T E C H N I C A L

R E P O R T

0207

KERNEL DENSITY ESTIMATION FOR SPATIAL PROCESSES : THE L_1 THEORY

M. HALLIN, Z. LU and L.T. TRAN

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Kernel Density Estimation for Spatial Processes : the L_1 Theory

Marc Hallin*
I.S.R.O., E.C.A.R.E.S., and Département de Mathématique
Université Libre de Bruxelles
Brussels, Belgium

Zudi Lu†
Academy of Mathematics and System Sciences
Chinese Academy of Sciences
Beijing, China

and

Lanh T. Tran
Department of Mathematics
Indiana University
Bloomington, USA

January 29, 2002

Abstract

The purpose of this paper is to investigate kernel density estimators for spatial processes with linear or nonlinear structures. Sufficient conditions for kernel estimators to converge in L_1 are obtained under extremely general, verifiable conditions. The results hold for mixing as well as for non-mixing processes. Potential applications include testing for spatial interaction, the spatial analysis of causality structures, the definition of leading/lagging sites, the construction of clusters of comoving sites, etc.

AMS 2000 subject classifications. Primary 62G07, 62G20; Secondary 62M10.
Key words and phrases. Bandwidth, kernel density estimator, L_1 theory, spatial linear or nonlinear processes.

1 Introduction

The applications of spatial statistical models are extremely numerous and diverse. Data, in a number of fields, are collected on the surface of the earth, thus involving two or three dimensional spatial coordinates, possibly more. The subject has generated an enormous literature that

*Research supported by an A.R.C. contract of the Communauté française de Belgique.
†Research supported by the National Science Foundation of China.

cannot be reviewed here; for background material on spatial statistics and spatial time series, the reader is referred to Anselin and Florex (1995), Basawa (1996a,b), Cressie (1991), Possolo (1991), Ripley (1981), Rosenblatt (1985), and Tjøstheim (1987).

In this paper, our goal is to study nonparametric density estimation for spatial data with linear or nonlinear structures in situations in which parametric estimation cannot be adopted with confidence.

Denote by $\mathbb{Z}^N$ the set of integer lattice points in the N-dimensional Euclidean space, where $N \geq 1$ and $\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$. A d-dimensional random field over $\mathbb{Z}^N$ is a $\mathbb{R}^d$-valued stochastic process $\{\mathbf{X_j}, \mathbf{j} \in \mathbb{Z}^N\}$ defined on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$. In the sequel, we always tacitly assume that $\{\mathbf{X_j}\}$ is strictly stationary; a point $\mathbf{j} \in \mathbb{Z}^N$ will be referred to as a *site*.

A random field $\{\mathbf{X_j}\}$ is called *linear* if there exist an integer $\tilde{d}$, a collection of $d \times \tilde{d}$ matrices $\mathbf{a_i}$, $\mathbf{i} \in \mathbb{Z}^N$, and an i.i.d. $\mathbb{R}^{\tilde{d}}$-valued random field $\{\mathbf{e_n}\}$ defined on $(\Omega, \mathcal{F}, \mathrm{P})$, with $\mathrm{E}[\mathbf{e_j}] = \mathbf{0}$ and $\mathrm{E}[\mathbf{e_j}\mathbf{e_j}'] := \sigma^2\mathbf{I}$, $\sigma < \infty$ such that

$$\mathbf{X_j} = \sum_{\mathbf{i} \in \mathbb{Z}^N} \mathbf{a_i}\mathbf{e_{j-i}}, \quad \mathbf{j} \in \mathbb{Z}^N, \tag{1.1}$$

where the series on the right-hand side is convergent in L_2.

Such models were considered as early as 1954 by Whittle (1954, 1963), who suggested a linear spatial autoregression model (see also Kulkarni 1992) whose stationary solution can be expressed (with $N = 2$ and $d = \tilde{d} = 1$) as a spatial moving average of the form (1.1). The problem of density estimation for linear random fields was studied in Hallin, Lu, and Tran (2001) from a L_2 point of view.

In this paper, we consider the much more general case of a nonlinear structure of the form

$$\mathbf{X_j} = g(\mathbf{e_{j-i}}, \mathbf{i} \in \mathbb{Z}^N), \quad \mathbf{j} \in \mathbb{Z}^N, \tag{1.2}$$

where g is a Borel-measurable function from $\left(\mathbb{R}^{\tilde{d}}\right)^{\mathbb{Z}^N}$ to $\mathbb{R}^d$. Let $\mathcal{I}_\mathbf{n}$, where $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{Z}^N$ is such that $n_k \geq 1$, $k = 1, \ldots, N$, be the rectangular region of $\mathbb{Z}^N$ defined by

$$\mathcal{I}_\mathbf{n} := \{\mathbf{i} \in \mathbb{Z}^N : 1 \leq i_k \leq n_k,\ k = 1, \ldots, N\};$$

the number of sites in $\mathcal{I}_\mathbf{n}$ is denoted as $\hat{\mathbf{n}} := \prod_{k=1}^N n_k$. We write that $\mathbf{n} \to \infty$ when

$$\min_{1 \leq k \leq N} \{n_k\} \to \infty \qquad \text{and} \qquad \max_{1 \leq j,\, k \leq N} |n_j / n_k| < C$$

for some fixed $0 < C < \infty$; the same letter C is used throughout for various positive constants, the value of which is unimportant.

Suppose that $\{\mathbf{X_j}\}$ has density f, and is observed on $\mathcal{I}_\mathbf{n}$. The kernel density estimator $f_\mathbf{n}$ of f is defined by

$$f_\mathbf{n}(\mathbf{x}) := (\hat{\mathbf{n}} b^d)^{-1} \sum_{\mathbf{j} \in \mathcal{I}_\mathbf{n}} K\Big((\mathbf{x} - \mathbf{X_j})/b\Big), \quad \mathbf{x} \in \mathbb{R}^d, \tag{1.3}$$

where $K\ : \mathbb{R}^d \to \mathbb{R}^+$ is a kernel function, and $b \equiv b_\mathbf{n}$ a sequence of bandwidths tending to zero as $\mathbf{n}$ tends to infinity.

The convergence to f of $f_\mathbf{n}$ can be considered from several point of views, but Devroye and Györfi (1985, page 1) convincingly pointed out that L_1 is the natural distance to be considered for densities. Accordingly, the main objective of this paper is to obtain weak conditions for the L_1 distance

$$J_{\mathbf{n}} := \int_{\mathbb{R}^d} |f_{\mathbf{n}}(\mathbf{x}) - f(\mathbf{x})| \, d\mathbf{x} \tag{1.4}$$

between $f_{\mathbf{n}}$ and f to converge to zero in probability.

The nonparametric estimation of a probability density f has a pretty long history in statistical inference. In the independent case with $N = 1$, the L_1 distance (1.4) tends to zero under very mild conditions (see Devroye 1983); this consistency result was employed by Chan and Tran (1992) in the problem of testing for serial dependence. The results obtained here similarly have potential usefulness in testing for spatial interaction.

