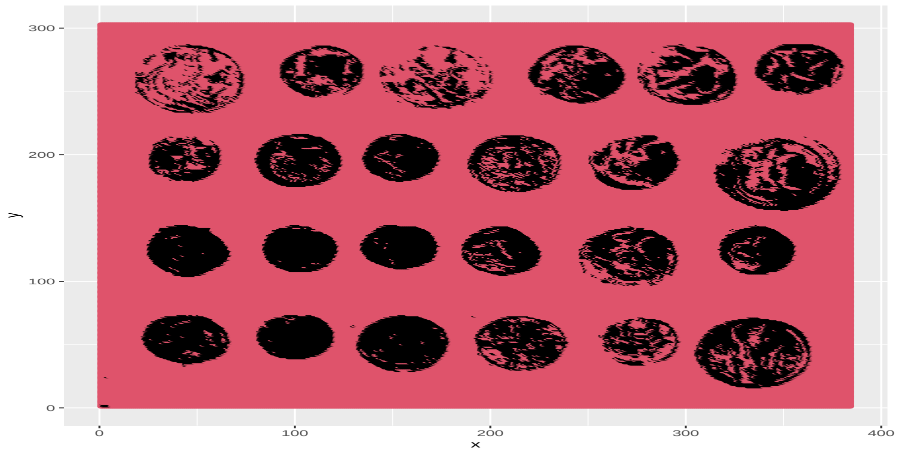

Fig. 7 Image segmentation in two zones: a background zone (pink color) and objects zone (black color).

From Figure 7, we can conclude that the objects (coins) can well be detected in the thresholded image. The slight overlap between the background and the coins is due to the fact that the background is not uniform.

5 Discussion on the choice of d and N

With regard to the kernel density estimator, it works well for low-dimensional problems, but are not effective for high dimensional problems. As noted by many authors, kernel methods suffers from the curse of dimensionality caused by the sparsity of data in high dimensional spaces. Or in other words, there will be only very few neighboring data points to any value x in a higher dimensional space, unless the sample size is extremely large. It has been shown that the best possible $AMISE$ rate is $O(\hat{\mathbf{n}}^{-4/(4+d)})$ which slightly increases as d increases. We believe that estimation of the density by kernel method is feasible in as many as six dimensions. Concerning the mode estimate, for a high dimensional sample space, in practice, the *argmax* is usually computed over a finite grid, but the grid size exponentially increases with the dimension d, which leads to time-consuming computations. However, for graphical exploratory purposes, it suffices to deal with $d \leq 3$. With regard to N, if we let for example $n_1 = n_2 = ... = n_N$, then $\hat{\mathbf{n}} = n^N$ and the $AMISE$ rate is $O(n^{-4N/(4+d)})$. This means that the $AMISE$ rate decreases exponentially to 0 as N increases. In practice, it is reasonable to take $N \leq 3$ for the spatial case and $N \leq 4$ for the spatio-temporal case.

6 Proofs

Before we start the proofs, we introduce the known spatial block decomposition of [7] which we will use several times. Let $\{Z_{\mathbf{i}},\ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}\}$ be any set of random variables observed on the region $\mathcal{I}_{\mathbf{n}}$. Without loss of generality, we assume that $n_k = 2q_k p$ for $k = 1, ..., N$ where q_k and p are positive integers. According to the block decomposition of [7], the random variables $Z_{\mathbf{i}}$ can be regrouped into $2^N q_1 \times ... \times q_N$ cubic blocks of side p. Denote

$$U(1, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = 2j_k p + 1 \\ k=1,...,N}}^{(2j_k+1)p} Z_{\mathbf{i}}$$

$$U(2, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = 2j_k p + 1 \\ k=1,...,N-1}}^{(2j_k+1)p} \sum_{i_N = (2j_N+1)p+1}^{2(j_N+1)p} Z_{\mathbf{i}}$$

$$U(3, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = 2j_k p + 1 \\ k=1,...,N-2}}^{(2j_k+1)p} \sum_{i_{N-1} = (2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N = 2j_N p+1}^{(2j_N+1)p} Z_{\mathbf{i}}$$

...

etc

...

$$U(2^N - 1, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = (2j_k+1)p + 1 \\ k=1,...,N-1}}^{2(j_k+1)p} \sum_{i_N = 2j_N p+1}^{(2j_N+1)p} Z_{\mathbf{i}}$$

$$U(2^N, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = (2j_k+1)p + 1 \\ k=1,...,N}}^{2(j_k+1)p} Z_{\mathbf{i}}.$$

For each integer $i = 1, ..., 2^N$, we define

$$T(\mathbf{n}, i) = \sum_{\substack{j_k = 1 \\ k=1,...,N}}^{q_k - 1} U(i, \mathbf{n}, \mathbf{j}).$$

Thus,

$$\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i}} = \sum_{i=1}^{2^N} T(\mathbf{n}, i). \tag{17}$$

Furthermore, for each $i = 1, ..., 2^N$,, $T(\mathbf{n}, i)$ is a sum of $r = q_1 \times ... \times q_N$ of the $U(i, \mathbf{n}, \mathbf{j})$'s. If for example we let $i = 1$, then $T(\mathbf{n}, 1)$ is a sum of $r = q_1 \times ... \times q_N$ of the $U(1, \mathbf{n}, \mathbf{j})$'s. The random term $U(1, \mathbf{n}, \mathbf{j})$ is measurable with the σ-field generated by $Z_{\mathbf{i}}$ with $\mathbf{i}$ belonging to the set of sites

$$\mathcal{S}_{\mathbf{j}} = \{\mathbf{i} : 2j_k p + 1 \le i_k \le (2j_k + 1)p, k = 1, ..., N\}. \tag{18}$$

For different values of $\mathbf{j} = (j_1, ..., j_N)$, the sets of sites (18) are separated by a distance of at least p, *i.e*,

$$\mathrm{dist}(\mathcal{S}_{\mathbf{j}}, \mathcal{S}_{\mathbf{j}'}) \geq p \quad \text{for any } \mathbf{j} \neq \mathbf{j}'. \tag{19}$$

Proof of Theorem 1 Let $c_1 = \sup_{x\in\mathbb{R}^d} K(x)$ and $c_2 = 2\mathbb{E}\|X\|$. We denote $a_{\mathbf{n}} = 8c_1c_2/(\epsilon b_{\mathbf{n}}^d)$ for some arbitrary number $\epsilon > 0$. Then,

$$\sup_{x\in\mathbb{R}^d} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)\| \leq \sup_{\|x\|\leq a_{\mathbf{n}}} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)| + \sup_{\|x\|> a_{\mathbf{n}}} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)|. \tag{20}$$

Consequently, we prove the theorem if we show that each term on the right-hand side of (20) tending almost surely to zero as $\mathbf{n} \to \infty$. We first show that, as $\mathbf{n} \to \infty$,

$$\sup_{\|x\|\leq a_{\mathbf{n}}} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)| \longrightarrow 0 \quad a.s. \tag{21}$$

