KERNEL REGRESSION ESTIMATION FOR RANDOM FIELDS

(Density Estimation for Random Fields)

by

Michel Carbon, Christian Francq and Lanh Tat Tran*

Université de Rennes 2, Ensai, Université de Lille 3, and Indiana University

Dedicated to Madan L. Puri

AMS 1980 subject classifications: primary: 62G05; secondary: 60J25.
Keywords and phrases: random field, kernel, regression estimation, bandwidth.

*Research of L. T. Tran was partially supported by the U. S. National Science Foundation Grant DMS-9403718.

ABSTRACT

Consider a stationary random field $\{X_{\mathbf{n}}\}$ indexed by N-dimensional lattice points, where $\{X_{\mathbf{n}}\}$ takes values in R^d. An important problem in spatial statistics is the estimation of the regression of $\{X_{\mathbf{n}}\}$ on the values of the random field at surrounding sites, say, $X_{\mathbf{n_1}}$, ..., $X_{\mathbf{n}_\ell}$. Note that $(X_{\mathbf{n_1}}, ..., X_{\mathbf{n}_\ell})$ is a ℓd-dimensional vector. Assume the existence of the regression function

$$r(x) = E\left\{\varphi(X_{\mathbf{n}}) \mid (X_{\mathbf{n_1}}, \ldots, X_{\mathbf{n}_\ell}) = x\right\},$$

where φ is a continuous real-valued function which is not necessarily bounded, and $x \in R^{\ell d}$. Kernel-type estimators of the regression function $r(x)$ are investigated. They are shown to converge uniformly on compact sets under general conditions. In addition, they can attain the optimal rates of convergence in L_∞. The results hold for a large class of spatial processes.

Send Correspondence to:
Lanh Tat Tran
Department of Mathematics
Indiana University
Bloomington, IN 47405
(812) 855-7489

1 Introduction

A random field is a collection of random variables indexed by points of Z^N. In the case $N = 1$, a random field reduces to a time series. The case $N > 1$ is useful for stochastic phenomena evolving in more than one direction and showing spatial interaction. The random field theory has been found to be useful in many disciplines, including spatial modeling or image analysis. As in time series analysis, in random field theory there exist situations in which parametric families cannot be adopted with confidence. In this paper we study density and regression nonparametric estimation.

Spatial data arise in a variety of fields and the statistical treatment of such data is the subject of an abundant literature. For background material on this subject, the reader is referred to Ripley (1981), Cressie (1991), Guyon (1995), Hallin, Lu and Tran (2004) and the references there in.

Denote the integer lattice points in the N-dimensional Euclidean space by $Z^N, N \geq 1$. Consider a strictly stationary random field $\{X_{\mathbf{n}}\}$ indexed by $\mathbf{n}$ in Z^N and defined on some probability space $(\Omega, \mathcal{F}, P)$. A point $\mathbf{n}$ in Z^N will be referred to as a site. We will write n instead of $\mathbf{n}$ when $N = 1$. For relevant works on random fields, see for example, Neaderhouser (1980), Bolthausen (1982), Guyon and Richardson (1984), Guyon (1987) and Nahapetian (1987).

Suppose $X_{\mathbf{n}}$ takes values in R^d. Consider the set of the nearest neighbors of the site $\mathbf{n}$. Since the k-th component of a nearest neighbor of $\mathbf{n} = (n_1, \ldots, n_N)$ takes one of the three values $n_k - 1$, n_k and $n_k + 1$, and since a neighbor of $\mathbf{n}$ can not be $\mathbf{n}$ itself, there exist $\ell := 3^N - 1$ nearest neighbors of $\mathbf{n}$. Denote $X_{(\mathbf{n})}$ the random vector of $R^{\tilde{d}}$ defined by $X_{(\mathbf{n})} = (X_{\mathbf{n_1}}, \ldots, X_{\mathbf{n}_\ell})$, where $\mathbf{n_i}$, $i = 1, \ldots, \ell$, are the nearest neighbors of $\mathbf{n}$ (in a given order).