The same problem has been studied quite extensively in the context of strongly mixing stationary processes ($N = 1$) (see Roussas 1988, Robinson 1983 and 1987, Ioannides and Roussas 1987, Masry and Györfi 1987, Yakowitz 1987, Boente and Fraiman 1988, Bosq 1989, Tran 1989, and Györfi, Härdle, Sarda, and Vieu 1990, to name only a few). For linear process (satisfying (1.1) with $N = 1$ and $d = \tilde{d} = 1$), results have been obtained by Chanda (1982), Tran (1992), and Hallin and Tran (1996) among others.

The case of random fields ($N > 1$) has not been studied much : the fact that the sites do not have a natural ordering indeed makes the problem technically more difficult. Tran (1990) and Tran and Yakowitz (1993) have investigated some aspects of the problem. More recently, Hallin, Lu and Tran (2001) have obtained the limiting distribution of kernel density estimators for linear random fields under general conditions.

To the best of our knowledge, the only paper dealing with a L_1 approach of kernel estimation problems in random fields is Carbon, Hallin and Tran (1996), under strong mixing assumptions. Although the strong mixing property is often reasonable, it is not satisfied by many processes of practical interest. A very simple example of a linear process that is not strong mixing is the autoregressive process $X_t = (1/2)X_{t-1} + e_t$ with $\mathrm{P}(e_t = \pm 1) = 1/2$ (here $d = \tilde{d} = 1$); cf. Andrews (1984). Moreover, it is generally impossible to check whether or not a process is strongly mixing.

In this paper, we relax both the linearity assumption of Hallin, Lu and Tran (2001), and the mixing assumptions of Carbon, Hallin and Tran (1996). Our results show that kernel density estimators of f converge in L_1 under general and rather simple conditions on the bandwidth b.

The paper is organized as follows. The main assumptions and consistency result (Theorem 2.1) are presented in Section 2. In Section 3, this result is considered (Corollaries 3.1 and 3.2) in the the particular case of linear random fields. The conditions for $J_{\mathbf{n}}$ to converge to zero are quite simple in this case. Section 4 contains a series of lemmas which are crucial for the proof of (Theorem 2.1). In Section 5, we devise a nonstandard blocking technique for spatial random variables, from which we establish an exponential inequality by Poissonization, as done by Devroye (1983). This exponential inequality in turn provides sharp bounds which are later utilized to obtain mild conditions for the convergence of $J_{\mathbf{n}}$. The proof of the main result then readily follows (Section 6).

2 Assumptions and main result

Assume throughout that $\{\mathbf{X_j}\}$, observed over $\mathcal{I}_{\mathbf{n}}$, satisfies (1.2). For any site $\mathbf{j} = (j_1, \ldots, j_N)$ and any $\mathbf{m} = (m_1, \ldots, m_N) \in \mathbb{Z}^N$ such that $m_k \geq 1$, k=1,..., N, set

$$\mathcal{R}(\mathbf{j}, \mathbf{m}) := \{\mathbf{i} \in \mathbb{Z}^N : |i_1 - j_1| \leq m_1, \ldots, |i_N - j_N| \leq m_N\},$$

and

$$\mathbf{X}_{\mathbf{j}}^{(\mathbf{m})} := g_{\mathbf{m}}(\mathbf{e_i}, \mathbf{i} \in \mathcal{R}(\mathbf{j}, \mathbf{m})) := \mathrm{E}\Big[\mathbf{X_j} | \mathbf{e_i}, \mathbf{i} \in \mathcal{R}(\mathbf{j}, \mathbf{m})\Big]. \tag{2.1}$$

Write $v(\mathbf{m}) = \mathrm{E}\|\mathbf{X_j} - \mathbf{X}_{\mathbf{j}}^{(\mathbf{m})}\|^2$, where $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^d$: the random field $\{\mathbf{X_j}\}$ is said to be *v-stable* (in L_2) with respect to $\{\mathbf{e_j}\}$ if $\lim_{\mathbf{m}\to\infty} v(\mathbf{m}) = 0$ as $\mathbf{m} \to \infty$. We call $v(\mathbf{m})$ the *stability coefficients* of $\{\mathbf{X_j}\}$ (with respect to $\{\mathbf{e_j}\}$).

It is clear that, when g is linear, that is, when $\{\mathbf{X_j}\}$ is a linear random field of the form (1.1), then

$$v(\mathbf{m}) = \sigma^2 \sum_{\mathbf{i}} \|\mathbf{a_i}\|^2, \quad \text{with} \quad \|\mathbf{a_i}\|^2 = \sum_{k=1}^{d} \sum_{\ell=1}^{\tilde{d}} (\mathbf{a_i})_{k\ell}^2, \tag{2.2}$$

where the summation $\sum_{\mathbf{i}}$ covers all sites $\mathbf{i} = (i_1, \ldots, i_N)$ such that $i_k > m_k$ for some $k = 1, \ldots, N$.

Our consistency result (Theorem 2.1) of course requires some assumptions. Assumptions 2.1 and 2.2 are very mild and standard conditions on the kernel K and the boundedness of f, respectively; the same assumptions are made in Carbon, Hallin, and Tran (1996). On top of these two conditions, we also need an assumption relating the asymptotic behavior of the bandwidth and the stability coefficients of the model. If the stability coefficients go to zero "slowly" (Assumption2.3), then the bandwidth should go to zero "fast" (Assumption 2.4). If If the stability coefficients go to zero "fast" (Assumption 2.5), then bandwidth decay can be "slower" (Assumption 2.6).

Assumption 2.1 The kernel function K in (1.3) is absolutely integrable, with $\int K(\mathbf{z})d\mathbf{z} = 1$ and $\int \|\mathbf{z}\| K(\mathbf{z})d\mathbf{z} < \infty$. In addition, $|K(\mathbf{x}) - K(\mathbf{y})| \le C\|\mathbf{x} - \mathbf{y}\|$ for some constant $C > 0$ and any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

Assumption 2.2 The density f of $X_{\mathbf{i}}$ is bounded.

Assumption 2.3 There exists a constant $a > 0$ such that $v(k\mathbf{1}) = o(k^{-a})$ as $k \to \infty$, where $\mathbf{1} := (1, \ldots, 1) \in \mathbb{Z}^N$.

Assumption 2.4 The bandwidth $b = b_{\mathbf{n}}$ tends to 0 slowly enough that

$$\lim_{\mathbf{n}\to\infty} \hat{\mathbf{n}} b_{\mathbf{n}}^{d+2(d+1)N/a} = \infty. \tag{2.3}$$

Note that when a is large, condition (2.3) is close to the condition that $\hat{\mathbf{n}} b_{\mathbf{n}}^d \to \infty$, which is the condition needed for $J_{\mathbf{n}}$ to converge to zero in probability in the i.i.d. case (see, for example, Devroye 1983).

Assumption 2.3 only requires that $v(k\mathbf{1})$ decays at algebraic (polynomial) rate as $k \to \infty$. If this decay is geometric, then the condition on the bandwidth can be much weaker, and almost the same as in the i.i.d. case. More precisely, let us replace Assumption 2.3 with

Assumption 2.5 There exists $0 < \rho < 1$ such that $v(k\mathbf{1}) = O(\rho^k)$ as $k \to \infty$.

