ELSEVIER

Statistics & Probability Letters 36 (1997) 115–125

STATISTICS & PROBABILITY LETTERS

Kernel density estimation for random fields (Density estimation for random fields)

Michel Carbon [a], Lanh Tat Tran [b,*,1], Berlin Wu [c]

[a] Université de Lille, France
[b] Department of Mathematics, Indiana University, Bloomington, IN 47405, USA
[c] National Chengchi University, Taiwan

Received 1 May 1996; received in revised form 1 March 1996

Abstract

Kernel-type estimators of the multivariate density of stationary random fields indexed by multidimensional lattice points space are investigated. Sufficient conditions for kernel estimators to converge uniformly are obtained. The estimators can attain the optimal rates L_∞ of convergence. The results apply to a large class of spatial processes. © 1997 Elsevier Science B.V.

AMS classification: Primary: 62G05; Secondary: 60J25

Keywords: Random field; Kernel; Bandwidth

1. Introduction

Our goal in this paper is to study density estimation for random variables which show spatial interaction. We sense a practical need for nonparametric spatial estimation for situations in which parametric families cannot be adopted with confidence.

Denote the integer lattice points in the N-dimensional Euclidean space by $\mathbb{Z}^N$, $N \geqslant 1$. Consider a strictly stationary random field $\{X_\mathbf{n}\}$ indexed by $\mathbf{n}$ in $\mathbb{Z}^N$ and defined on some probability space $(\Omega, \mathscr{F}, P)$. A point $\mathbf{n}$ in $\mathbb{Z}^N$ will be referred to as a site. We will write n instead of $\mathbf{n}$ when $N=1$. For two finite sets of sites S and S', the Borel fields $\mathscr{B}(S) = \mathscr{B}(X_\mathbf{n},\ \mathbf{n} \in S)$ and $\mathscr{B}(S') = \mathscr{B}(X_\mathbf{n},\ \mathbf{n} \in S')$ are the σ-fields generated by the random variables $X_\mathbf{n}$ with $\mathbf{n}$ ranging over S and S', respectively. Denote the Euclidean distance between S and S' by $\operatorname{dist}(S,S')$. We will assume that $X_\mathbf{n}$ satisfies the following mixing condition: there exists a function $\varphi(t) \downarrow 0$ as $t \to \infty$, such that whenever $S, S' \subset \mathbb{Z}^N$,

$$\begin{aligned}\alpha(\mathscr{B}(S), \mathscr{B}(S')) &= \sup\{|P(AB) - P(A)P(B)|,\ A \in \mathscr{B}(S),\ B \in \mathscr{B}(S')\} \\ &\leqslant h(\operatorname{Card}(S), \operatorname{Card}(S'))\varphi(\operatorname{dist}(S,S')), \qquad (1.1)\end{aligned}$$

* Corresponding author. Tel.: +1 812 855 7489.

[1] Research partially supported by the U.S. National Science Foundation Grant DMS-9403718.

PII S0167-7152(97)00054-0

where Card(S) denotes the cardinality of S. Here h is a symmetric positive function nondecreasing in each variable. Throughout the paper, assume that h satisfies either

$$h(n,m) \leqslant \min\{m,n\} \tag{1.2}$$

or

$$h(n,m) \leqslant C(n+m+1)^{\tilde{k}} \tag{1.3}$$

for some $\tilde{k} \geqslant 1$ and some $C > 0$. If $h \equiv 1$, then $X_{\mathbf{n}}$ is called strongly mixing. Conditions (1.2) and (1.3) are the same as the mixing conditions used by Neaderhouser (1980) and Takahata (1983), respectively, and are weaker than the uniform mixing condition used by Nahapetian (1980). They are satisfied by many spatial models. Examples can be found in Neaderhouser (1980), Rosenblatt (1985) and Guyon (1987). For relevant works on random fields, see, e.g., Neaderhouser (1980), Bolthausen (1982), Guyon and Richardson (1984), Guyon (1987) and Nahapetian (1987).

Denote by $I_{\mathbf{n}}$ a rectangular region defined by

$$I_{\mathbf{n}} = \{\mathbf{i}\colon\ \mathbf{i} \in \mathbb{Z}^N,\ 1 \leqslant i_k \leqslant n_k,\ k = 1,\ldots,N\}.$$

Assume that we observe $\{X_{\mathbf{n}}\}$ on $I_{\mathbf{n}}$. Suppose $X_{\mathbf{n}}$ takes values in $\mathbb{R}^d$ and has density f. The letter C will be used to denote constants whose values are unimportant. We write $\mathbf{n} \to \infty$ if

$$\min\{n_k\} \to \infty \quad \text{and} \quad |n_j/n_k| < C \tag{1.4}$$

for some $0 < C < \infty$, $1 \leqslant j,k \leqslant N$. Define $\hat{\mathbf{n}} = n_1 \ldots n_N$. All limits are taken as $\mathbf{n} \to \infty$ unless indicated otherwise. We use x to denote a fixed point of $\mathbb{R}^d$. The integer part of a number a is denoted by $[a]$.