To do so, we denote $h_{\mathbf{n}} = \epsilon b_{\mathbf{n}}^{d+1}/3R$ where $R > 0$ is the constant defined in **H4**, $s_{\mathbf{n}} = 2a_{\mathbf{n}}h_{\mathbf{n}}^{-1}$ and $\nu_{\mathbf{n}} = \lfloor s_{\mathbf{n}}^d \rfloor$ where $\lfloor . \rfloor$ denotes the integer part. Since $\boldsymbol{B}(0, a_{\mathbf{n}})$ is compact, it can be covered by $\nu_{\mathbf{n}}$ balls centered at $y_j,\ j = 1, ..., \nu_{\mathbf{n}}$ with radius $h_{\mathbf{n}}$. Taking into account the new conditions on $b_{\mathbf{n}}$,one can easily prove (21) by using the same argument as in (Theorem 3.3, [12]). It remains to show that as $\mathbf{n} \to \infty$,

$$\sup_{\|x\|> a_{\mathbf{n}}} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)| \longrightarrow 0 \quad a.s. \tag{22}$$

To do that, for any $\epsilon > 0$, we can write

$$\begin{aligned} &\mathbb{P}\left(\sup_{\|x\|> a_{\mathbf{n}}} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)| \geq \epsilon\right) \\ &\quad \leq \mathbb{P}\left(\sup_{\|x\|> a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) + \mathbb{P}\left(\sup_{\|x\|> a_{\mathbf{n}}} \mathbb{E}|f_{\mathbf{n}}(x)| \geq \epsilon/2\right). \end{aligned} \tag{23}$$

We will find an upper bound for each term on the right-hand side of (23). Let us first deal with the first term. Clearly,

$$\begin{aligned} \mathbb{P}\left(\sup_{\|x\|> a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) &\leq \mathbb{P}\left\{\sup_{\|x\|> a_{\mathbf{n}}} \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} \left| \sum_{\|x - X_{\mathbf{i}}\| > a_{\mathbf{n}}/2} K\Big(\frac{x - X_{\mathbf{i}}}{b_{\mathbf{n}}}\Big)\right| \geq \epsilon/4\right\} \\ &\quad + \mathbb{P}\left\{\sup_{\|x\|> a_{\mathbf{n}}} \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} \left| \sum_{\|x - X_{\mathbf{i}}\| \leq a_{\mathbf{n}}/2} K\Big(\frac{x - X_{\mathbf{i}}}{b_{\mathbf{n}}}\Big)\right| \geq \epsilon/4\right\}. \end{aligned}$$

However, by **H5**, for each $x \in \mathbb{R}^d$,

$$\begin{aligned}
\frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} & \Big| \sum_{\|x-X_{\mathbf{i}}\|>a_{\mathbf{n}}/2} K\Big(\frac{x-X_{\mathbf{i}}}{b_{\mathbf{n}}}\Big)\Big| \\
& \leq \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} \sum_{\|x-X_{\mathbf{i}}\|>a_{\mathbf{n}}/2} \left\{\frac{\|x-X_{\mathbf{i}}\|}{b_{\mathbf{n}}}\right\}^{d+1} \Big| K\Big(\frac{x-X_{\mathbf{i}}}{b_{\mathbf{n}}}\Big)\Big| \left\{\frac{b_{\mathbf{n}}}{\|x-X_{\mathbf{i}}\|}\right\}^{d+1} \\
& \leq \frac{C}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} \sum_{\|x-X_{\mathbf{i}}\|>a_{\mathbf{n}}/2} \left\{\frac{b_{\mathbf{n}}}{\|x-X_{\mathbf{i}}\|}\right\}^{d+1} \leq \frac{C}{a_{\mathbf{n}}^{d+1}},
\end{aligned}$$

with $C>0$ is generic constant. Since $a_{\mathbf{n}} \to \infty$, then for $\hat{\mathbf{n}}$ large enough, we have

$$\mathbb{P}\Big\{\sup_{\|x\|>a_{\mathbf{n}}} \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d}\Big| \sum_{\|x-X_{\mathbf{i}}\|>a_{\mathbf{n}}/2} K\Big(\frac{x-X_{\mathbf{i}}}{b_{\mathbf{n}}}\Big)\Big| \geq \epsilon/4\Big\} = 0$$

Hence, for $\hat{\mathbf{n}}$ large enough,

$$\mathbb{P}\left(\sup_{\|x\|>a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) \leq \mathbb{P}\Big\{\sup_{\|x\|>a_{\mathbf{n}}} \frac{1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d}\Big| \sum_{\|x-X_{\mathbf{i}}\|\leq a_{\mathbf{n}}/2} K\Big(\frac{x-X_{\mathbf{i}}}{b_{\mathbf{n}}}\Big)\Big| \geq \epsilon/4\Big\}.$$

Since for $\|x\| > a_{\mathbf{n}}$ and $\|x - X_{\mathbf{i}}\| \leq a_{\mathbf{n}}/2$, $\|X_{\mathbf{i}}\| \geq a_{\mathbf{n}}/2$, then,

$$\mathbb{P}\left(\sup_{\|x\|>a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) \leq \mathbb{P}\Big\{\frac{c_1}{\hat{\mathbf{n}} b_{\mathbf{n}}^d} \sum_{\mathbf{i}\in I_{\mathbf{n}}} \mathbb{I}_{\{\|X_{\mathbf{i}}\|>a_{\mathbf{n}}/2\}} \geq \epsilon/4\Big\},$$

where $\mathbb{I}_A$ denotes the indicator function of the set A. Markov's inequality yields

$$\mathbb{P}\big(\|X_{\mathbf{i}}\| \geq a_{\mathbf{n}}/2\big) \leq 2a_{\mathbf{n}}^{-1}\mathbb{E}\|X_{\mathbf{i}}\| \leq C b_{\mathbf{n}}^d \to 0 \text{ as } \mathbf{n} \to \infty.$$

Therefore, for $\hat{\mathbf{n}}$ large enough, we can write

$$\mathbb{P}\left(\sup_{\|x\|>a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) \leq \mathbb{P}\left(\Big|\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \Delta_{\mathbf{i}}\Big| \geq \epsilon/8\right), \tag{24}$$

with, for each $\mathbf{i} \in \mathcal{I}_{\mathbf{n}}$,

$$\Delta_{\mathbf{i}} = c_1(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1}\left(\mathbb{I}_{\{\|X_{\mathbf{i}}\|\geq a_{\mathbf{n}}/2\}} - \mathbb{P}\left(\|X_{\mathbf{i}}\| \geq a_{\mathbf{n}}/2\right)\right).$$

Now, we apply the above block decomposition of [7] to the random variables $\Delta_{\mathbf{i}}$. For this aim, we let $Z_{\mathbf{i}} = \Delta_{\mathbf{i}}$ in the block decomposition. Then,

$$|Z_{\mathbf{i}}| \leq C(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} \text{ and } \mathbb{E}Z_{\mathbf{i}} = 0.$$

Thus, by (17),

$$\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i}} = \sum_{i=1}^{2^N} T(\mathbf{n}, i).$$