Then, Assumption 2.4 can be weakened into

Assumption 2.6 The bandwidth $b_{\mathbf{n}}$ is such that

$$b_{\mathbf{n}}^d (\ln \hat{\mathbf{n}}) \to 0 \quad \text{and} \quad \hat{\mathbf{n}} b_{\mathbf{n}}^d (\ln \hat{\mathbf{n}})^{-N} \to \infty. \tag{2.4}$$

as $\mathbf{n} \to \infty$.

Under these assumptions, we now can state the main result of this paper, which establishes the L_1 consistency of $\hat{f}_{\mathbf{n}}$.

Theorem 2.1 *Suppose that either Assumptions* 2.1, 2.2, 2.3, *and* 2.4 *or Assumptions* 2.1, 2.2, 2.5, *and* 2.6 *are satisfied : then* $J_{\mathbf{n}} \stackrel{\mathrm{P}}{\to} 0$ *as* $\mathbf{n} \to \infty$.

3 Linear random fields

In this section, we show that the conditions on the stability coefficients in Assumptions 2.3 and 2.5 are satisfied by a large class of linear random fields (of the form (1.1)). For simplicity, we let $d = \tilde{d} = 1$ (writing $X_{\mathbf{j}}$ instead of $\mathbf{X}_{\mathbf{j}}$, $a_{\mathbf{i}}$ instead of $a_{\mathbf{i}}$, etc.). Suppose $a_{\mathbf{i}}$ tends to zero at algebraic rate, with

$$|a_{\mathbf{i}}| = O(|i_1|^{-\delta_1} \dots |i_N|^{-\delta_N}) \tag{3.1}$$

as $\mathbf{i} \to \infty$, where $\delta_k > 1/2$ for $k = 1, \cdots, N$. From (2.2),

$$\begin{aligned} v(\mathbf{m}) &\leq O(1)(\sum_{i_1=m_1}^{\infty} i_1^{-2\delta_1} + \dots + \sum_{i_N=m_N}^{\infty} i_N^{-2\delta_N}) \\ &\leq O(i_1^{-2\delta_1+1} + \dots + i_N^{-2\delta_N+1}). \end{aligned}$$

Choose $0 < a < 2\min_{1\leq k\leq N} \delta_k - 1$: then,

$$k^a v(k\mathbf{1}) = k^a O(k^{-2\min_{1\leq k\leq N} \delta_k + 1}) \to 0,$$

so that Assumption 2.3 is satisfied. We have thus proved the following corollary to Theorem 2.1.

Corollary 3.1 *Suppose the linear random field* $\{X_{\mathbf{j}}\}$ *defined in* (1.1) *(with* $d = \tilde{d} = 1$*) satisfies condition* (3.1). *Suppose that Assumptions* 2.1, 2.2, *and* 2.4 *also hold. Then* $J_{\mathbf{n}} \stackrel{\mathrm{P}}{\to} 0$ *as* $\mathbf{n} \to \infty$.

Next, if $a_{\mathbf{i}}$ decays at a geometric rate, that is, if

$$|a_{\mathbf{i}}| = O(\rho_1^{|i_1|} \dots \rho_N^{|i_N|}) \tag{3.2}$$

as $\mathbf{i} \to \infty$, where $0 < \rho_k < 1$ for $k = 1, \cdots, N$, then

$$v(\mathbf{m}) \leq O(\rho_1^{m_1} + \dots + \rho_N^{m_N}),$$

and thus $v(k\mathbf{1}) = O(\rho^k) \to 0$, with $0 < \rho = \max_{1\leq k\leq N} \rho_i < 1$. Hence Assumption 2.5 is satisfied, and the following corollary to Theorem 2.1 holds.

Corollary 3.2 *Suppose the linear random field* $\{X_{\mathbf{j}}\}$ *defined in* (1.1) *(with* $d = \tilde{d} = 1$*) satisfies condition* (3.2). *Suppose that Assumption* 2.1, 2.2, *and* 2.6 *also hold. Then* $J_{\mathbf{n}} \stackrel{\mathrm{P}}{\to} 0$ *as* $\mathbf{n} \to \infty$.

4 Preliminaries and lemmas

The proof of Proposition 2.1 relies on a series of lemmas, which we now state and prove. For any positive constant u, define

$$J_{\mathbf{n}1}(u) := \int_{\|\mathbf{x}\|>u} |f_{\mathbf{n}}(\mathbf{x}) - f(\mathbf{x})|d\mathbf{x}, \quad \text{and} \quad J_{\mathbf{n}2}(u) := \int_{\|\mathbf{x}\|\leq u} |f_{\mathbf{n}}(\mathbf{x}) - f(\mathbf{x})|d\mathbf{x}.$$

Note that, for any $u \geq 0$, $J_{\mathbf{n}}$ decomposes into $J_{\mathbf{n}} = J_{\mathbf{n}1}(u) + J_{\mathbf{n}2}(u)$.

Lemma 4.1 *Let $\varepsilon > 0$ be an arbitrarily small positive number. Then, $\lim_{\mathbf{n}\to\infty} \mathrm{E}\left[J_{\mathbf{n}1}(u)\right] < \varepsilon$ for all u larger than some $U(\varepsilon)$.*

Proof. Since $\int f(\mathbf{x})d\mathbf{x} = 1$ and $\int K(\mathbf{x})d\mathbf{x} < \infty$, for $u > 0$ sufficiently large, we have

$$\int_{\|\mathbf{x}\|>u/2} f(\mathbf{x})d\mathbf{x} < \varepsilon/3 \quad \text{and} \quad \int_{\|\mathbf{x}\|>u} K(\mathbf{x})d\mathbf{x} < \varepsilon/3, \tag{4.1}$$

hence

$$\begin{aligned} \mathrm{E}[J_{\mathbf{n}1}(u)] &\leq \mathrm{E}\left[\int_{\|\mathbf{x}\|>u} f_{\mathbf{n}}(\mathbf{x})d\mathbf{x}\right] + \int_{\|\mathbf{x}\|>u} f(\mathbf{x})d\mathbf{x} \\ &\leq \int \int_{\|\mathbf{x}\|>u} b^{-d}K((\mathbf{x}-\mathbf{y})/b)d\mathbf{x}f(\mathbf{y})d\mathbf{y} + \varepsilon/3. \end{aligned} \tag{4.2}$$