The kernel density estimator $f_{\mathbf{n}}$ of $f(x)$ is defined by

$$f_{\mathbf{n}}(x) = (\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} \sum_{\mathbf{j} \in I_{\mathbf{n}}} K((x - X_{\mathbf{j}})/b_{\mathbf{n}}),$$

where $b_{\mathbf{n}}$ is a sequence of bandwidths tending to zero as $\mathbf{n}$ tends to infinity.

Density estimation for weakly dependent processes $\{X_n\}$ has generated a considerable amount of interest. See for example, Roussas (1969, 1988), Robinson (1983), and Masry and Györfi (1987). The reason is partly due to the fact that many stochastic processes, among them various useful time series models, satisfy the strong mixing property; and the strong mixing property is relatively easy to check.

The asymptotic normality of $f_{\mathbf{n}}$ has been established by Tran (1990). Recently, Tran and Yakowitz (1993) have investigated nearest neighbors estimators for random fields.

2. Assumptions

Assumption 1. The kernel function K is a density function on $\mathbb{R}^d$ and $\int \|z\| K(x)\,\mathrm{d}x < \infty$. In addition K satisfies a Lipschitz condition $|K(x) - K(y)| < C\|x - y\|$, where $\|\cdot\|$ is the usual norm on $\mathbb{R}^d$.

Since K is Lipschitz and integrable, $\sup_{x \in R^d} K(x) < \tilde{K}$, for some constant $\tilde{K}$.

Assumption 2. The density f satisfies a Lipschitz condition $|f(x) - f(y)| < C\|x - y\|$.

Assumption 3. The joint probability density $f_{\mathbf{i},\mathbf{j}}(x,y)$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ exists and satisfies

$$|f_{\mathbf{i},\mathbf{j}}(x,y) - f(x)f(y)| \leqslant C$$

for some constant C and for all x, y and $\mathbf{i}$, $\mathbf{j}$.

We will implicitly assume that Assumptions 1–3 hold. The letter D denotes an arbitrary compact set in $\mathbb{R}^d$. Denote

$$\Psi_{\mathbf{n}} = (\log \hat{\mathbf{n}}(\hat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1})^{1/2}. \tag{2.1}$$

3. The polynomial case

In this section, we will consider the case where $\varphi(i)$ tends to zero at a polynomial rate, that is,

$$\varphi(i) \leqslant Ci^{-\theta}, \tag{3.1}$$

for some $\theta > 0$.

Denote

$$\theta_1 = \frac{d(dN+2N+\theta)}{\theta-(d+2)N}, \qquad \theta_2 = \frac{N(d+2)-\theta}{\theta-(d+2)N},$$

$$\theta_3 = \frac{d(dN+3N+\theta)}{\theta-(d+1+2\tilde{k})N}, \qquad \theta_4 = \frac{N(d+1)-\theta}{\theta-(d+1+2\tilde{k})N}.$$

Our main effort is devoted to proving the following result giving sharp rates of convergence of $f_{\mathbf{n}}$ to f. Its proof is deferred to Section 7.

Theorem 3.1. *Suppose* (3.1) *holds.*

(i) *If* (1.2) *is satisfied and*

$$\hat{\mathbf{n}}b_{\mathbf{n}}^{\theta_1}(\log \hat{\mathbf{n}})^{\theta_2} \to \infty, \tag{3.2}$$

(ii) *or if* (1.3) *is satisfied and*

$$\hat{\mathbf{n}}b_{\mathbf{n}}^{\theta_3}(\log \hat{\mathbf{n}})^{\theta_4} \to \infty \tag{3.3}$$

then

$$\sup_{x\in D} |f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)| = \mathrm{O}(\Psi_{\mathbf{n}}) \quad \textit{in probability.} \tag{3.4}$$

Remark 3.1. For (3.2) to hold it is necessary that $\theta > (d+2)N$ since $b_{\mathbf{n}}$ goes to zero. Hence $\theta_1 > d$ and (3.2) implies

$$\hat{\mathbf{n}}b_{\mathbf{n}}^d \to \infty. \tag{3.5}$$

Similarly, it can be shown that (3.3) implies (3.5).

The convergence of $Ef_{\mathbf{n}}(x)$ to $f(x)$ behaves as in the independent case and is now given.