Then,

$$\mathbb{P}\left(\left|\sum_{\mathbf{i}\in I_{\mathbf{n}}} Z_{\mathbf{i}}\right| \geq \epsilon/8\right) \leq \sum_{i=1}^{2^N} \mathbb{P}\left(|T(\mathbf{n},i)| \geq \epsilon/2^{N+3}\right). \tag{25}$$

Therefore, it suffices to find an upper bound for

$$\mathbb{P}\left(|T(\mathbf{n},1)| \geq \epsilon/2^{N+3}\right).$$

To elaborate, enumerate the r.v.'s $U(1,\mathbf{n},\mathbf{j})$ and the corresponding sets of sites $\mathcal{S}_{\mathbf{j}}$ in an arbitrary manner (see (18)) and refer to them respectively as $V_1, V_2, ..., V_r$ and $\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_r$. Approximate $V_1, V_2, ..., V_r$ by the r.v.'s $V_1^*, V_2^*, ..., V_r^*$ using Lemma 4.5 of [7]. Therefore, for each $k = 1, ..., r$,

$$|V_k| \leq Cp^N(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} := M_{\mathbf{n}} \tag{26}$$

and by (19),

$$\sum_{i=1}^{r} \mathbb{E}|V_i - V_i^*| \leq 2rM_{\mathbf{n}} h((r-1)p^N, p^N)\varphi(p). \tag{27}$$

We have the following inequality

$$\begin{aligned}&\mathbb{P}\big(\big|T(\mathbf{n},1)\big| \geq \epsilon/2^{N+3}\big)\\ &\leq \mathbb{P}\left(\sum_{i=1}^{r}|V_i - V_i^*| \geq \epsilon/2^{N+4}\right) + \mathbb{P}\left(\left|\sum_{i=1}^{r} V_i^*\right| \geq \epsilon/\times 2^{N+4}\right)\end{aligned} \tag{28}$$

Markov's inequality yields

$$\mathbb{P}\Big(\sum_{i=1}^{r}|V_i - V_i^*| \geq \epsilon/(3\times 2^{N+1})\Big) \leq C\hat{\mathbf{n}}^{\xi} b_{\mathbf{n}}^{-d}\varphi(p), \tag{29}$$

and Bernstein's inequality yields

$$\mathbb{P}\left(\left|\sum_{i=1}^{r} V_i^*\right| \geq \epsilon/2^{N+4})\right) \leq 2\exp\left\{-\frac{\epsilon'^2}{4r\mathrm{var}\,\mathrm{V}_1 + 2\mathrm{M}'_{\mathbf{n}}\epsilon'}\right\} \tag{30}$$

where $\epsilon' = \epsilon/(3\times 2^{N+4})$. Let $\delta = b_{\mathbf{n}}^{-d/N}$, then,

$$\mathrm{var}\mathrm{V}_1 \leq \mathrm{p}^{\mathrm{N}}\mathrm{var}\,\mathrm{Z}_{\mathbf{1}} + \sum_{\mathbf{i},\mathbf{j}:\ \|\mathbf{i}-\mathbf{j}\|\geq\delta} |\mathrm{cov}(\mathrm{Z}_{\mathbf{i}},\mathrm{Z}_{\mathbf{j}})| + \sum_{\mathbf{i},\mathbf{j}:\ 0<\|\mathbf{i}-\mathbf{j}\|<\delta} |\mathrm{cov}(\mathrm{Z}_{\mathbf{i}},\mathrm{Z}_{\mathbf{j}})|. \tag{31}$$

We will find an upper bound for each term on the right-hand side of (31). By Markov's inequality,

$$\begin{aligned}\mathrm{var}Z_{\mathbf{1}} &\leq c_1^2(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-2}\mathbb{P}\big(\|X_{\mathbf{1}}\| \geq a_{\mathbf{n}}/2\big) \leq 2c_1^2 a_{\mathbf{n}}^{-1}(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-2}\mathbb{E}\|X_{\mathbf{1}}\|\\ &\leq C(\hat{\mathbf{n}}^2 b_{\mathbf{n}}^d)^{-1}.\end{aligned} \tag{32}$$

In the other hand, by Markov's inequality and **H3**, we have for any $\mathbf{i} \neq \mathbf{j}$,

$$
\begin{aligned}
&|\operatorname{cov}(Z_{\mathbf{i}}, Z_{\mathbf{j}})| \\
&= c_1^2(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-2} \Big| \mathbb{P}\Big(\|X_{\mathbf{i}}\| > a_{\mathbf{n}}/2, \|X_{\mathbf{j}}\| > a_{\mathbf{n}}/2\Big) - \mathbb{P}\big(\|X_{\mathbf{i}}\| > a_{\mathbf{n}}/2\big)\mathbb{P}\big(\|X_{\mathbf{j}}\| > a_{\mathbf{n}}/2\big) \Big| \\
&\leq 2c_1^2(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-2}\mathbb{P}\big(\|X_{\mathbf{i}}\| > a_{\mathbf{n}}/2\big) \\
&\leq 6c_1^2 a_{\mathbf{n}}^{-2}(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-2}\mathbb{E}\|X_{\mathbf{1}}\|^2 \leq C\hat{\mathbf{n}}^{-2}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\sum_{\mathbf{i},\mathbf{j}:\ \|\mathbf{i}-\mathbf{j}\| \geq \delta} |\operatorname{cov}(Z_{\mathbf{i}}, Z_{\mathbf{j}})| &\leq C\hat{\mathbf{n}}^{-2} \sum_{\substack{i_k: \|\mathbf{i}\| \geq \delta \\ k=1,\dots,N}} \varphi(\|\mathbf{i}\|) \\
&\leq Cp^N \hat{\mathbf{n}}^{-2} \sum_{i \geq \delta} i^{N-1}\varphi(i) \leq Cp^N \hat{\mathbf{n}}^{-2} \int_\delta^{+\infty} u^{N-\theta-1} du \\
&\leq Cp^N \hat{\mathbf{n}}^{-2}\delta^{\theta-N} \leq Cp^N(\hat{\mathbf{n}}^2 b_{\mathbf{n}}^d)^{-1}, \qquad (33)
\end{aligned}
$$

since $\delta = b_{\mathbf{n}}^{-d/N}$. Moreover, we can easily show that

$$
\sum_{\mathbf{i},\mathbf{j}:\ 0 < \|\mathbf{i}-\mathbf{j}\| < \delta} |\operatorname{cov}(Z_{\mathbf{i}}, Z_{\mathbf{j}})| \leq Cp^N(\hat{\mathbf{n}}^2 b_{\mathbf{n}}^d)^{-1}. \qquad (34)
$$

Combining (31)-(34), we have

$$
\mathrm{var V_1} \leq \mathrm{Cp^N}(\hat{\mathbf{n}}^2 \mathrm{b}_{\mathbf{n}}^{\mathrm{d}})^{-1} \qquad (35)
$$

Hence, by (26), (30) and (35), we obtain the following inequality

$$
\mathbb{P}\left(\left|\sum_{i=1}^r V_i^*\right| \geq \epsilon/2^{N+4})\right) \leq 2\exp\left(-C\hat{\mathbf{n}} b_{\mathbf{n}}^d\right). \qquad (36)
$$