Letting $\mathbf{t} := (\mathbf{x}-\mathbf{y})/b$ and noting that $b \to 0$ as $\mathbf{n} \to \infty$, (4.1) and (4.2) yield

$$\begin{aligned} \mathrm{E}[J_{\mathbf{n}1}(u)] &\leq \int \int_{\|\mathbf{y}+b\mathbf{t}\|>u} K(\mathbf{t})d\mathbf{x}f(\mathbf{y})d\mathbf{y} + \varepsilon/3 \\ &\leq \int_{\|\mathbf{t}\|>u} K(\mathbf{t})d\mathbf{t} + \int \int_{\|\mathbf{y}\|+b\|\mathbf{t}\|>u,\, \|\mathbf{t}\|\leq u} K(\mathbf{t})d\mathbf{x}f(\mathbf{y})d\mathbf{y} + \varepsilon/3 \\ &\leq \int \int_{\|\mathbf{t}\|\leq u} K(\mathbf{t})d\mathbf{x}I_{(\|\mathbf{y}\|+bu>u)}(\mathbf{y})f(\mathbf{y})d\mathbf{y} + 2\varepsilon/3 \\ &\leq \int_{\|\mathbf{y}\|>u/2} f(\mathbf{y})d\mathbf{y} + 2\varepsilon/3 < \varepsilon, \end{aligned}$$

which completes the proof. □

Still for $u > 0$, define

$$J^{(1)}_{\mathbf{n}2}(u) := \int_{\|\mathbf{x}\|\leq u} |f_{\mathbf{n}}(\mathbf{x}) - \mathrm{E}f_{\mathbf{n}}(\mathbf{x})|d\mathbf{x} \quad \text{and} \quad J^{(2)}_{\mathbf{n}2} := \int_{\mathbb{R}^d} |\mathrm{E}f_{\mathbf{n}}(\mathbf{x}) - f(\mathbf{x})|d\mathbf{x}.$$

Obviously, $J_{\mathbf{n}2}(u) \leq J^{(1)}_{\mathbf{n}2}(u) + J^{(2)}_{\mathbf{n}2}$.

Lemma 4.2 *For all $u > 0$, $J^{(2)}_{\mathbf{n}2}(u) \to 0$ as $\mathbf{n} \to \infty$.*

Proof. See Lemma 2.1 of Carbon, Hallin, and Tran (1996, p.159), or Lemma 1 of Devroye (1983, page 897). □

Clearly, we have, for all $\mathbf{m}$,

$$\begin{aligned} J^{(1)}_{\mathbf{n}2}(u) &\leq \int_{\|\mathbf{x}\|\leq u} |f_{\mathbf{n}}(\mathbf{x}) - f^{(\mathbf{m})}_{\mathbf{n}}(\mathbf{x})|d\mathbf{x} + \int_{\|\mathbf{x}\|\leq u} |f^{(\mathbf{m})}_{\mathbf{n}}(\mathbf{x}) - \mathrm{E}f^{(\mathbf{m})}_{\mathbf{n}}(\mathbf{x})|d\mathbf{x} \\ &\qquad + \int_{\|\mathbf{x}\|\leq u} |\mathrm{E}f^{(\mathbf{m})}_{\mathbf{n}}(\mathbf{x}) - \mathrm{E}f_{\mathbf{n}}(\mathbf{x})|d\mathbf{x} \\ &:= I_{\mathbf{n}1}(u) + I_{\mathbf{n}2}(u) + I_{\mathbf{n}3}(u), \end{aligned}$$

say, where $f^{(\mathbf{m})}_{\mathbf{n}}(\mathbf{x}) := (\hat{\mathbf{n}}b^d)^{-1}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} K((\mathbf{x}-\mathbf{X}^{(\mathbf{m})}_{\mathbf{j}})/b)$, with $\mathbf{X}^{(\mathbf{m})}_{\mathbf{j}}$ defined in (2.1). Consider a sequence $\mathbf{m} = \mathbf{m}(\mathbf{n})$, $\mathbf{n} \in \mathbb{Z}^N$, such that, for $\mathbf{n} \to \infty$, $\mathbf{m}(\mathbf{n}) \to \infty$. In the sequel, we will make some statements requiring that

$$b^{-2(d+1)}v(\mathbf{m}(\mathbf{n})) \to 0, \tag{4.3}$$

and/or

$$\hat{\mathbf{m}}/(\hat{\mathbf{n}}b^d) \to 0, \quad \text{and} \quad \hat{\mathbf{m}}\exp\{-C\hat{\mathbf{n}}\varepsilon^2/\hat{\mathbf{m}}\} \to 0. \tag{4.4}$$

for all $\varepsilon > 0$. Such sequences $\mathbf{m}(\mathbf{n})$ need not exist; later on (Section 6), however, we will show that their existence follows from either Assumption 2.3 or Assumption 2.5 (for (4.3)), either Assumption 2.4 or Assumption 2.6 (for (4.4)).

Lemma 4.3 *Let the sequence* $\mathbf{m} = \mathbf{m}(\mathbf{n})$ *satisfy* (4.3). *Then, for all* u, $I_{\mathbf{n}1}(u) \xrightarrow{\mathrm{P}} 0$ *as* $\mathbf{n} \to \infty$.

Proof. Let λ denote the d-dimensional Lebesgue measure. By the Lipschitz continuity of $K(\cdot)$ and the definition of $v(\mathbf{m})$,

$$\begin{aligned} \mathrm{E}\left[I_{\mathbf{n}1}(u)\right] &\leq \mathrm{E}\left[\int_{\|\mathbf{x}\|\leq u} (\hat{\mathbf{n}}b^d)^{-1}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\left|K((\mathbf{x}-\mathbf{X}_{\mathbf{j}})/b) - K((\mathbf{x}-\mathbf{X}^{(\mathbf{m})}_{\mathbf{j}})/b)\right| d\mathbf{x}\right] \\ &\leq C\mathrm{E}\left[(\hat{\mathbf{n}}b^d)^{-1}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\|\mathbf{X}_{\mathbf{j}} - \mathbf{X}^{(\mathbf{m})}_{\mathbf{j}}\|\right]/b\lambda(\|\mathbf{x}\|\leq u) \\ &= Cb^{-(d+1)}\mathrm{E}\left[\|\mathbf{X}_{\mathbf{j}} - \mathbf{X}^{(\mathbf{m})}_{\mathbf{j}}\|\right]\lambda(\|\mathbf{x}\|\leq u) \\ &\leq Cb^{-(d+1)}\left(\mathrm{E}\|\mathbf{X}_{\mathbf{j}} - \mathbf{X}^{(\mathbf{m})}_{\mathbf{j}}\|^2\right)^{1/2}\lambda(\|\mathbf{x}\|\leq u) \\ &\leq (b^{-2(d+1)}v(\mathbf{m}(\mathbf{n})))^{1/2}\lambda(\|\mathbf{x}\|\leq u), \end{aligned}$$

a quantity which, in view of (4.3), converges to zero as $\mathbf{n} \to \infty$. □

Lemma 4.4 *Under the same assumptions as in Lemma* 4.3, $I_{\mathbf{n}3}(u) \xrightarrow{\mathrm{P}} 0$ *for all* u, *as* $\mathbf{n} \to \infty$.

Proof. Thee proof follows along the same lines as for Lemma 4.3. □

The same property also holds for $I_{\mathbf{n}2}(u)$ provided that $\mathbf{m} = \mathbf{m}(\mathbf{n})$ satisfies (4.4), but the proof is more complicated.

Lemma 4.5 *Let the sequence* $\mathbf{m} = \mathbf{m}(\mathbf{n})$ *satisfy* (4.4). *Then,* $I_{\mathbf{n}2}(u) \xrightarrow{\mathrm{P}} 0$ *for all* u, *as* $\mathbf{n} \to \infty$.