Lemma 3.1. *Assume*

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{2+d} (\log \hat{\mathbf{n}})^{-1} \to 0. \tag{3.6}$$

Then

$$|Ef_{\mathbf{n}}(x) - f(x)| = o(\Psi_{\mathbf{n}}).$$

Proof. A simple computation shows that (3.6) implies that $b_{\mathbf{n}} = o(\Psi_{\mathbf{n}})$. By Assumptions 1 and 2,

$$|Ef_{\mathbf{n}}(x) - f(x)| \leqslant Cb_{\mathbf{n}} \int \|z\| K(z)\, dz \leqslant Cb_{\mathbf{n}} = o(\Psi_{\mathbf{n}}). \qquad \square$$

A combination of Theorem 3.1 and Lemma 3.1 yields,

Theorem 3.2. *Suppose* (3.1) *holds. If* (3.2) *and* (3.6) *hold, or* (3.3) *and* (3.6) *hold, then*

$$\sup_{x \in D} |f_{\mathbf{n}}(x) - f(x)| = O(\Psi_{\mathbf{n}}) \quad \textit{in probability}.$$

Example 3.1. Consider the case where φ satisfies (1.2) and choose the popular bandwidth $b_{\mathbf{n}} = \hat{\mathbf{n}}^{-a}$ where $a > 0$ is specified below. Clearly, (3.2) is satisfied if

$$a < \frac{\theta - (d+2)N}{d(dN + 2N + \theta)} \equiv \bar{a}$$

and 3.6 is satisfied if

$$a > 1/(2+d) \equiv \underline{a}.$$

Such an a exists if $\bar{a} > \underline{a}$. Suppose $\theta = 20$, $d = 1$, $N = 2$ then $\bar{a} = \frac{7}{8}$ and $\underline{a} = \frac{1}{3}$ so any $\frac{1}{3} < a < \frac{7}{8}$ would work.

Consider the case $N = 1$. We write $\hat{\mathbf{n}}$ as n. Taking the limit of $\bar{a}$ as $\theta \to \infty$, it is easily seen that Condition 3.2 is marginally close to the condition $nb_n^d \to \infty$ normally assumed in the i.i.d. case.

Define

$$\theta_1^* = \frac{d(dN + 2N + \theta)}{\theta - (d+4)N}, \qquad \theta_2^* = \frac{N(d+2) - \theta}{\theta - (d+4)N},$$

$$\theta_3^* = \frac{d(dN + 3N + \theta)}{\theta - (d + 2 + 2\tilde{k})N}, \qquad \theta_4^* = \frac{N(d+1) - \theta}{\theta - (d + 2 + 2\tilde{k})N}.$$

Let ε be an arbitrary small positive number and denote $g(\mathbf{n}) = \prod_{i=1}^{N} (\log n_i)(\log\log n_i)^{1+\varepsilon}$. Clearly, $\sum 1/(\hat{\mathbf{n}} g(\mathbf{n})) < \infty$, where the summation is over all $\mathbf{n}$ in $\mathbb{Z}^N$.

Theorem 3.3. *Suppose* (3.1) *and* (3.6) *hold.*
(i) *If* 1.2 *is satisfied and*

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_1^*} (\log \hat{\mathbf{n}})^{\theta_2^*} g(\mathbf{n})^{-2N/(\theta - (d+4)N)} \to \infty, \tag{3.7}$$

(ii) *or if* 1.3 *is satisfied and*

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_3^*} (\log \hat{\mathbf{n}})^{\theta_4^*} g(\mathbf{n})^{-2N/(\theta - (d+2+2\tilde{k})N)} \to \infty. \tag{3.8}$$

Then

$$\sup_{x\in D}|f_{\mathbf{n}}(x)-f(x)|=\mathrm{O}(\Psi_{\mathbf{n}})\quad a.s.$$

The proof of Theorem 3.3 is presented in Section 7.

4. Preliminaries

The next lemma can be found in Tran (1990) (see also Ibragimov and Linnik, 1971).

Lemma 4.1. (i) *Suppose* (1.1) *holds. Denote by* $\mathscr{L}_r(\mathscr{F})$ *the class of* $\mathscr{F}$*-measurable r.v.'s* X *satisfying* $\|X\|_r=(E|X|^r)^{1/r}<\infty$. *Suppose* $X\in\mathscr{L}_r(\mathscr{B}(S))$ *and* $Y\in\mathscr{L}_s(\mathscr{B}(S'))$. *Assume also that* $1\leqslant r,s,t<\infty$ *and* $r^{-1}+s^{-1}+t^{-1}=1$. *Then*

$$|\mathrm{EXY\text{-}EXEY}|\leqslant C\|X\|_r\|Y\|_s\{h(\mathrm{Card}(S),\mathrm{Card}(S'))\varphi(\mathrm{dist}\,(S,S'))\}^{1/t} \tag{4.1}$$

(ii) *For r.v.'s bounded with probability* 1, *the right-hand side of* (4.1) *can be replaced by* $Ch(\mathrm{Card}(S),\mathrm{Card}(S'))\varphi(\mathrm{dist}\,(S,S'))$.