Combining (24)-(25), (28)-(29) and (36), we get

$$
\mathbb{P}\left(\sup_{\|x\| > a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) \leq 2^{N+1}\exp\left(-C\hat{\mathbf{n}} b_{\mathbf{n}}^d\right) + C\hat{\mathbf{n}}^\xi b_{\mathbf{n}}^{-d}\varphi(p). \qquad (37)
$$

Choosing $p = \left(\hat{\mathbf{n}} b_{\mathbf{n}}^d/(\log \hat{\mathbf{n}})\right)^{1/2N}$, (7)-(8) yield

$$
\sum_{\mathbf{n} \in \mathbb{N}^{*N}} \mathbb{P}\left(\sup_{\|x\| > a_{\mathbf{n}}} |f_{\mathbf{n}}(x)| \geq \epsilon/2\right) < \infty. \qquad (38)
$$

It remains to show that the second term on the right-hand side of the inequality (23) vanishes as $\mathbf{n} \to \infty$, *i.e.*,

$$
\mathbb{P}\left(\sup_{\|x\| > a_{\mathbf{n}}} \mathbb{E}|f_{\mathbf{n}}(x)| \geq \epsilon/2\right) \longrightarrow 0. \qquad (39)
$$

For each $x \in \mathbb{R}^d$,

$$\begin{aligned}\mathbb{E}|f_{\mathbf{n}}(x)| &\leq \int_{\mathbb{R}^d} K(u) f(x - b_{\mathbf{n}} u) du \\ &= \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})} K(u) f(x - b_{\mathbf{n}} u) du + \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})^c} K(u) f(x - b_{\mathbf{n}} u) du. \end{aligned} \quad (40)$$

Thus, if $\|x\| > a_{\mathbf{n}}$ and $u \in \boldsymbol{B}(0, b_{\mathbf{n}}^{-1})$, then $\|x - b_{\mathbf{n}} u\| \geq a_{\mathbf{n}} - 1$. Consequently, by **H1**,

$$\begin{aligned}\sup_{\|x\| > a_{\mathbf{n}}} \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})} K(u) f(x - b_{\mathbf{n}} u) d &\leq \sup_{\|x\| > a_{\mathbf{n}} - 1} f(x) \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})} K(u) du \\ &\leq C \sup_{\|x\| > a_{\mathbf{n}} - 1} f(x) \longrightarrow 0 \text{ as } \mathbf{n} \to \infty. \end{aligned} \quad (41)$$

In the other hand, we have, by **H5**,

$$\begin{aligned} &\sup_{\|x\| > a_{\mathbf{n}}} \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})^c} K(u) f(x - b_{\mathbf{n}} u) du \\ &\leq \sup_u (\|u\|^{d+1} K(u)) \sup_{\mathbb{R}^d} f \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})} \frac{du}{\|u\|^{d+1}} \\ &\leq C \int_{\boldsymbol{B}(0, b_{\mathbf{n}}^{-1})} \frac{du}{\|u\|^{d+1}} \longrightarrow 0 \text{ as } \mathbf{n} \to \infty. \end{aligned} \quad (42)$$

Combining (40)-(42), we have

$$\mathbb{E}|f_{\mathbf{n}}(x)| \longrightarrow 0 \text{ as } \mathbf{n} \to \infty. \quad (43)$$

By (23), (38) and (43) together with Borel-Cantelli lemma, we get (22). Finally, by (20)-(22), the proof is completed. □

Proof of Theorem 4 Since

$$\mathbb{E} f_{\mathbf{n}}(x) = \int_{\mathbb{R}^d} K(u) f(x - b_{\mathbf{n}} u) du,$$

we have for any $\delta > 0$,

$$\begin{aligned} \sup_{x \in \boldsymbol{B}(\boldsymbol{m}, \delta)^c} \mathbb{E} f_{\mathbf{n}}(x) &\leq \sup_{x \in \boldsymbol{B}(\boldsymbol{m}, \delta)^c} \int_{\|u\| \leq 1} K(u) f(x - b_{\mathbf{n}} u) du + \\ &\qquad \sup_{x \in \boldsymbol{B}(\boldsymbol{m}, \delta)^c} \int_{\|u\| > 1} K(u) f(x - b_{\mathbf{n}} u) du \\ &\leq \sup_{x \in \boldsymbol{B}(\boldsymbol{m}, \delta)^c} \int_{\|u\| \leq 1} K(u) f(x - b_{\mathbf{n}} u) du + f(\boldsymbol{m}) \int_{\|u\| > 1} K(u) du. \end{aligned} \quad (44)$$

Hence, if $\|u\| \leq 1$ and $x \in \boldsymbol{B}(\boldsymbol{m}, \delta)^c$, since $b_{\mathbf{n}} \to 0$, then for $\hat{\mathbf{n}}$ large enough, $\|x - b_{\mathbf{n}}u - \boldsymbol{m}\| > \delta - b_{\mathbf{n}}\|u\| > \delta/2$. Consequently, by (9),

$$\sup_{x\in \boldsymbol{B}(\boldsymbol{m},\delta)^c} \int_{\|u\|\leq 1} K(u) f(x - b_{\mathbf{n}}u) du \leq \sup_{x\in \boldsymbol{B}(\boldsymbol{m},\delta/2)^c} f(x) \int_{\|u\|\leq 1} K(u) du$$
$$< f(\boldsymbol{m}) \int_{\|u\|\leq 1} K(u) du.$$

Then, (44) yields

$$\limsup_{\mathbf{n}} \sup_{x\in \boldsymbol{B}(\boldsymbol{m},\delta)^c} \mathbb{E} f_{\mathbf{n}}(x) < f(\boldsymbol{m}) \int_{\mathbb{R}^d} K(u) du = f(\boldsymbol{m}). \tag{45}$$

Since, by Theorem 1,

$$\sup_{x\in\mathbb{R}^d} |f_{\mathbf{n}}(x) - \mathbb{E} f_{\mathbf{n}}(x)| \longrightarrow 0 \;\; a.s.$$

by (45), we can write

$$\limsup_{\mathbf{n}} \sup_{\boldsymbol{B}(\boldsymbol{m},\delta)^c} f_{\mathbf{n}} < f(\boldsymbol{m}) \;\; a.s. \tag{46}$$

Since f is continuous in a neighborhood of $\boldsymbol{m}$, by Corollary 2, there exists $\delta_0 > 0$ such that

$$\sup_{x\in \boldsymbol{B}(\boldsymbol{m},\delta_0)} |f_{\mathbf{n}}(x) - f(x)| \longrightarrow 0 \;\; a.s. \tag{47}$$

By (46)-(47), we have for any $\delta \leq \delta_0$,

$$\limsup_{\mathbf{n}} \sup_{\boldsymbol{B}(\boldsymbol{m},\delta)^c} f_{\mathbf{n}} < \limsup_{\mathbf{n}} \sup_{\boldsymbol{B}(\boldsymbol{m},\delta)} f_{\mathbf{n}} \;\; a.s.$$