Proof. For any given $\varepsilon > 0$, let the positive constants M, L, N_0, $a_1, \ldots, a_{N_0}$, and the N_0-tuple of disjoint finite rectangles $A_1, \ldots, A_{N_0}$ in $\mathbb{R}^d$ be such that the function $K^\star(\mathbf{x}) := \sum_{i=1}^{N_0} a_i I_{A_i}(\mathbf{x})$ satisfies

$$|K^\star| \leq M, \quad K^\star(\mathbf{x}) = 0 \text{ for } \mathbf{x} \notin [-L, L]^d, \quad \text{and} \quad \int |K(\mathbf{x}) - K^\star(\mathbf{x})|d\mathbf{x} < \varepsilon.$$

For all rectangle A in $\mathbb{R}^d$, set $\mu_{\mathbf{n}}^{(\mathbf{m})}(A) := \hat{\mathbf{n}}^{-1}\sum_{\mathbf{i}\in I_{\mathbf{n}}} I_{(\mathbf{X}_{\mathbf{i}}^{(\mathbf{m})}\in A)}$ and $\mu^{(\mathbf{m})}(A) := \mathrm{P}(\mathbf{X}_{\mathbf{i}}^{(\mathbf{m})} \in A)$. Then,

$$f_{\mathbf{n}}^{(\mathbf{m})}(\mathbf{x}) = b^{-d}\int K((\mathbf{x}-\mathbf{y})/b)\mu_{\mathbf{n}}^{(\mathbf{m})}(d\mathbf{y})$$

and

$$\mathrm{E}\left[f_{\mathbf{n}}^{(\mathbf{m})}(\mathbf{x})\right] = b^{-d}\int K((\mathbf{x}-\mathbf{y})/b)\mu^{(\mathbf{m})}(d\mathbf{y}).$$

Similarly, defining

$$f_{\mathbf{n}}^{\star(\mathbf{m})}(\mathbf{x}) := b^{-d}\int K^{\star}((\mathbf{x}-\mathbf{y})/b)\mu_{\mathbf{n}}^{(\mathbf{m})}(d\mathbf{y}),$$

we have

$$\mathrm{E}f_{\mathbf{n}}^{\star(\mathbf{m})}(\mathbf{x}) = b^{-d}\int K^{\star}((\mathbf{x}-\mathbf{y})/b)\mu^{(\mathbf{m})}(d\mathbf{y}).$$

Thus, $I_{\mathbf{n}2}(u)$ decomposes into

$$\begin{aligned} I_{\mathbf{n}2}(u) &= \int_{\|\mathbf{x}\|\le u} |f_{\mathbf{n}}^{(\mathbf{m})}(\mathbf{x}) - f_{\mathbf{n}}^{\star(\mathbf{m})}(\mathbf{x})|d\mathbf{x} + \int_{\|\mathbf{x}\|\le u} |f_{\mathbf{n}}^{\star(\mathbf{m})}(\mathbf{x}) - \mathrm{E}f_{\mathbf{n}}^{\star(\mathbf{m})}(\mathbf{x})|d\mathbf{x} \\ &\qquad + \int_{\|\mathbf{x}\|\le u} |\mathrm{E}f_{\mathbf{n}}^{\star(\mathbf{m})}(\mathbf{x}) - \mathrm{E}f_{\mathbf{n}}^{(\mathbf{m})}(\mathbf{x})|d\mathbf{x} \\ &:= I_{\mathbf{n}2}^{(1)}(u) + I_{\mathbf{n}2}^{(2)}(u) + I_{\mathbf{n}2}^{(3)}(u), \end{aligned}$$

say. By the same argument as in the proof of Lemma 3.2 of Devroye and Gyorfi (1985), it is easily proved that $I_{\mathbf{n}2}^{(1)}(u) < \varepsilon$ and $I_{\mathbf{n}2}^{(3)}(u) < \varepsilon$. For $I_{\mathbf{n}2}^{(2)}(u)$, we have

$$I_{\mathbf{n}2}^{(2)}(u) \le Mb^{-d}\sum_{i=1}^{N_0}\int_{\|\mathbf{x}\|\le u} |\mu_{\mathbf{n}}^{(\mathbf{m})}(\mathbf{x}+bA_i) - \mu^{(\mathbf{m})}(\mathbf{x}+bA_i)|d\mathbf{x}.$$

So, in order to prove that $I_{\mathbf{n}2}^{(2)}(u) \xrightarrow{\mathrm{P}} 0$, it suffices to show that, for all finite rectangle A of $\mathbb{R}^d$,

$$I_{\mathbf{n}2}^{A}(u) := b^{-d}\int_{\|\mathbf{x}\|\le u} |\mu_{\mathbf{n}}^{(\mathbf{m})}(\mathbf{x}+bA) - \mu^{(\mathbf{m})}(\mathbf{x}+bA)|d\mathbf{x} \xrightarrow{\mathrm{P}} 0.$$

Consider a partition of $\mathbb{R}^d$ into sets B that are d-fold products of intervals of the form $[(i-1)b/N_0,\, ib/N_0)$, where i is an integer, and N_0 a positive number to be chosen later. Call this partition Ψ. Let $\tilde{a}_1, ..., \tilde{a}_d$ be positive numbers with $\tilde{a}_i \ge 2/N_0$ for all $1 \le i \le d$. Define

$$A := \prod_{i=1}^{d}[\tilde{x}_i, \tilde{x}_i + \tilde{a}_i) \quad \text{and} \quad A^* := \prod_{i=1}^{d}\Big[\tilde{x}_i + (1/N_0),\, \tilde{x}_i + \tilde{a}_i - (1/N_0)\Big),$$

where $\tilde{x}_i$, $i = 1, \cdots, d$, are real constants, $C_{\mathbf{x}} := \mathbf{x} + bA - \bigcup\{B \in \Psi : B \subseteq \mathbf{x} + bA\}$, and $C_{\mathbf{x}}^* := \mathbf{x} + b(A - A^*)$. It is easy to see that $C_{\mathbf{x}} \subseteq C_{\mathbf{x}}^*$. Thus,

$$\begin{aligned} I_{\mathbf{n}2}^{A}(u) &\le b^{-d}\int \sum_{\substack{B\in\Psi \\ B\subseteq \mathbf{x}+bA}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)|d\mathbf{x} + b^{-d}\int\{\mu^{(\mathbf{m})}(C_{\mathbf{x}}^*) + \mu_{\mathbf{n}}^{(\mathbf{m})}(C_{\mathbf{x}}^*)\}d\mathbf{x} \\ &:= q_{1\mathbf{n}} + q_{2\mathbf{n}}, \end{aligned} \tag{4.5}$$

say. Let us show that $\mathrm{E}q_{1\mathbf{n}}$ and $\mathrm{E}q_{2\mathbf{n}}$ can be made arbitrarily small for sufficiently large N_0 and $\mathbf{n}$.
(a) Starting with $q_{2\mathbf{n}}$, note that

$$C^*_{\mathbf{x}} = \mathbf{x} + b\prod_{i=1}^{d}\Big\{[\tilde{x}_i,\, \tilde{x}_i + 1/N_0) \cup [\tilde{x}_i + \tilde{a}_i - 1/N_0,\, \tilde{x}_i + \tilde{a}_i)\Big\} := \mathbf{x} + b\tilde{C}.$$