Denote by $K_{\mathbf{n}}(x)$ the averaging kernel

$$K_{\mathbf{n}}(x)=(1/b_{\mathbf{n}}^d)K(x/b_{\mathbf{n}}).$$

Then

$$f_{\mathbf{n}}(x)=\frac{1}{\hat{\mathbf{n}}}\sum K_{\mathbf{n}}(x-X_{\mathbf{j}}).$$

Define

$$\Delta_{\mathbf{i}}(x)=\hat{\mathbf{n}}^{-1}(K_{\mathbf{n}}(x-X_{\mathbf{i}})-EK_{\mathbf{n}}(x-X_{\mathbf{i}})), \tag{4.2}$$

$$I_{\mathbf{n}}(x)=\sum_{\mathbf{i}\in I_{\mathbf{n}}}E(\Delta_{\mathbf{i}}(x))^2 \quad\text{and}\quad R_{\mathbf{n}}(x)=\sum_{\substack{\mathbf{j}\in I_{\mathbf{n}}\\ i_k\neq j_k \text{ for some } k}}\sum_{\mathbf{i}\in I_{\mathbf{n}}}|\mathrm{Cov}\{\Delta_{\mathbf{i}}(x),\Delta_{\mathbf{j}}(x)\}|.$$

Lemma 4.2. *If* (3.1) *holds for* $\theta>2N$, *then*

$$\sum_{i=1}^{\infty}i^{N-1}(\varphi(i))^a<\infty \tag{4.3}$$

for some $0<a<\frac{1}{2}$.

Proof. Choose an a such that $\frac{1}{2}>a>N/\theta$. Then (4.3) holds. □

Lemma 4.3. *If* (3.1) *holds for* $\theta>2N$, *then*

$$\lim\hat{\mathbf{n}}b_{\mathbf{n}}^d(I_{\mathbf{n}}(x)+R_{\mathbf{n}}(x))<C,$$

where C *is a constant independent of* x.

The proof of Lemma 4.3 follows from Lemma 2.2 of Tran (1990). The following lemma of Rio (1993) will be needed in the sequel. Its proof is found in Rio (1995) (see Theorem 4).

Lemma 4.4. *Suppose $\mathscr{A}$ is a σ-field of $(\Omega, \mathscr{F}, P)$ and X is a real-valued random variable taking a.s. its values in $[a,b]$. Suppose furthermore that there exists a random variable U with uniform distribution over $[0,1]$, independent of $\mathscr{A} \vee \sigma(X)$. Then there exists some random variable X^* independent of $\mathscr{A}$ and with the same distribution as X such that*

$$E|X - X^*| \leqslant 2(b-a)\alpha(\mathscr{A}, \sigma(X)).$$

Moreover, X^* is a $\mathscr{A} \vee \sigma(X) \vee \sigma(U)$-measurable random variable.

The approximation of strongly mixing r.v.'s by independent ones used later is presented below. Its proof is given in Section 7.

Lemma 4.5. *Suppose $S_1, S_2, \ldots, S_r$ be sets containing m sites each with $\operatorname{dist}(S_i, S_j) \geqslant \delta$ for all $i \neq j$ where $1 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant r$. Suppose $Y_1, Y_2, \ldots, Y_r$ is a sequence of real-valued r.v.'s measurable with respect to $\mathscr{B}(S_1), \mathscr{B}(S_2), \ldots, \mathscr{B}(S_r)$, respectively, and Y_i takes values in $[a,b]$. Then there exists a sequence of independent r.v.'s Y_1^*, $Y_2^*, \ldots, Y_r^*$ independent of $Y_1, Y_2, \ldots, Y_r$ such that Y_i^* has the same distribution as Y_i and satisfies*

$$\sum_{i=1}^{r} E|Y_i - Y_i^*| \leqslant 2r(b-a)h((r-1)m, m)\varphi(\delta). \tag{4.4}$$