The main goal of this paper is to estimate the regression of $\varphi(X_{\mathbf{n}})$ on $X_{(\mathbf{n})}$, where φ is a continuous real-valued function. In some applications it could be desirable to make a prediction of $\varphi(X_{\mathbf{n}})$ based on other values of the random field, for instance the values of the random field on a given subset of nearest neighbors of $\mathbf{n}$. Therefore consider a finite set of ℓ translations $t_1, \ldots, t_\ell$ into Z^N, where ℓ is an arbitrary positive integer. The choice of the translations is not treated in this paper. This important issue is similar to that of the order identification in ARMA time series analysis. To our knowledge, there is no satisfactory answer to this problem in our general framework.

Suppose that these translations are distinct and different from the identity. So $t(\mathbf{i}) := \{t_1(\mathbf{i}), \ldots, t_\ell(\mathbf{i})\}$ does not contain $\mathbf{i}$ and $\text{Card}\{t(\mathbf{i})\} = \ell$.

Denote

$$X_{(\mathbf{n})} = \left(X_{t_1(\mathbf{n})}, \ldots, X_{t_\ell(\mathbf{n})}\right),$$

and suppose $X_{(\mathbf{n})}$ has density f. Assume the existence of a function r satisfying $r(x) = E\left\{\varphi(X_{\mathbf{n}}) \mid X_{(\mathbf{n})} = x\right\}$, $x \in R^{\tilde{d}}$, with $\tilde{d} = \ell d$.

Without loss of generality, consider a site $\mathbf{n}$ with positive components. Denote by $I_{\mathbf{n}}$ a rectangular region defined by

$$I_{\mathbf{n}} = \{\mathbf{i} : \mathbf{i} \in Z^N, 1 \le i_k \le n_k, k = 1, \ldots, N\}$$

and consider the set

$$\mathcal{I}_{\mathbf{n}} = \{t_j(\mathbf{i}) : \mathbf{i} \in I_{\mathbf{n}}, 1 \le j \le \ell\} = \bigcup_{\mathbf{i} \in I_{\mathbf{n}}} t(\mathbf{i}).$$

Assume that we observe $\{X_{\mathbf{j}}\}$ on $I_{\mathbf{n}}$ and $\mathcal{I}_{\mathbf{n}}$. The letter C will be used to denote constants whose values are unimportant. We write $\mathbf{n} \to \infty$ if

$$\min\{n_k\} \to \infty \quad \text{and} \quad |n_j/n_k| < C \tag{1.1}$$

for some $0 < C < \infty$, $1 \le j, k \le N$. Define $\hat{\mathbf{n}} = n_1 \cdots n_N = \mathrm{Card}(I_{\mathbf{n}})$. All limits are taken as $\mathbf{n} \to \infty$ unless indicated otherwise. We use x to denote a fixed point of $R^{\tilde{d}}$.

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

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

and, for $f_{\mathbf{n}}(x) \neq 0$, the Nadaraya-Watson estimator $r_{\mathbf{n}}(x)$ of $r(x)$ is defined by

$$r_{\mathbf{n}}(x) = \psi_{\mathbf{n}}(x)/f_{\mathbf{n}}(x), \quad \psi_{\mathbf{n}}(x) = (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \sum_{\mathbf{j} \in I_{\mathbf{n}}} \varphi(X_{\mathbf{j}}) K\left\{(x - X_{(\mathbf{j})})/b_{\mathbf{n}}\right\},$$

where K is a kernel function and $b_{\mathbf{n}}$ is a sequence of bandwidths tending to zero as $\mathbf{n}$ tends to infinity. Recently, Hallin, Lu and Tran (2004) have shown that $r_{\mathbf{n}}(x)$ has an asymptotic normal distribution under general assumptions.

The rest of the paper proceeds as follows. Section 2 presents the assumptions and states our mains results. We give conditions ensuring the uniform consistency on any compact set of the kernel density estimator $f_{\mathbf{n}}$ and of the regression estimator $r_{\mathbf{n}}(x)$. Weak consistency and strong consistency are considered, and the rates of convergence are provided. Technical lemmas are collected in Section 3. Section 4 is devoted to the proof of the consistency of the kernel density estimator. The proof of the consistency of the regression estimator is given in Section 5.