From (4.5) and (4.6), and the fact that, for any Borel set $\tilde{C}$, and any probability measure ν on the Borel sets of $\mathbb{R}^d$, $\int \nu(\mathbf{x} + b\tilde{C})d\mathbf{x} = \lambda(b\tilde{C})$,

$$\mathrm{E}q_{2\mathbf{n}} \le b^{-d}\int(\mathrm{E}\mu^{(\mathbf{m})}_{\mathbf{n}}(C^*_x) + \mu^{(\mathbf{m})}(C^*_x))dx = 2b^{-d}\int \mu^{(\mathbf{m})}(C^*_x)dx = 2b^{-d}\lambda(b\tilde{C}) < \varepsilon$$

for N_0 large enough.
(b) Turning to $q_{1\mathbf{n}}$, let $\tilde{a} > 0$, $S := \prod_{i=1}^d[-\tilde{a}, \tilde{a}]$, and $T := \prod_{i=1}^d[-2\tilde{a}, 2\tilde{a}]$. Choose $\tilde{a}$ large enough so that $\mu^{(\mathbf{m})}(S^c) < \varepsilon/2$ for $\mathbf{m} = \mathbf{m}(\mathbf{n})$ large enough (S^c stands for the complement of S in $\mathbb{R}^d$). Let $\emptyset$ denote the empty set. In the following, we let $\mathbf{n}$, hence $\mathbf{m}$, be large enough. Define

$$\mathcal{F} := \{B \in \Psi : B \cap S \neq \emptyset,\ \mu^{(\mathbf{m})}(B) < (\epsilon/2)\lambda(B)/\lambda(T)\}, \quad \text{and} \quad E := S^c \cup (\cup_{B\in\mathcal{F}}B),$$

where λ denotes the d-dimensional Lebesgue measure. Note that $\cup_{B\in\mathcal{F}}B \subseteq T$ for $b < N_0\tilde{a}$: thus, $\mu^{(\mathbf{m})}(\cup_{B\in\mathcal{F}}B) < \epsilon/2$ and $\mu^{(\mathbf{m})}(E) < \epsilon$.

Next, define $\mathcal{G} := \{B \in \Psi : B \cap S \neq \emptyset, B \notin \mathcal{F}\}$. Clearly, $\mathcal{G} \cup \mathcal{F} = \{B \in \Psi : B \cap S \neq \emptyset\}$, $\Psi - \mathcal{G} = \{B \in \Psi,\ B \cap S = \emptyset \text{ or } B \in \mathcal{F}\}$, and $\cup_{B\in(\Psi-\mathcal{G})}B \subseteq E$. Since the collection of sets $B \in \Psi$ which are subsets of $x + bA$ is no larger than the collection of all sets $B \in \Psi$,

$$\begin{aligned}
q_{1\mathbf{n}} &= b^{-d}\int \sum_{\substack{B\in\Psi \\ B\subseteq \mathbf{x}+bA}} \left|\mu^{(\mathbf{m})}_{\mathbf{n}}(B) - \mu^{(\mathbf{m})}(B)\right| d\mathbf{x} \qquad (4.6)\\
&= b^{-d}\int \sum_{\substack{B\in\Psi \\ B\subseteq \mathbf{x}+bA}} \left|\mu^{(\mathbf{m})}_{\mathbf{n}}(B\cap(\mathbf{x}+bA)) - \mu^{(\mathbf{m})}(B\cap(\mathbf{x}+bA))\right| d\mathbf{x}\\
&\le b^{-d}\int \sum_{B\in\Psi} \left|\mu^{(\mathbf{m})}_{\mathbf{n}}(B\cap(\mathbf{x}+bA)) - \mu^{(\mathbf{m})}(B\cap(\mathbf{x}+bA))\right| d\mathbf{x}\\
&= b^{-d}\int \sum_{B\in\Psi} \Big\{(\mu^{(\mathbf{m})}_{\mathbf{n}} - \mu^{(\mathbf{m})})^+(B\cap(\mathbf{x}+bA)) + (\mu^{(\mathbf{m})}_{\mathbf{n}} - \mu^{(\mathbf{m})})^-(B\cap(\mathbf{x}+bA))\Big\} d\mathbf{x}.
\end{aligned}$$

Hence, by the definitions of $\mu^{(\mathbf{m})}_{\mathbf{n}}$ and $\mu^{(\mathbf{m})}$,

$$\begin{aligned}
q_{1\mathbf{n}} &\le b^{-d}\sum_{B\in\Psi}\Big\{\int\int I_B(\mathbf{y})I_{\mathbf{x}+bA}(\mathbf{y})(\mu^{(\mathbf{m})}_{\mathbf{n}} - \mu^{(\mathbf{m})})^+(d\mathbf{y})d\mathbf{x}\\
&\qquad + \int\int I_B(\mathbf{y})I_{\mathbf{x}+bA}(\mathbf{y})(\mu^{(\mathbf{m})}_{\mathbf{n}} - \mu^{(\mathbf{m})})^-(d\mathbf{y})\, d\mathbf{x}\Big\}\\
&= b^{-d}\sum_{B\in\Psi}\Big\{\lambda(bA)\int I_B(\mathbf{y})(\mu^{(\mathbf{m})}_{\mathbf{n}} - \mu^{(\mathbf{m})})^+(d\mathbf{y}) + \lambda(bA)\int I_B(\mathbf{y})(\mu^{(\mathbf{m})}_{\mathbf{n}} - \mu^{(\mathbf{m})})^-(d\mathbf{y})\Big\}\\
&\le C\sum_{B\in\Psi}|\mu^{(\mathbf{m})}_{\mathbf{n}}(B) - \mu^{(\mathbf{m})}(B)|,
\end{aligned}$$

since $b^{-d}\lambda(bA)$ is bounded by a constant C. It follows that

$$\begin{aligned} q_{1n} &\leq C \sum_{B\in\mathcal{G}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)| + C \sum_{B\in\Psi-\mathcal{G}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)| \\ &\leq C\Big[\sum_{B\in\mathcal{G}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)| + \mu_{\mathbf{n}}^{(\mathbf{m})}(E) - \mu^{(\mathbf{m})}(E) + 2\mu^{(\mathbf{m})}(E)\Big]. \end{aligned}$$

From the definition of E, it is now easy to see that $\mu^{(\mathbf{m})}(E) < \varepsilon$, so that $q_{1\mathbf{n}}$ tends to 0 in probability if

$$\sum_{B\in\mathcal{G}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)| \xrightarrow{\mathrm{P}} 0, \tag{4.7}$$

and

$$|\mu_{\mathbf{n}}^{(\mathbf{m})}(E) - \mu^{(\mathbf{m})}(E)| \xrightarrow{\mathrm{P}} 0. \tag{4.8}$$

In order to complete the proof of Lemma 4.5, it is thus sufficient to show that condition (4.4) implies the convergences (4.7) and (4.8). This final part of the proof is postponed to Section 6, and relies on a delicate blocking argument and an exponential inequality, which we rather develop in a separate section.