5. Uniform convergence of the kernel estimator

Choose

$$\ell = b_{\mathbf{n}}^{(d+1)} \Psi_{\mathbf{n}}. \tag{5.1}$$

Since D is compact, it can be covered with, say v cubes I_k having sides of length ℓ and center at x_k. Clearly,

$$v \leqslant C(b_{\mathbf{n}}^{(d+1)} \Psi_{\mathbf{n}})^{-d}. \tag{5.2}$$

Define

$$Q_{1\mathbf{n}} = \max_{1 \leqslant k \leqslant v} \sup_{x \in I_k} |f_{\mathbf{n}}(x) - f_{\mathbf{n}}(x_k)|,$$

$$Q_{2\mathbf{n}} = \max_{1 \leqslant k \leqslant v} \sup_{x \in I_k} |Ef_{\mathbf{n}}(x_k) - Ef_{\mathbf{n}}(x)|,$$

$$Q_{3\mathbf{n}} = \max_{1 \leqslant k \leqslant v} |f_{\mathbf{n}}(x_k) - Ef_{\mathbf{n}}(x_k)|.$$

Then

$$\sup_{x \in D} |f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)| \leqslant Q_{1\mathbf{n}} + Q_{2\mathbf{n}} + Q_{3\mathbf{n}}.$$

Lemma 5.1. *Both $Q_{1\mathbf{n}}$ and $Q_{2\mathbf{n}}$ equal* $\mathrm{O}(\Psi_{\mathbf{n}})$ *a.s.*

Proof. By Assumption 1, the kernel K satisfies the Lipschitz condition. Therefore

$$|f_{\mathbf{n}}(x) - f_{\mathbf{n}}(x_k)| \leqslant C b_{\mathbf{n}}^{-(d+1)} \|x - x_k\| \leqslant C b_{\mathbf{n}}^{-(d+1)} \ell = \mathrm{O}(\Psi_{\mathbf{n}}) \quad \text{a.s.}$$

The lemma easily follows. □

We proceed to show

$$Q_{3\mathbf{n}} = \mathrm{O}(\Psi_{\mathbf{n}}) \quad \text{in probability.} \tag{5.3}$$

Define

$$S_{\mathbf{n}}(x) = \sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k} \Delta_{\mathbf{i}}(x). \tag{5.4}$$

Then

$$S_{\mathbf{n}}(x) = f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x). \tag{5.5}$$

Showing (5.3) is equivalent to showing that

$$\max_{1\leqslant k\leqslant v} |S_{\mathbf{n}}(x_k)| = \mathrm{O}(\Psi_{\mathbf{n}}) \quad \text{a.s.} \tag{5.6}$$

Without loss of generality assume that $n_i = 2pq_i$ for $1 \leqslant i \leqslant N$. The random variables $\Delta_{\mathbf{i}}(x)$ can be grouped into $2^N q_1 \times q_2 \times \cdots \times q_N$ cubic blocks of side p. Denote

$$U(1,\mathbf{n},\mathbf{j},x) = \sum_{\substack{i_k=2j_k p+1\\k=1,\ldots,N}}^{(2j_k+1)p} \Delta_{\mathbf{i}}(x), \tag{5.7}$$

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

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

$$U(4,\mathbf{n},\mathbf{j},x) = \sum_{\substack{i_k=2j_k p+1\\k=1,\ldots,N-2}}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{i}}(x),$$

and so on. Note that

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

Finally,

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

For each integer $1 \leqslant i \leqslant 2^N$, define

$$T(\mathbf{n},i,x) = \sum_{\substack{j_k=0\\k=1,\ldots,N}}^{q_k-1} U(i,\mathbf{n},\mathbf{j},x).$$

Clearly,

$$S_{\mathbf{n}}(x)=\sum_{i=1}^{2^N} T(\mathbf{n},i,x). \tag{5.8}$$

The blocking idea here is reminiscent of the blocking scheme in Tran (1990) and Politis and Romano (1993).

By (5.8), to establish (5.6), it is sufficient to show that

$$\max_{1\leqslant k\leqslant v}|T(\mathbf{n},i,x_k)|=\mathrm{O}(\Psi_{\mathbf{n}}) \quad \text{a.s.} \tag{5.9}$$

for each $1\leqslant i\leqslant 2^N$. Without loss of generality we will show (5.9) for $i=1$. Now, $T(\mathbf{n},1,x)$ is the sum of

$$r=q_1\times q_2\times\cdots\times q_N \tag{5.10}$$

of the $U(1,\mathbf{n},\mathbf{j},x)$'s. Note that $U(1,\mathbf{n},\mathbf{j},x)$ is measurable with the σ-field generated by $X_{\mathbf{i}}$ with $\mathbf{i}$ belonging to the set of sites

$$\{\mathbf{i}: 2j_kp+1\leqslant i_k\leqslant(2j_k+1)p,\ k=1,\dots,N\}.$$

These sets of sites are separated by a distance of at least p. Enumerate the r.v.'s $U(1,\mathbf{n},\mathbf{j},x)$ and the corresponding σ-fields with which they are measurable in an arbitrary manner and refer to them respectively as $Y_1,Y_2,\dots,Y_r$ and $S_1,S_2,\dots,S_r$. Approximate $Y_1,Y_2,\dots,Y_r$ by the r.v.'s $Y_1^*,Y_2^*,\dots,Y_r^*$ as was done in Lemma 4.5. Clearly,

$$|Y_i|<Cp^N(\hat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1}\tilde{K}. \tag{5.11}$$