Finally, if we let δ tend to 0, the proof is completed. □

Proof of Theorem 5 We will show that

$$\sup_{x\in D} \left| \frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} \right| = O\left(\psi_{\mathbf{n}}\right), \tag{48}$$

with $\psi_{\mathbf{n}}$ is defined in (10). Choose $l_{\mathbf{n}} = b_{\mathbf{n}}^{d+2}\psi_{\mathbf{n}}$. Since D is compact, it can be covered by

$$\nu'_{\mathbf{n}} = \lfloor (b_{\mathbf{n}}^{d+2}\psi_{\mathbf{n}})^{-d} \rfloor \tag{49}$$

balls centered at $y_j,\ j = 1, ..., \nu'_{\mathbf{n}}$ with radius $l_{\mathbf{n}}$. Define

$$Q_{\mathbf{n},1} = \max_{1\leq t\leq \nu'_{\mathbf{n}}} \sup_{x\in D} \left| \frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \frac{\partial f_{\mathbf{n}}(y_t)}{\partial x_k} \right|,$$
$$Q_{\mathbf{n},2} = \max_{1\leq t\leq \nu'_{\mathbf{n}}} \sup_{x\in D} \left| \mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \mathbb{E}\frac{\partial f_{\mathbf{n}}(y_t)}{\partial x_k} \right|,$$
$$Q_{\mathbf{n},3} = \max_{1\leq t\leq \nu'_{\mathbf{n}}} \left| \frac{\partial f_{\mathbf{n}}(y_t)}{\partial x_k} - \mathbb{E}\frac{\partial f_{\mathbf{n}}(y_t)}{\partial x_k} \right|.$$

Hence, we can write

$$\sup_{x\in D}\left|\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k}-\mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k}\right|\leq Q_{\mathbf{n},1}+Q_{\mathbf{n},2}+Q_{\mathbf{n},3}. \tag{50}$$

Since by $\mathbf{H'4}$, $\frac{\partial K(x)}{\partial x_k}$ is Lipschzian, we get

$$\left|\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k}-\frac{\partial f_{\mathbf{n}}(y_t)}{\partial x_k}\right|\leq Cb_{\mathbf{n}}^{-(d+1)}\|x-y_t\|\leq Cb_{\mathbf{n}}^{-(d+1)}l_{\mathbf{n}}\leq C\psi_{\mathbf{n}}$$

for some generic constant $C>0$. Then,

$$Q_{n,1}=O(\psi_{\mathbf{n}})\text{ and }Q_{n,2}=O(\psi_{\mathbf{n}})\quad\text{a.s.} \tag{51}$$

To accomplish the proof of (48), it remains to show that $Q_{\mathbf{n},3}=O(\psi_{\mathbf{n}})$ almost surely, *i.e.*,

$$\max_{1\leq t\leq \nu'_{\mathbf{n}}}|S_{\mathbf{n}}(y_t)|=O(\psi_{\mathbf{n}})\quad a.s.\text{ for each } y\in\{y_1,...,y_{\nu'_{\mathbf{n}}}\}, \tag{52}$$

with

$$\begin{aligned}S_{\mathbf{n}}(y)&=\frac{\partial f_{\mathbf{n}}(y)}{\partial x_k}-\mathbb{E}\frac{\partial f_{\mathbf{n}}(y)}{\partial x_k}\\&=(\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-1}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\left\{\frac{\partial}{\partial x_k}K\left(\frac{y-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right)-\mathbb{E}\frac{\partial}{\partial x_k}K\left(\frac{y-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right)\right\}:=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\Delta'_{\mathbf{i}}(y),\end{aligned}$$

where

$$\Delta'_{\mathbf{i}}(y)=(\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-1}\left\{\frac{\partial}{\partial x_k}K\left(\frac{y-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right)-\mathbb{E}\frac{\partial}{\partial x_k}K\left(\frac{y-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right)\right\}.$$

Therefore, $\mathbb{E}\Delta'_{\mathbf{i}}(y)=0$ and $|\Delta'_{\mathbf{i}}(y)|\leq C(\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-1}$ by $\mathbf{H'5}$. Now, we apply the above block decomposition of [7] to the random variables $\Delta'_{\mathbf{i}}(y)$. For this aim, we set $Z_{\mathbf{i}}(y)=\Delta'_{\mathbf{i}}(y)$ in the block decomposition. Thus, by (17),

$$S_{\mathbf{n}}(y)=\sum_{i=1}^{2^N}T(\mathbf{n},i,y).$$

Here we let $U(i,\mathbf{n},\mathbf{j})$ and $T(\mathbf{n},i)$ in the block decomposition be functions of y since we will take the maximum of $T(\mathbf{n},i,y)$ with respect to y belonging to $y_1,...,y_{\nu'_{\mathbf{n}}}$. Consequently, to prove (48), it suffices to show for example

$$\max_{1\leq t\leq \nu'_{\mathbf{n}}}|T(\mathbf{n},1,y_t)|=O(\psi_{\mathbf{n}})\quad a.s. \tag{53}$$

To do that, for a fixed t chosen, enumerate the r.v.'s $U(1,\mathbf{n},\mathbf{j},y_t)$ and the corresponding sets of sites $\mathcal{S}_{\mathbf{j}}$ (see (18)) in an arbitrary manner and refer to them respectively as $W_1,W_2,...,W_r$ and $\mathcal{S}_1,\mathcal{S}_2,...,\mathcal{S}_r$. Approximate $W_1,W_2,...,W_r$

by the r.v.'s $W_1^*, W_2^*, ..., W_r^*$ using Lemma 4.5 of [7]. Clearly, since $U(1, \mathbf{n}, \mathbf{j}, y_t)$ is a sum of p^N random variables $Z_{\mathbf{i}}(y_t)$, for each $k = 1, ..., r$,

$$|W_k| \leq Cp^N(\hat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1}. \tag{54}$$

Let $\epsilon_{\mathbf{n}} = \eta\psi_{\mathbf{n}}$ where $\eta > 0$ is a constant to be chosen later. We have

$$\begin{aligned}&\mathbb{P}\big(\big|T(\mathbf{n}, i, y_t))\big| \geq \epsilon_{\mathbf{n}}\big)\\ &\leq \mathbb{P}\left(\sum_{i=1}^{r}|W_i - W_i^*| \geq \epsilon_{\mathbf{n}}/2\right) + \mathbb{P}\left(\left|\sum_{i=1}^{r} W_i^*\right| \geq \epsilon_{\mathbf{n}}/2\right)\end{aligned} \tag{55}$$

We will find an upper bound for each term on the right-hand side of (55). For the first term, we get by Markov's inequality together with (19), Lemma 4.5 of [7] and (54)

$$\mathbb{P}\Big(\sum_{i=1}^{r}|W_i - W_i^*| \geq \epsilon_{\mathbf{n}}/2)\Big) \leq C\epsilon_{\mathbf{n}}^{-1}\hat{\mathbf{n}}^{\xi}b_{\mathbf{n}}^{-(d+1)}\varphi(p). \tag{56}$$