5 Blocking and exponential inequality.

This section is devoted to a nonstandard blocking of spatial Bernoulli random variables, which is crucial in proving that (4.4) implies (4.7) and (4.8), hence in the end of the proof of Lemma 4.5. Without any loss of generality, we may assume that $\mathbf{m}(\mathbf{n}) := (m_1, \dots, m_N) \in \mathbb{Z}^N$ is such that $n_k = (2m_k + 1)q_k(\mathbf{n})$, where q_k, $k = 1, \dots, N$, are positive integers. For all $\boldsymbol{\ell} := (\ell_1, \dots, \ell_N)$ with $l_k \in \{0, 1, \dots, 2m_k\}$, $k = 1, \dots, N$, define

$$Z_{\boldsymbol{\ell}}(B) := \sum_{j_1=1}^{q_1} \cdots \sum_{j_N=1}^{q_N} I_{\{\mathbf{X}^{(\mathbf{m})}_{2m_1(j_1-1)+j_1+\ell_1, \cdots, 2m_N(j_N-1)+j_N+\ell_N} \in B\}}.$$

Then, $Z_{\boldsymbol{\ell}}(B)$ is the sum of a block of independent Bernoulli random variables, and

$$\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B) = (1/\hat{\mathbf{n}}) \sum_{\ell_1=0}^{2m_1} \cdots \sum_{\ell_N=0}^{2m_N} (Z_{\boldsymbol{\ell}}(B) - \mathrm{E}Z_{\boldsymbol{\ell}}(B)).$$

In the notation we introduced in the proof of Lemma 4.5, we then have the two following exponential inequalities.

Lemma 5.1 *For all* $\varepsilon > 0$,

$$\begin{aligned} \mathrm{P}\left[|\mu_{\mathbf{n}}^{(\mathbf{m})}(E) - \mu^{(\mathbf{m})}(E)| > \varepsilon\right] &\leq 3(\prod_{k=1}^{N}(2m_k+1)) \left\{-(\prod_{k=1}^{N} q_k)\varepsilon^2/25\right\} \\ &= 3(\prod_{k=1}^{N}(2m_k+1)) \exp\left\{-(\prod_{k=1}^{N} n_k/(2m_k+1))\varepsilon^2/25\right\}. \end{aligned}$$

Lemma 5.2 *If* $\hat{\mathbf{n}}b^d/\hat{\mathbf{m}} \to \infty$*, then, for all* $\varepsilon > 0$*,*

$$\begin{aligned} \mathrm{P}\left(\sum_{B\in\mathcal{G}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)| > \varepsilon\right) &\le 3(\prod_{k=1}^{N}(2m_k+1))\exp\left\{-(\prod_{k=1}^{N} q_k)\varepsilon^2/25\right\} \\ &= 3(\prod_{k=1}^{N}(2m_k+1))\exp\left\{-(\prod_{k=1}^{N} n_k/(2m_k+1))\varepsilon^2/25\right\}. \end{aligned}$$

Proof. We only prove Lemma 5.2; the proof for Lemma 5.1 is entirely similar. Let $\mathbf{q} := (q_1, \ldots, q_N)$. Since the number of elements in $\mathcal{G}$ is bounded by Cb^{-d}, and since $Cb^{-d}/\hat{\mathbf{q}} = O(\hat{\mathbf{m}}b^{-d}/\hat{\mathbf{n}}) \to 0$ in view of the assumption made,

$$\begin{aligned} \mathrm{P}\left(\sum_{B\in\mathcal{G}} |\mu_{\mathbf{n}}^{(\mathbf{m})}(B) - \mu^{(\mathbf{m})}(B)| > \varepsilon\right) &\le \mathrm{P}\left(\sum_{B\in\mathcal{G}}\left|\sum_{\ell_1=0}^{2m_1}\cdots\sum_{\ell_N=0}^{2m_N}(Z_{\boldsymbol{\ell}}(B) - \mathrm{E}Z_{\boldsymbol{\ell}}(B)\right| \ge \hat{\mathbf{n}}\varepsilon\right) \\ &\le \left(\prod_{k=1}^{N}(2m_k+1)\right)\mathrm{P}\left(\sum_{B\in\mathcal{G}}|Z_{\mathbf{0}}(B) - \mathrm{E}Z_{\mathbf{0}}(B)| \ge \hat{\mathbf{q}}\varepsilon\right) \\ &\le \left(\prod_{k=1}^{N}(2m_k+1)\right) 3\exp\{-\hat{\mathbf{q}}\varepsilon^2/25\}, \end{aligned}$$

where the last inequality is obtained from Lemma 3 of Devroye (1983, page 898). □

Proof of Lemma 4.5 (continued). Assume that condition (4.4) holds. Noting that the exponential bound, in Lemmas 5.2 and 5.1, is $O(\hat{\mathbf{m}}\exp\{-C\hat{\mathbf{n}}\varepsilon^2/\hat{\mathbf{m}}\})$, (4.7) and (4.8) easily follow from these two lemmas. □

Summing up, piecing together Lemmas 4.1 through 4.5, we have proved the following result.

Lemma 5.3 *Assume that Assumptions* 2.1 *and* 2.2 *are satisfied. If there exists a sequence* $\hat{\mathbf{m}}(\hat{\mathbf{n}})$ *such that conditions* (4.3) *and* (4.4) *hold, then* $J_{\mathbf{n}} \xrightarrow{\mathrm{P}} 0$ *as* $\mathbf{n} \to \infty$*.*

6 Proof of the main result.

In view of Lemma 5.3, the proofs of the two consistency results of Theorem 2.1 are now straightforward.

Proof of Theorem 2.1. The proof simply consists in exhibiting a particular sequence $\hat{\mathbf{m}}(\hat{\mathbf{n}})$ satisfying conditions *(4.3)* and *(4.4)*.
(a) First consider the system of Assumptions 2.1-2.4. Set $\mathbf{m}(\mathbf{n}) := (m_1^a(\mathbf{n}), \ldots, m_N^a(\mathbf{n}))$, with $m_k^a(\mathbf{n}) := b^{-2(d+1)}$ for $k = 1, \cdots, N$, where a is the positive constant in Assumption 2.3 : this sequence $\mathbf{m}(\mathbf{n})$ clearly satisfies condition (4.3).

As for condition (4.4), note that, for the same sequence $\mathbf{m}(\mathbf{n})$,

$$(\ln \hat{\mathbf{m}})/b^{-d} = (2(d+1)/a)(\ln b^{-1})b^d \to 0\,,$$

hence $(\hat{\mathbf{n}}/\hat{\mathbf{m}})/(\ln \hat{\mathbf{m}}) \ge (\hat{\mathbf{n}}/\hat{\mathbf{m}})/b^{-d}$ diverges to infinity, if $\hat{\mathbf{n}}b^d/\hat{\mathbf{m}} \to \infty$, a condition which is satisfied under Assumption 2.4. Thus condition (4.4) is a consequence of Assumption 2.4.

Hence, conditions (4.3) and (4.4) hold, and the theorem follows from Lemma 5.3.