Denote

$$\varepsilon_{\mathbf{n}}=\eta\Psi_{\mathbf{n}},$$

where η is a constant to be chosen later.

Define $\beta_{1\hat{\mathbf{n}}}=b_{\mathbf{n}}^{-d}h(\hat{\mathbf{n}},p^N)\varphi(p)\Psi_{\mathbf{n}}^{-1}$. Then the following results holds whose justification is given in Section 7.

Lemma 5.2. *Given an arbitrarily large positive constant a, there exists a positive constant C such that*

$$P\left[\max_{1\leqslant k\leqslant v}|T(\mathbf{n},1,x_k)|>\varepsilon_{\mathbf{n}}\right]\leqslant Cv\left(\hat{\mathbf{n}}^{-a}+\beta_{1\hat{\mathbf{n}}}\right).$$

Theorem 3.1 follows easily from the development of the lemmas of this section. Its proof is given in Section 7.

6. The exponential case

Assume that for some $s>0$,

$$\varphi(i)=C\exp\{-si\}. \tag{6.1}$$

Theorem 6.1. *Assume* (6.1) *holds.*

(i) *If*

$$\hat{\mathbf{n}}b_{\mathbf{n}}^d(\log\hat{\mathbf{n}})^{-2N-1}\to\infty, \tag{6.2}$$

then $\sup_{x\in D}|f_{\mathbf{n}}(x)-Ef_{\mathbf{n}}(x)|=\mathrm{O}(\Psi_{\mathbf{n}})$ *a.s.*

(ii) *If in addition* (3.6) *holds, then* $\sup_{x\in D}|f_{\mathbf{n}}(x)-f(x)|=\mathrm{O}(\Psi_{\mathbf{n}})$ *a.s.*

7. Proofs

Proof of Theorem 3.1. To complete the proof, we will show that $v\hat{\mathbf{n}}^{-a} \to 0$ and $v\beta_{1\hat{\mathbf{n}}} \to 0$. Using (5.2) and (2.1),

$$v \leqslant C\,\hat{\mathbf{n}}^{d/2} b_{\mathbf{n}}^{-d((d/2)+1)} (\log \hat{\mathbf{n}})^{-d/2}.$$

Relation (3.5) implies $\hat{\mathbf{n}} > C b_{\mathbf{n}}^{-d}$ or $b_{\mathbf{n}}^{-d((d/2)+1)} \leqslant C\,\hat{\mathbf{n}}^{(d/2)+1}$. Therefore

$$v\,\hat{\mathbf{n}}^{-a} \leqslant C\,\hat{\mathbf{n}}^{d+1-a} (\log \hat{\mathbf{n}})^{-d/2},$$

which goes to 0 if $a > d+1$.

Using (5.2),

$$v\beta_{1\hat{\mathbf{n}}} \sim b_{\mathbf{n}}^{-d(d+2)} h(\hat{\mathbf{n}}, p^N) p^{-\theta} \Psi_{\mathbf{n}}^{-(d+1)}. \tag{7.1}$$

A computation using (1.2), (2.1) and (7.1) shows that (3.2) is equivalent to $(v\beta_{1\hat{\mathbf{n}}})^{-1} \to \infty$ which implies $v\beta_{1\hat{\mathbf{n}}} \to 0$. Analogously, (1.3), (2.1), and (7.1) show that (3.3) is equivalent to $v\beta_{1\hat{\mathbf{n}}} \to 0$.

Proof of Lemma 4.5. Suppose δ_j, $j \geqslant 1$ is a sequence of i.i.d. uniform $[0,1]$ r.v.'s independent of Y_j, $j \geqslant 1$. Define $Y_1^* = Y_1$. By Lemma (4.4), for every $i \geqslant 2$, there exists a measurable function f_i such that $Y_i^* = f_i(Y_1, \ldots, Y_i, \delta_i)$. In addition, each Y_i^* is independent of $Y_1, \ldots, Y_{i-1}$, has the same distribution as Y_i and satifies

$$E|Y_i - Y_i^*| \leqslant 2(b-a)\alpha(\sigma(Y_\ell\colon \ell < i-1), \sigma(Y_i)) \leqslant 2(b-a)h((i-1)m, m)\varphi(\delta).$$

The last inequality follows by using (1.1). For $1 \leqslant i \leqslant r$, we have $h((i-1)m, m) \leqslant h(rm, m)$ since h is non-decreasing in each variable as stated in the introduction and (4.4) follows by summing up on $1 \leqslant i \leqslant r$.