For the second term, set $\lambda_{\mathbf{n}} = (\hat{\mathbf{n}}b^{d+2}\log\hat{\mathbf{n}})^{1/2}$. One can easily verify that for $\hat{\mathbf{n}}$ large enough, $|\lambda_{\mathbf{n}}W_1| < 1/2$ by (54). Furthermore, $\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} = \eta$. Applying Markov's inequality, we have by the strict stationarity

$$\mathbb{P}\left(\left|\sum_{i=1}^{r} W_i^*\right| \geq \epsilon_{\mathbf{n}}/2)\right) \leq 2\exp\left(-\lambda_{\mathbf{n}}(\epsilon_{\mathbf{n}}/2) + r\lambda_{\mathbf{n}}^2\text{var}W_1\right) \tag{57}$$

Let us find an upper bound to varW_1. To do that, set $\delta = b_{\mathbf{n}}^{-d/N}$. We can write

$$\text{varW}_1 \leq p^N \text{var}\,\Delta'_{\mathbf{1}}(\text{y}) \tag{58}$$

$$+ \sum_{\mathbf{i},\mathbf{j}:\ \|\mathbf{i}-\mathbf{j}\|\geq\delta}|\text{cov}(\Delta'_{\mathbf{i}}(y), \Delta'_{\mathbf{j}}(y))| + \sum_{\mathbf{i},\mathbf{j}:\ 0<\|\mathbf{i}-\mathbf{j}\|<\delta}|\text{cov}(\Delta'_{\mathbf{i}}(y), \Delta'_{\mathbf{j}}(y))|. \tag{59}$$

Since f is bounded by assumption, we have by $\mathbf{H'5}$

$$\begin{aligned}\text{var}\Delta'_{\mathbf{1}}(y) &= (\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-2}\text{var}\left\{\frac{\partial}{\partial x_k}K\Big(\frac{y - X_{\mathbf{1}}}{b_{\mathbf{n}}}\Big)\right\}\\ &\leq (\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-2}\mathbb{E}\left\{\frac{\partial}{\partial x_k}K\Big(\frac{y - X_{\mathbf{1}}}{b_{\mathbf{n}}}\Big)\right\}^2\\ &\leq (\hat{\mathbf{n}}^2b_{\mathbf{n}}^{d+2})^{-1}\sup_{\mathbb{R}^d} f\int_{\mathbb{R}^d}\left\{\frac{\partial}{\partial x_k}K(u)\right\}^2 du \leq \frac{C}{\hat{\mathbf{n}}^2b_{\mathbf{n}}^{d+2}}.\end{aligned} \tag{60}$$

We also have, for any $\mathbf{i} \neq \mathbf{j}$, by $\mathbf{H2}$ and $\mathbf{H'5}$

$$\begin{aligned}&\big|\text{cov}\big(\Delta'_{\mathbf{i}}(y), \Delta'_{\mathbf{j}}(y)\big)\big|\\ &\leq (\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-2}\int_{\mathbb{R}^d}\frac{\partial}{\partial x_k}K\Big(\frac{y-u}{b_{\mathbf{n}}}\Big)\frac{\partial}{\partial x_k}K\Big(\frac{y-v}{b_{\mathbf{n}}}\Big)\big|f_{\mathbf{i},\mathbf{j}}(u,v) - f(u)f(v)\big|dudv\\ &\leq C(\hat{\mathbf{n}}b_{\mathbf{n}}^{d+1})^{-2}\Big\{\int_{\mathbb{R}^d}\frac{\partial}{\partial x_k}K\Big(\frac{y-u}{b_{\mathbf{n}}}\Big)du\Big\}^2 \leq \frac{C}{\hat{\mathbf{n}}^2b_{\mathbf{n}}^2}.\end{aligned}$$

Then, for $\delta = 1/b_{\mathbf{n}}^{-d/N}$,

$$\begin{aligned}\sum_{\mathbf{i},\mathbf{j}:\ \|\mathbf{i}-\mathbf{j}\|\geq\delta} |\mathrm{cov}(\Delta'_{\mathbf{i}}(y),\Delta'_{\mathbf{j}}(y))| &\leq C(\hat{\mathbf{n}}b_{\mathbf{n}})^{-2} \sum_{\substack{i_k:\|\mathbf{i}\|\geq\delta\\ k=1,\dots,N}} \varphi(\|\mathbf{i}\|)\\ &\leq Cp^N(\hat{\mathbf{n}}b_{\mathbf{n}})^{-2}\sum_{i\geq\delta} i^{N-1}\varphi(i) \leq Cp^N(\hat{\mathbf{n}}b_{\mathbf{n}})^{-2}\int_{\delta}^{+\infty} v^{N-\theta-1}dv\\ &\leq Cp^N(\hat{\mathbf{n}}b_{\mathbf{n}})^{-2}\delta^{\theta-N} \leq Cp^N(\hat{\mathbf{n}}^2 b_{\mathbf{n}}^{d+2})^{-1}, \end{aligned} \tag{61}$$

and

$$\begin{aligned}\sum_{\mathbf{i},\mathbf{j}:\ 0<\|\mathbf{i}-\mathbf{j}\|<\delta} |\mathrm{cov}(\Delta'_{\mathbf{i}}(y),\Delta'_{\mathbf{j}}(y))| &\leq C(\hat{\mathbf{n}}b_{\mathbf{n}})^{-2} \sum_{\mathbf{i},\mathbf{j}:\ 0<\|\mathbf{i}-\mathbf{j}\|<\delta} 1\\ &\leq C(\hat{\mathbf{n}}b_{\mathbf{n}})^{-2}(2\delta p)^N \leq Cp^N(\hat{\mathbf{n}}^2 b_{\mathbf{n}}^{d+2})^{-1}. \end{aligned} \tag{62}$$

Combining (57)-(62), we obtain

$$\mathbb{P}\left(\left|\sum_{i=1}^{r} W_i^*\right| \geq \epsilon_{\mathbf{n}}/2)\right) \leq 2\exp\left\{(-\eta+C)\log\hat{\mathbf{n}}\right\}. \tag{63}$$

Finally, by (55)-(57) and (63), we get

$$\begin{aligned}&\mathbb{P}\big(\big|\max_{1\leq t\leq\nu'} T(\mathbf{n},1,y_t))\big| \geq \epsilon_{\mathbf{n}}\big)\\ &\quad\leq 2\nu'_{\mathbf{n}}\exp\{(-\eta/2+C)\log\hat{\mathbf{n}}\} + C\nu'_{\mathbf{n}}\hat{\mathbf{n}}^{\xi} b_{\mathbf{n}}^{-(d+1)}\varphi(p)\epsilon_{\mathbf{n}}^{-1}. \end{aligned} \tag{64}$$

If we choose $p = \psi_{\mathbf{n}}^{-1/N}$ and η large enough, then by (64) and (4,10), we get

$$\sum_{\mathbf{n}\in\mathbb{N}^{*N}} \mathbb{P}\left(\left|\max_{1\leq t\leq\nu'} T(\mathbf{n},1,y_t))\right| \geq \epsilon_{\mathbf{n}}\right) < \infty. \tag{65}$$