2 Assumptions and main results

We first introduce some mixing assumptions. For two finite sets of sites S and S', the Borel fields $\mathcal{B}(S) = \mathcal{B}(X_{\mathbf{n}}, \mathbf{n} \in S)$ and $\mathcal{B}(S') = \mathcal{B}(X_{\mathbf{n}}, \mathbf{n} \in S')$ 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 dist (S, S') and assume that $X_{\mathbf{n}}$ satisfies the following mixing condition: there exists a function $\kappa(t) \downarrow 0$ as $t \to \infty$, such that whenever $S, S' \subset Z^N$,

$$(2.1)\quad \alpha\left\{\mathcal{B}(S), \mathcal{B}(S')\right\} = \sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(S), B \in \mathcal{B}(S')\}$$
$$\leq h\left\{\mathrm{Card}(S), \mathrm{Card}(S')\right\} \kappa\left\{\mathrm{dist}\,(S, S')\right\},$$

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

$$(2.2)\qquad h(n, m) \leq \min\{m, n\}$$

or

$$(2.3)\qquad h(n, m) \leq C(n + m + 1)^{\tilde{k}}$$

for some $\tilde{k} > 0$ and some $C > 0$. If $h \equiv 1$, then $X_{\mathbf{n}}$ is called strongly mixing. Conditions (2.2) and (2.3) 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).

Now we need regularity assumptions on the kernel and on the density f.

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

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

We also need an assumption on the joint density of $X_{(\mathbf{i})}$ and $X_{(\mathbf{j})}$. This assumption is the same as Assumption 3 in Carbon, Tran and Wu (1997), except that we have to assume that $t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset$ (when $t(\mathbf{i}) \cap t(\mathbf{j}) \neq \emptyset$, the joint distribution of $X_{(\mathbf{i})}$ and $X_{(\mathbf{j})}$ does not admit a density with respect to the Lebesgue measure).

Assumption 3. If $t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset$, the joint probability density $f_{X_{(\mathbf{i})}, X_{(\mathbf{j})}}(x, y)$ of $X_{(\mathbf{i})}$ and $X_{(\mathbf{j})}$ exists and satisfies

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

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

We need to introduce additional notations. Let D be an arbitrary compact set in $R^{\tilde{d}}$ and denote

$$\Psi_{\mathbf{n}} = \left(\log \hat{\mathbf{n}} (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1}\right)^{1/2}. \tag{2.4}$$

For the mixing coefficients defined by (2.1), we will distinguish two rates of convergence. We will consider the case where $\kappa(i)$ tends to zero at a polynomial rate, that is,

$$\kappa(i) \leq C i^{-\theta}, \tag{2.5}$$

for some $\theta > 0$. We will also consider the case where $\kappa(i)$ tends to zero at the exponential rate, that is,

$$\kappa(i) \leq C \exp\{-si\}, \tag{2.6}$$

for some $s > 0$.

For mixing coefficients with polynomial decreasing rate (2.5), the constraints on the bandwidth will be related to θ by means of

$$\theta_1 = \frac{\tilde{d}(\theta + (\tilde{d}+3)N)}{\theta - (\tilde{d}+3)N}, \quad \theta_2 = \frac{-\theta + (\tilde{d}+1)N}{\theta - (\tilde{d}+3)N},$$

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

$$\theta_5 = \frac{\tilde{d}(\theta + (\tilde{d}+3)N)}{\theta - (\tilde{d}+5)N}, \quad \theta_6 = \frac{-\theta + (\tilde{d}+1)N}{\theta - (\tilde{d}+5)N},$$

$$\theta_7 = \frac{\tilde{d}(\theta + (\tilde{d}+3)N)}{\theta - (\tilde{d}+3+2\tilde{k})N}, \quad \theta_8 = \frac{-\theta + (\tilde{d}+1)N}{\theta - (\tilde{d}+3+2\tilde{k})N},$$

$$\theta_9 = \frac{-2N}{\theta - (\tilde{d}+5)N}, \quad \theta_{10} = \frac{-2N}{\theta - (\tilde{d}+3+2\tilde{k})N}.$$

Let ϵ be an arbitrary small positive number and denote

$$g(\mathbf{n}) = \prod_{i=1}^{N} (\log n_i)(\log\log n_i)^{1+\epsilon}. \tag{2.7}$$

Note that $\sum \frac{1}{\hat{\mathbf{n}} g(\mathbf{n})} < \infty$, where the summation is over all $\mathbf{n}$ in Z^N. In the sequel, it will be required that the bandwidths sequence satisfy one or several of the following conditions:

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{2+\tilde{d}} (\log \hat{\mathbf{n}})^{-1} \to 0, \tag{2.8}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_1} (\log \hat{\mathbf{n}})^{\theta_2} \to \infty, \tag{2.9}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_3} (\log \hat{\mathbf{n}})^{\theta_4} \to \infty, \tag{2.10}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}} \to \infty, \tag{2.11}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}} (\log \hat{\mathbf{n}})^{-2N-1} \to \infty, \tag{2.12}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_5} (\log \hat{\mathbf{n}})^{\theta_6} g(\mathbf{n})^{\theta_9} \to \infty, \tag{2.13}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_7} (\log \hat{\mathbf{n}})^{\theta_8} g(\mathbf{n})^{\theta_{10}} \to \infty. \tag{2.14}$$

The bandwidth condition (2.8) imposes that $b_{\mathbf{n}} \to 0$ sufficiently quickly. The other bandwidth conditions impose that $b_{\mathbf{n}} \to 0$ not too quickly. For the consistency of the density estimator we require (2.8) and a second bandwidth condition among (2.9)–(2.14). Obviously, this second bandwidth condition is more restrictive for the strong consistency than for the weak consistency. This bandwidth condition is also different under the Takahata condition (2.3) and under the Neaderhouser condition (2.2).

Theorem 2.1. *Assume that Assumptions 1–3 hold.*

(i) Suppose the polynomial mixing rate (2.5) holds. If

- $\theta > (\tilde{d}+3)N$, *the Neaderhouser condition (2.2) holds, the bandwidth conditions (2.9) and (2.8) hold*

or if

- $\theta > (\tilde{d}+1+2\tilde{k})N$, *the Takahata condition (2.3) holds, the bandwidth conditions (2.10) and (2.8) hold* *then for all compact set D in $R^{\tilde{d}}$*

$$\sup_{x\in D}|f_{\mathbf{n}}(x)-f(x)|=O(\Psi_{\mathbf{n}}) \quad \textit{in probability.} \tag{2.15}$$

If

- $\theta>(\tilde{d}+5)N$, *the Neaderhouser condition (2.2) holds, the bandwidth conditions (2.13) and (2.8) hold*

or if

- $\theta>(\tilde{d}+3+2\tilde{k})N$, *the Takahata condition (2.3) holds, the bandwidth conditions (2.14) and (2.8) hold*

then

$$\sup_{x\in D}|f_{\mathbf{n}}(x)-f(x)|=O(\Psi_{\mathbf{n}}) \quad \textit{a.s.} \tag{2.16}$$

(ii) Now assume that the exponential mixing rate (2.6) holds. If the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds, and if the bandwidth conditions (2.12) and (2.8) hold then

$$\sup_{x\in D}|f_{\mathbf{n}}(x)-f(x)|=O(\Psi_{\mathbf{n}}) \quad \textit{a.s.} \tag{2.17}$$

Carbon, Tran and Wu (1997) obtained the uniform consistency of kernel density estimators under the assumption that, when $\mathbf{i}\neq\mathbf{j}$, the joint probability density $f_{X_{\mathbf{i}},X_{\mathbf{j}}}(x,y)$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ exists. It is clear that this assumption does not hold when $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ are replaced by $X_{(\mathbf{i})}$ and $X_{(\mathbf{j})}$, since $X_{(\mathbf{i})}$ and $X_{(\mathbf{j})}$ may have a common component, even when $\mathbf{i}\neq\mathbf{j}$. For this reason the proof of Theorem 2.1 cannot be deduced from Carbon, Tran and Wu (1997).

The following assumptions are additional conditions required for the convergence of $r_{\mathbf{n}}(x)$ to $r(x)$.

Assumption 4. $\inf_{x\in D} f(x)>0$.

Assumption 5. Suppose that

$$\sum_{i=1}^{\infty} i^{N-1}\{\kappa(i)\}^{a}<\infty \tag{2.18}$$

for some $0 < a < 1/2$, and for some $s > s_0 := 2 + 2a(1-2a)^{-1}$,

$$E|\varphi(X_{\mathbf{j}})|^s < \infty, \quad \sup_u \int |\varphi(v)|^s f_{X_{\mathbf{j}},X_{(\mathbf{j})}}(u,v)dv < C,$$

where $f_{X_{\mathbf{j}},X_{(\mathbf{j})}}$ denotes the joint density of $(X_{\mathbf{j}}, X_{(\mathbf{j})})$.