(b) Next consider the system of Assumptions 2.1-2.2-2.5-2.6. Set $\mathbf{m}(\mathbf{n}) := (m_1(\mathbf{n}), \ldots, m_N(\mathbf{n}))$, with $m_k(\mathbf{n}) := (\hat{\mathbf{n}}b^d/\ln^\beta \hat{\mathbf{n}})^{\alpha/N}$ for $k = 1, \cdots, N$, where $0 < \alpha < 1$ is a positive constant and $\beta = -(N/\alpha)(1-\alpha)$. Then $\hat{\mathbf{m}} = (\hat{\mathbf{n}}b^d/\ln^\beta \hat{\mathbf{n}})^\alpha$, and

$$(\ln \hat{\mathbf{m}})/b^{-d} = \alpha(b^d \ln \hat{\mathbf{n}} + b^d \ln b^d - \beta b^d \ln\ln \hat{\mathbf{n}}) \to 0$$

if $b^d \ln \hat{\mathbf{n}} \to 0$. Thus

$$(\hat{\mathbf{n}}/\hat{\mathbf{m}})/(\ln \hat{\mathbf{m}}) \geq (\hat{\mathbf{n}}/\hat{\mathbf{m}})/b^{-d} = (\hat{\mathbf{n}}b^d/(\ln \hat{\mathbf{n}})^N)^{1-\alpha} \to \infty,$$

and hence condition (4.4) is satisfied under Assumption 2.6.

For condition (4.3), note that, by Assumption 2.5,

$$\begin{aligned} b^{-2(d+1)}v(\mathbf{m}) &= b^{-2(d+1)}\rho^{(\hat{\mathbf{n}}b^d/\ln^\beta \hat{\mathbf{n}})^{\alpha/N}} \\ &= \exp\{-(\hat{\mathbf{n}}b^d/\ln^\beta \hat{\mathbf{n}})^{\alpha/N} \ln \rho^{-1} + \ln b^{-2(d+1)}\} \to 0 \end{aligned}$$

if $(\hat{\mathbf{n}}b^d/\ln^\beta \hat{\mathbf{n}})^{\alpha/N}/\ln b^{-1} \to \infty$. Since $\hat{\mathbf{n}}b^d \to \infty$, so $\ln \hat{\mathbf{n}} > d \ln b^{-1}$ for $\mathbf{n}$ large enough. Thus $\hat{\mathbf{n}}b^d$ is greater than

$$d(\hat{\mathbf{n}}b^d/\ln^\beta \hat{\mathbf{n}})^{\alpha/N}/\ln \hat{\mathbf{n}} = d(\hat{\mathbf{n}}b^d/\ln^N \hat{\mathbf{n}})^{\alpha/N},$$

which tends to ∞ by Assumption 2.6. This, along with Lemma 5.3, completes the proof. □

References

[1] Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability* **21**, 930-934.

[2] Anselin, L. and Florax, R.J.G.M. (1995). *New Directions in Spatial Econometrics*, Springer, Berlin.

[3] Basawa, I.V. (1996a). Special Issue on Spatial Statistics, Part I, *Journal of Statistical Planning and Inference* **50**, 311-411.

[4] Basawa, I.V. (1996b). Special Issue on Spatial Statistics, Part II, *Journal of Statistical Planning and Inference* **51**, 1-97.

[5] Boente, G. and Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *Journal of Multivariate Analysis* **25**, 90-99.

[6] Bosq, D. (1989). Estimation et prévision nonparamétrique d'un processus stationnaire. *Comptes Rendus de l'Académie des Sciences de Paris* T.I **308**, 453-456.

[7] Carbon, M., Hallin, M., and Tran, L.T. (1996). Kernel density estimation for random fields: the L_1 theory, *Journal of Nonparametric Statistics* **6**, 157-170.

[8] Chan, N. H. and Tran, L.T. (1992). Nonparametric tests for serial dependence. *Journal of Time Series Analysis* **13**, 19-28.

[9] Chanda, K.C. (1983). Density estimation for linear processes, *Annals of the Institute of Statistical Mathematics* **35**, Part A, 439-446.

[10] Cressie, N.A.C. (1991). *Statistics for Spatial Data*, Wiley, New York.

[11] Devroye, L. (1983). The equivalence of weak, strong, and complete convergence in L_1 for kernel density estimates. *Annals of Statistics* **11**, 896-904.

[12] Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation : the L_1 View*, Wiley, New York.

[13] Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1990). *Nonparametric Curve Estimation From Time Series.* Lecture Notes in Statistics 60, Springer Verlag, New York.

[14] Hallin, M., Z. Lu, and Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7**, 657-668.

[15] Hallin, M. and Tran, L. T. (1996). Kernel density estimation for linear processes : asymptotic normality and optimal bandwidth selection. *Annals of the Institute of Statistical Mathematics* **48**, 429-449.

[16] Ioannides, D. A. and Roussas, G. G. (1987). Note on the uniform convergence of density estimates for mixing random variables. *Statistics and Probability Letters* **5**, 279-285.

[17] Izenman, A. J. (1991). Recent developments in nonparametric density estimation. *Journal of the American Statistical Association* **86**, 205-224.

[18] Kulkarni, P.M. (1992). Estimation of parameters of a two-dimensional spatial autoregressive model with regression. *Statistics and Probability Letters* **15**, 157-162.

[19] Masry, E. and Györfi (1987). Strong consistency and rates for recursive density estimators for stationary mixing processes. *Journal of Multivariate Analysis* **22**, 79-93.

[20] Possolo, A. (1991). *Spatial Statistics and Imaging*, I.M.S. Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, California.

[21] Ripley, B. (1981). *Spatial Statistics*, Wiley, New York.

[22] Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* **4**, 185-207.

[23] Robinson, P.M. (1987). Time series residuals with application to probability density estimation. *Journal of Time Series Analysis* **8**, 329-344.

[24] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*, Birkhauser, Boston.

[25] Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *Journal of Statistical Planning and Inference* **18**, 135-149.

[26] Tjøstheim, D. (1987). Spatial series and time series : similarities and differences. In *Spatial Processes and Spatial Time Series Analysis* (F. Droesbeke, Ed.), FUSL, Brussels, pp. 217-228.

[27] Tran, L. T. (1989). The L_1 convergence of kernel density estimates under dependence. *Canadian Journal of Statistics* **17**, 197-208.

[28] Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis* **34**, 37-53.

[29] Tran, L. T. (1992). Kernel density estimation for linear processes. *Stochastic Processes and their Applications* **41**, 281-296.

[30] Tran, L. T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *Journal of Multivariate Analysis* **44**, 23-46.

[31] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434-449.

[32] Whittle, P. (1963). Stochastic process in several dimensions. *Bulletin of the International Statistical Institute* **40**, 974-85.

[33] Yakowitz, S. (1987). Nearest-neighbor methods for time series analysis. *Journal of Time Series Analysis* **8**, 235-247.

Institut de Statistique, Campus de la Plaine CP 210,
Université Libre de Bruxelles, B-1050 Bruxelles
Belgium

Academy of Mathematics and System Sciences,
Chinese Academy of Sciences, Beijing 100080,
China

Department of Mathematics,
Indiana University, Bloomington, IN 47405
U.S.A.