Finally, (53) is immediate by (65) together with Borel-Cantelli Lemma and the proof is completed. □

Proof of Theorem 6 The proof is inspired from [20] in which independent univariate case. Let $0 < M < \infty$ and $\epsilon > 0$ be a positive constant for which (15) in Lemma 2 is satisfied. Then,

$$\begin{aligned}\mathbb{P}\left(\|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| \geq M\psi_{\mathbf{n}}\right) &\leq \mathbb{P}\left(\|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| \geq \epsilon\right)\\ &\quad + \mathbb{P}\left(\|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| \geq M\psi_{\mathbf{n}}, \|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| < \epsilon\right). \end{aligned} \tag{66}$$

For the first on the right-hand side of (66), Theorem 4 yields

$$\sum_{\mathbf{n}\in\mathbb{N}^{*N}} \mathbb{P}\left(\|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| \geq \epsilon\right) < \infty. \tag{67}$$

Let us now turn to the other term. Set

$$D_{\mathbf{n}} = \left\{x\in D:\ M\psi_{\mathbf{n}} \leq \|x - \boldsymbol{m}\| < \epsilon\right\}.$$

By definition of mode estimator, we have

$$\mathbb{P}\left(\|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| \geq M\psi_{\mathbf{n}}, \|\hat{\boldsymbol{m}}_{\mathbf{n}} - \boldsymbol{m}\| < \epsilon\right) = \mathbb{P}(\hat{\boldsymbol{m}}_{\mathbf{n}} \in D_{\mathbf{n}})$$
$$\leq \mathbb{P}\left(\sup_{x \in D_{\mathbf{n}}} f_{\mathbf{n}}(x) \geq f_{\mathbf{n}}(\boldsymbol{m})\right)$$
$$\leq \mathbb{P}\left(\sup_{D_{\mathbf{n}}} \frac{f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x) - (f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}))}{\|x - \boldsymbol{m}\|} \geq \inf_{D_{\mathbf{n}}} \frac{\mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(x)}{\|x - \boldsymbol{m}\|}\right) := R_{\mathbf{n}}.$$

We will show that for a fixed k chosen (where $k = 1, ..., d$) with $|x_k - \boldsymbol{m}_k| > 0$,

$$R_{\mathbf{n}} \leq \mathbb{P}\left(\sup_{x \in D} \left|\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k}\right| \geq C\psi_{\mathbf{n}}\right), \tag{68}$$

for some generic constant $C > 0$. However, we have

$$\sup_{D_{\mathbf{n}}} \frac{f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x) - (f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}))}{\|x - \boldsymbol{m}\|} \leq \sup_{D_{\mathbf{n}}} \frac{f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x) - (f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}))}{|x_k - \boldsymbol{m}_k|},$$

since $\|x - \boldsymbol{m}\| = \sqrt{\sum_{j=1}^{d}(x_j - m_j)^2} \geq |x_k - m_k|$ with $\boldsymbol{m} = (m_1, ..., m_d)$. Theorems 1-2 and Corollary 2 yield

$$\sup_{D_{\mathbf{n}}} |f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x)| = \sup_{D_{\mathbf{n}}} |f_{\mathbf{n}}(x) - f(x)| + o(1).$$

Thus,

$$\sup_{D_{\mathbf{n}}} \frac{f_{\mathbf{n}}(x) - \mathbb{E}f_{\mathbf{n}}(x) - (f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}))}{\|x - \boldsymbol{m}\|}$$
$$\leq \sup_{D_{\mathbf{n}}} \frac{f_{\mathbf{n}}(x) - f(x) - (f_{\mathbf{n}}(\boldsymbol{m}) - f(\boldsymbol{m}))}{|x_k - \boldsymbol{m}_k|} + o(1)$$
$$\leq \sup_{x \in D} \left|\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k}\right| + o(1). \tag{69}$$

Let us now turn to the term $\mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(x)/\|x - \boldsymbol{m}\|$ which can be splitted into three parts. Using Taylors expansion of order 2 together with the symmetry of K, we get

$$\mathbb{E}f_{\mathbf{n}}(x) = \int_{\mathbb{R}^d} K(u) f(x - b_{\mathbf{n}}u)du = f(x) + \frac{b_{\mathbf{n}}^2}{2} \sum_{1 \leq i,j \leq d} \frac{\partial^2}{\partial x_i \partial x_j} f(x) + o(b_{\mathbf{n}}^2).$$

Thus,

$$\frac{\mathbb{E}f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E}f_{\mathbf{n}}(x)}{\|x - \boldsymbol{m}\|}$$
$$= \frac{f(\boldsymbol{m}) - f(x)}{\|x - \boldsymbol{m}\|} + O(b_{\mathbf{n}}^2) \frac{\sum_{1 \leq i,j \leq d} \left(\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{m}) - \frac{\partial^2}{\partial x_i \partial x_j} f(x)\right)}{\|x - \boldsymbol{m}\|} + o(b_{\mathbf{n}}^2) \frac{1}{\|x - \boldsymbol{m}\|}$$
$$:= T_1 + T_2 + T_3. \tag{70}$$

By Lemma 2, we have

$$T_1 = \frac{f(\boldsymbol{m}) - f(x)}{\|x - \boldsymbol{m}\|} \geq C\|x - \boldsymbol{m}\| \geq C\psi_{\mathbf{n}}. \tag{71}$$

Since $x \in D_{\mathbf{n}}$, then by $\mathbf{H'1}$, we can easily prove

$$|T_2| = O(b_{\mathbf{n}}^2) \frac{\sum_{1\leq i,j\leq d} \left| \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{m}) - \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right|}{\|x - \boldsymbol{m}\|} \leq \frac{C b_{\mathbf{n}}^2}{\|x - \boldsymbol{m}\|} = o(\psi_{\mathbf{n}}). \tag{72}$$

Clearly,

$$T_3 = o(b_{\mathbf{n}}^2) \frac{1}{\|x - \boldsymbol{m}\|} = o(\psi_{\mathbf{n}}). \tag{73}$$

Combining (70)-(73), we have for $\hat{\mathbf{n}}$ large enough

$$\inf_{D_{\mathbf{n}}} \frac{\mathbb{E} f_{\mathbf{n}}(\boldsymbol{m}) - \mathbb{E} f_{\mathbf{n}}(x)}{\|x - \boldsymbol{m}\|} \geq C\psi_{\mathbf{n}}. \tag{74}$$

Then, (69) and (74) yield (68). By Theorem 5, we get

$$\sum_{\mathbf{n}\in \mathbb{N}^{*N}} R_{\mathbf{n}} < \infty. \tag{75}$$

Finally, according to (66)-(68), (75) and Borel-Cantelli Lemma, the proof is completed. □

Proof of Lemma 1 By strict stationarity, we can write, for each $k = 1, ..., d$,

$$\begin{aligned}
\mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} &= \frac{1}{b_{\mathbf{n}}^{d+1}} \int_{\mathbb{R}^d} \frac{\partial}{\partial x_k} K\left(\frac{x-u}{b_{\mathbf{n}}}\right) f(u)du \\
&= \frac{1}{b_{\mathbf{n}}} \int_{\mathbb{R}^d} \frac{\partial}{\partial x_k} K(v) f(x - b_{\mathbf{n}} v)dv.
\end{aligned}$$