It will be shown that if the polynomial mixing rate (2.5) holds for some $\theta > 2N$, or the exponential mixing rate (2.6) holds, then we have (2.18). Note that when the exponential mixing rate (2.6) holds, Assumption 5 is satisfied whenever

$$E|\varphi(X_{\mathbf{j}})|^s < \infty, \quad \sup_u \int |\varphi(v)|^s f_{X_{\mathbf{j}},X_{(\mathbf{j})}}(u,v)dv < C,$$

for some $s > 2$.

Assumption 6. If $t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset$, $\mathbf{j} \notin t(\mathbf{i})$ and $\mathbf{i} \notin t(\mathbf{j})$, the conditional probability density $f_{X_{(\mathbf{i})},X_{(\mathbf{j})}|X_{\mathbf{i}},X_{\mathbf{j}}}$ of $(X_{(\mathbf{i})}, X_{(\mathbf{j})})$ given $(X_{\mathbf{i}}, X_{\mathbf{j}})$ and the conditional density of $X_{(\mathbf{i})}$ given $X_{\mathbf{j}}$ exist and satisfy

$$f_{X_{(\mathbf{i})},X_{(\mathbf{j})}|X_{\mathbf{i}},X_{\mathbf{j}}}(u,v|y,z) \le C \quad \text{and} \quad f_{X_{(\mathbf{i})}|X_{\mathbf{j}}}(y|u) \le C$$

for all u, v, y, z and all i, j.

Assumption 7. For all $x, y \in R^{\tilde{d}}$,

$$|\psi(x) - \psi(y)| \le C\|x - y\|, \quad \text{where} \quad \psi(x) = r(x)f(x).$$

Consider s satisfying Assumption 5. Similar to the θ_i's defined before Theorem 2.1, we introduce coefficients θ_i^* in order to defined constraints on the rate of convergence of the bandwidth. The θ_i^*'s are functions of θ and s because the bandwidth obviously depends on the decreasing rate of the mixing coefficients and on the moment condition given in Assumption 5. More precisely, define

$$\theta_1^* = \frac{s\tilde{d}(\theta + (\tilde{d}+3)N)}{\theta(s-1) - N\{3s + \tilde{d}(s+2)\}}, \quad \theta_2^* = \frac{-s(\theta - (\tilde{d}+1)N)}{\theta(s-1) - N\{3s + \tilde{d}(s+2)\}},$$

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

$$\theta_5^* = \frac{-s(\theta - (\tilde{d}+1)N)}{\theta(s-1) - N\{(2\tilde{k}+1)s + \tilde{d}(s+2)\}}, \quad \theta_6^* = \frac{-\theta - 2N\tilde{d}}{\theta(s-1) - N\{(2\tilde{k}+1)s + \tilde{d}(s+2)\}},$$

$$\theta_7^* = \frac{s\tilde{d}(\theta + (\tilde{d}+3)N)}{\theta(s-1) - N\{5s + \tilde{d}(s+2)\}}, \quad \theta_8^* = \frac{-s(\theta - (\tilde{d}+1)N)}{\theta(s-1) - N\{5s + \tilde{d}(s+2)\}},$$

$$\theta_9^* = \frac{-\theta - 2N\tilde{d} - 2sN}{\theta(s-1) - N\{5s + \tilde{d}(s+2)\}}, \quad \theta_{10}^* = \frac{s\tilde{d}(\theta + (\tilde{d}+3)N)}{\theta(s-1) - N\{(2\tilde{k}+3)s + \tilde{d}(s+2)\}},$$

$$\theta_{11}^* = \frac{-s(\theta - (\tilde{d}+1)N)}{\theta(s-1) - N\{(2\tilde{k}+3)s + \tilde{d}(s+2)\}}, \quad \theta_{12}^* = \frac{-\theta - 2N\tilde{d} - 2sN}{\theta(s-1) - N\{(2\tilde{k}+3)s + \tilde{d}(s+2)\}}.$$