It remains to show that $Y_1^*, \ldots, Y_r^*$ are independent. To prove this it is sufficient to show that Y_i^* and $(Y_1^*, \ldots, Y_{i-1}^*)$ are independent for $i > 1$. Note that $(Y_1, \ldots, Y_i)$, $\delta_1, \ldots, \delta_i$ are independent. Thus $(Y_1, \ldots, Y_i, \delta_i)$, $\delta_1, \ldots, \delta_{i-1}$ are independent. Since Y_i^* is a measurable function of $Y_1, \ldots, Y_i, \delta_i$, it follows that $(Y_i^*, Y_1, \ldots, Y_{i-1}), \delta_1, \ldots, \delta_{i-1}$ are independent. Now Y_i^* is independent of $Y_1, \ldots, Y_{i-1}$. Hence $Y_i^*, (Y_1, \ldots, Y_{i-1}), \delta_1, \ldots, \delta_{i-1}$ are independent. Finally Y_i^* and $(Y_1^*, \ldots, Y_{i-1}^*)$ are independent since $(Y_1^*, \ldots, Y_{i-1}^*)$ is measurable with respect to the σ-field generated by $Y_1, \ldots, Y_{i-1}, \delta_1, \ldots, \delta_{i-1}$.

Proof of Lemma 5.2. Since $T(\mathbf{n}, 1, x)$ is equal to $\sum_{i=1}^r Y_i$, we have

$$P[|T(\mathbf{n},1,x)| > \varepsilon_{\mathbf{n}}] \leqslant P\left[\left|\sum_{i=1}^r Y_i^*\right| > \varepsilon_{\mathbf{n}}/2\right] + P\left[\sum_{i=1}^r |Y_i - Y_i^*| > \varepsilon_{\mathbf{n}}/2\right]. \tag{7.2}$$

We now proceed to obtain bounds for the two terms on the right hand side of (7.2).

By Markov's inequality and using (4.4), (5.11) and recall that the sets of sites with respect to which the Y_i's are measurable are separated by a distance of at least p,

$$P\left[\sum_{i=1}^r |Y_i - Y_i^*| > \varepsilon_{\mathbf{n}}\right] \leqslant Crp^N(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} h(\hat{\mathbf{n}}, p^N)\varphi(p)\varepsilon_{\mathbf{n}}^{-1} \sim \beta_{1\hat{\mathbf{n}}}. \tag{7.3}$$

Set

$$\lambda_{\mathbf{n}} = (\hat{\mathbf{n}} b_{\mathbf{n}}^d \log \hat{\mathbf{n}})^{1/2}, \tag{7.4}$$

$$p = \left[\left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^d}{4\lambda_{\mathbf{n}} \tilde{K}} \right)^{1/N} \right] \sim \left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^d}{\log \hat{\mathbf{n}}} \right)^{1/2N}. \tag{7.5}$$

A simple computation yields,

$$\lambda_{\mathbf{n}} \varepsilon_{\mathbf{n}} = \eta \log \hat{\mathbf{n}},$$

and by Lemma (4.3)

$$\lambda_{\mathbf{n}}^2 \sum_{i=0}^{r} E(Y_i^*)^2 \leqslant C \hat{\mathbf{n}} b_{\mathbf{n}}^d (I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x)) \log \hat{\mathbf{n}} < C \log \hat{\mathbf{n}}.$$

Using (5.11), we have $|\lambda_{\mathbf{n}} Y_i^*| < 1/2$ for large $\hat{\mathbf{n}}$. Applying Berstein's inequality,

$$\begin{aligned} P\left[\left| \sum_{i=0}^{r} Y_i^* \right| > \varepsilon_{\mathbf{n}} \right] &\leqslant 2 \exp\left(-\lambda_{\mathbf{n}} \varepsilon_{\mathbf{n}} + \lambda_{\mathbf{n}}^2 \sum_{i=0}^{r} E(Y_i^*)^2 \right) \\ &\leqslant 2 \exp((-\eta + C) \log \hat{\mathbf{n}}) \leqslant \hat{\mathbf{n}}^{-a}, \end{aligned} \tag{7.6}$$

for sufficiently large $\hat{\mathbf{n}}$.