Using integration by parts together with $\mathbf{H'6}$, we get by Taylor expansion and the symmetry of K,

$$\begin{aligned}
&\mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} = \int_{\mathbb{R}^d} K(v) \frac{\partial}{\partial x_k} f(x - b_{\mathbf{n}} v)dv \\
&= \int_{\mathbb{R}^d} K(v) \left\{ \frac{\partial f(x)}{\partial x_k} + b_{\mathbf{n}} \sum_{1\leq l\leq d} v_l \frac{\partial^2 f(x)}{\partial x_l \partial x_k} + \frac{b_{\mathbf{n}}^2}{2} \sum_{1\leq l,s\leq d} v_s v_l \frac{\partial^3 f(x - \zeta b_{\mathbf{n}} v)}{\partial x_s \partial x_l \partial x_k} + o(b_{\mathbf{n}}^2) \right\} dv \\
&= \frac{\partial f(x)}{\partial x_k} + \frac{b_{\mathbf{n}}^2}{2} \sum_{1\leq l,s\leq d} \frac{\int_{\mathbb{R}^d} v_s v_l K(v)\, \partial^3 f(x - \zeta b_{\mathbf{n}} v)dv}{\partial x_s \partial x_l \partial x_k} + o(b_{\mathbf{n}}^2),
\end{aligned}$$

for some $0 < \zeta < 1$. Finally, since by $\mathbf{H'1}$, the density f has all partial derivatives bounded, then by $\mathbf{H'7}$, we have for each $k = 1, ..., d$,

$$\sup_{x\in D} \left| \mathbb{E}\frac{\partial f_{\mathbf{n}}(x)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right| = O\left(b_{\mathbf{n}}^2\right),$$

and the proof is concluded because $b_{\mathbf{n}}^2 = O(\psi_{\mathbf{n}})$ by assumption on $b_{\mathbf{n}}$.

Proof of Lemma 2 (see [8], Lemma 5) Since f is twice differentiable, then ∇f is uniformly continuous on $\boldsymbol{B}(\boldsymbol{m}, \epsilon_0)$ for some $\epsilon_0 > 0$. Furthermore, for any $v \in \mathbb{R}^d$ with $\|v\| = 1$, we have $\lim_{t\to 0} v^T \nabla f(x + tv)/t = v^T \nabla^2 f(x)v$ with

$$-\lambda_1 \leq v^T \nabla^2 f(x)v \leq -\lambda_2,$$

where v^T denotes the transpose of v, and $-\lambda_1,\ -\lambda_2 < 0$ are eigenvalues of the Hessian $\nabla^2 f(x)$. Recall that all eigenvalues of $\nabla^2 f(x)$ are negative according to **H′3**. Consequently, there exists $\epsilon > 0$ small enough such that for any $\|v\| = 1$ and $t < \epsilon$,

$$-2C_2 \leq v^T \nabla f(x + tv)/t \leq -2C_1, \tag{76}$$

for some $0 < C_1 \leq C_2$. Now, for each $x \in \boldsymbol{B}(\boldsymbol{m}, \epsilon)$, let $v = (x - \boldsymbol{m})/\|x - \boldsymbol{m}\|$. Hence, by Taylor expansion, we can write for ϵ small enough,

$$f(x) - f(\boldsymbol{m}) = \int_0^{\|x-\boldsymbol{m}\|} v^T \nabla f(x + tv)dt.$$

Finally, by (76), we get

$$-C_2\|x - \boldsymbol{m}\|^2 \leq f(x) - f(\boldsymbol{m}) \leq -C_1\|x - \boldsymbol{m}\|^2$$

and the proof is completed. □

Acknowledgments

The author would like to thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. Special thanks also go to the staff of UMR MISTEA for encouraging my research.

References

1. C. Abraham, G. Biau, and B. Cadre. Simple estimation of the mode of a multivariate density. *The Canadian Journal of Statistics*, 31:23–34, 2003.
2. C. Abraham, G. Biau, and B. Cadre. On the asymptotic properties of a simple estimate of the mode. *ESAIM: PS*, 8:1–11, 2004.
3. G. Biau. Optimal asymptotic quadratic errors of density estimators on random fields. *Statistics & Probability Letters*, 60:297–307, 2002.
4. G. Biau and B. Cadre. Nonparametric spatial prediction. *Statistical Inference for Stochastic Processes*, 7:327–349, 2004.
5. D. Bosq. *Nonparametric statistics for stochastic processes.* Springer, New York, 1998.
6. M. Carbon, C. Francq, and L. T. Tran. Kernel regression estimation for random fields. *Journal of Statistical Planning and Inference*, 137:778–798, 2007.
7. M. Carbon, L. Tran, and B. Wu. Kernel density estimation for random fields (density estimation for random fields). *Statistics & Probability Letters*, 36:115–125, 1997.
8. S. Dasgupta and S. Kpotufe. Optimal rates for k-nn density and mode estimation. *In Advances in Neural Information Processing Systems*, pages 2555–2563, 2014.
9. L. Devroye. Recursive estimation of the mode of a multivariate density. *The Canadian Journal of Statistics*, 7:159–167, 1979.

10. X. Guyon. Estimation dun champ par pseudo-vraisemblance conditionnelle: Etude asymptotique et application au cas markovien. *Proc.6th Franco-Belgian Meeting of Statisticians*, 1987.
11. E. Hwang and D.W. Shin. Kernel estimators of mode under ψ-weak dependence. *Annals of the Institute of Statistical Mathematics*, 68:301–327, 2016.
12. B.Garel M. Carbon and L.T. Tran. Frequency polygons for weakly dependent processes. *Statistics & Probability Letters*, 33:1–13, 1997.
13. A. Mokkadem and M. Pelletier. The law of the iterated logarithm for the multivariate kernel mode estimator. *ESAIM:P.S*, 7:1–21, 2003.
14. C. C. Nedearhouser. Convergence of blocks spins defined by a random fields. *Journal of Statistical Physics*, 22:673–684, 1980.
15. E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076, 1962.
16. D. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, New York, 1984.
17. B. L. S. Prakasa Rao. *Nonparametric Functional Estimation*. Z. W. Birnbaum E. Lukacs, 1983.
18. J. P. Romano. On weak convergence and optimality of kernel density estimates of the mode. *The Annals of Statistics*, 16(2):629–647, 1988.
19. M. Rosenblatt. *Stationary sequences and random fields*. Birkhäuser, Boston, 1985.
20. X. Shi, Y. Wu, and B. Miao. A note of the convergence rate of the kernel density estimator of the mode. *Statistics & Probability Letters*, 79:18661871, 2009.
21. B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.
22. L.T. Tran. Kernel density estimation on random fields. *Journal of Multivariate Analysis*, 34:37–53, 1990.
23. A. Younso. On the consistency of a new kernel rule for spatially dependent data. *Statistics & Probability Letters*, 131:64–71, 2017.