Combining (7.2), (7.3) and (7.6),

$$P\left[\sup_{x \in D} |T(\mathbf{n}, 1, x)| > \varepsilon_{\mathbf{n}} \right] \leqslant C v (\hat{\mathbf{n}}^{-a} + \beta_{1\hat{\mathbf{n}}}).$$

Proof of Theorem 3.3. (i) Condition (3.7) is equivalent to

$$v \beta_{1\hat{\mathbf{n}}} \hat{\mathbf{n}} g(\mathbf{n}) \to 0,$$

which entails $\sum v_{\mathbf{n} \in Z^N} \beta_{1\hat{\mathbf{n}}} < \infty$. The theorem follows by the Borel–Cantelli lemma.

(ii) The proof of (ii) is similar to that of (i) and is omitted.

Proof of Theorem 6.1. We prove (i) only since the proof of (ii) is the same as the proof of (ii) of Theorem 3.1. Condition (6.2) implies that

$$(\hat{\mathbf{n}} b_{\mathbf{n}}^d / \log \hat{\mathbf{n}})^{1/2N} (\log \hat{\mathbf{n}})^{-1} \to \infty.$$

Suppose a is an arbitrarily given large positive number a. For all $\hat{\mathbf{n}}$ except finitely many,

$$(\hat{\mathbf{n}} b_{\mathbf{n}}^d / \log \hat{\mathbf{n}})^{1/2N} \geqslant (a/s) \log \hat{\mathbf{n}}.$$

Therefore

$$\varphi(p) \leqslant C \exp\{-sp\} \leqslant C \exp\{-s(a/s) \log \hat{\mathbf{n}}\} = C \hat{\mathbf{n}}^{-a}, \tag{7.7}$$

where the value of p is given in (7.5). If necessary, the constant C in inequality (7.7) can be increased so that the inequality holds for all $\hat{\mathbf{n}}$. Using (7.1) and (7.7), it is easy to show that $\sum_{\mathbf{n} \in Z^N} v \beta_{1\hat{\mathbf{n}}} < \infty$. The theorem follows by the Borel-Cantelli lemma.

Acknowledgements

The authors would like to thank the Associate Editor and the referee for many useful comments.

References

Bolthausen, E., 1982. On the central limit theorem for stationary random fields. Ann. Probab. 10, 1047–1050.

Guyon, X., Richardson, S., 1984. Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants. Z. Wahrsch. Verw. Gebiete 66, 297–314.

Guyon, X., 1987. Estimation d'un champ par pseudo-vraisemblance conditionnelle: Etude asymptotique et application au cas Markovien. Proc. 6th Franco–Belgian Meeting of Statisticians.

Ibragimov, I.A., Linnik, Yu.V., 1971. Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen.

Masry, E., Györfi, 1987. Strong consistency and rates for recursive density estimators for stationary mixing processes. J. Multivariate Anal. 22, 79–93.

Nahapetian, B.S., In: Dobrushin, R.L., Sinai, Ya. G. (Eds.), 1980. The central limit theorem for random fields with mixing conditions. Adv. in Probability, vol 6, Multicomponent systems, pp. 531–548.

Nahapetian, B.S., 1987. An approach to proving limit theorems for dependent random variables. Theory Probab. Appl. 32, 535–539.

Neaderhouser, C.C., 1980. Convergence of block spins defined on random fields. J. Statist. Phys. 22, 673–684.

Politis, D.N., Romano, J.P., 1993. Nonparametric resampling for homogeneous strong mixing random fields. J. Multivariate Anal. 47, 301–328.

Rio, E., 1995. The functional law of the iterated logarithm for stationary strongly mixing sequences. Ann. Probab. 23, 1188–1203.

Robinson, P.M., 1983. Nonparametric estimators for time series. J. Time Series Anal. 4, 185–207.

Rosenblatt, M., 1985. Stationary sequences and random fields. Birkhauser, Boston.

Roussas, G.G., 1969. Nonparametric estimation of the transition distribution of a Markov process. Ann. Inst. Statist. Math. 21, 73–87.

Roussas, G.G., 1988. Nonparametric estimation in mixing sequences of random variables. J. Statist. Plann. Inference 18, 135–149.

Takahata, H., 1983. On the rates in the central limit theorem for weakly dependent random fields. Z. Wahrsch. Verw. Gebiete 62, 477–480.

Tran, L.T., 1990. Kernel density estimation on random fields. J. Multivariate Anal. 34, 37–53.

Tran, L.T., Yakowitz, S., 1993. Nearest neighbor estimators for random fields. J. Multivariate Anal. 44, 23–46.