With the notation (2.7), consider the following additional restrictions on the bandwidths sequence:

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_1^*} (\log \hat{\mathbf{n}})^{\theta_2^*} \{g(\mathbf{n})\}^{\theta_3^*} \to \infty, \tag{2.19}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_4^*} (\log \hat{\mathbf{n}})^{\theta_5^*} \{g(\mathbf{n})\}^{\theta_6^*} \to \infty, \tag{2.20}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_7^*} (\log \hat{\mathbf{n}})^{\theta_8^*} g(\mathbf{n})^{\theta_9^*} \to \infty, \tag{2.21}$$

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_{10}^*} (\log \hat{\mathbf{n}})^{\theta_{11}^*} g(\mathbf{n})^{\theta_{12}^*} \to \infty. \tag{2.22}$$

The following theorem deals with the regression estimator. As for the density estimator considered in Theorem 2.1, the constraints on the bandwidth are different for the polynomial mixing rate (2.5) and the exponential mixing rate (2.6), for the Neaderhouser condition (2.2) and the Takahata condition (2.3), and for the weak and strong consistency.

Theorem 2.2. *Assume that Assumptions 1–7 hold.*

(i) Suppose the polynomial mixing rate (2.5) holds. Let s satisfying Assumption 5. If

- $\theta > N\{3s + \tilde{d}(s+2)\}/(s-1)$, *the Neaderhouser condition (2.2) holds, the bandwidth conditions (2.19) and (2.8) hold*

or if

- $\theta > N\{(2\tilde{k}+1)s + \tilde{d}(s+2)\}/(s-1)$, *the Takahata condition (2.3) holds, the bandwidth conditions (2.20) and (2.8) hold*

then

$$\sup_{x \in D} |r_{\mathbf{n}}(x) - r(x)| = O(\Psi_{\mathbf{n}}) \quad \textit{in probability.} \tag{2.23}$$

If

– *$\theta > N\{5s + \tilde{d}(s + 2)\}/(s - 1)$, the Neaderhouser condition (2.2) holds, the bandwidth conditions (2.21) and (2.8) hold*

or if

– *$\theta > N\{(2\tilde{k} + 3)s + \tilde{d}(s+2)\}/(s-1)$, the Takahata condition (2.3) holds, the bandwidth conditions (2.22) and (2.8) hold*

then

$$\sup_{x \in D} |r_{\mathbf{n}}(x) - r(x)| = O(\Psi_{\mathbf{n}}) \quad a.s. \tag{2.24}$$

(ii) Now assume that the exponential mixing rate (2.6) holds. If the Neaderhouser condition (2.2) holds or the Takahata condition (2.3) holds, and if the bandwidth conditions (2.12) and (2.8) hold then

$$\sup_{x \in D} |f_{\mathbf{n}}(x) - f(x)| = O(\Psi_{\mathbf{n}}) \quad a.s. \tag{2.25}$$

Note that, when $s \to \infty$, the conditions given in Theorem 2.2 are marginally close to those given in Theorem 2.1.

3 Preliminary lemmas

This section is a collection of technical lemmas which will be used to prove the consistency results stated in Theorems 2.1 and 2.2.

The first lemma entails that under Assumption 1 $\sup_{x \in R^{\tilde{d}}} K(x) < \tilde{K}$, for some constant $\tilde{K}$.

Lemma 3.1. *Any integrable positive function satisfying a Lipschitz condition is bounded.*

Proof. Let g be a function from $\mathbb{R}^k$ to $[0, \infty)$ such that $\int_{\mathbb{R}^k} g(x)dx < \infty$ and $|g(y) - g(x)| < C\|x - y\|$ for some constant C. We argue by contradiction. If g is not bounded then

$$\forall A > 0, \quad \exists\, y \in \mathbb{R}^k \quad \text{such that} \quad g(y) > A.$$

Let $B_y(\epsilon) = \{x \in \mathbb{R}^k : \|x - y\| < \epsilon\}$ be the sphere of center y and radius $\epsilon > 0$. We have

$$\forall x \in B_y(\epsilon), \quad g(x) \geq g(y) - |g(y) - g(x)| > A - C\epsilon.$$

Thus

$$\int_{\mathbb{R}^k} g(x)dx \geq \int_{B_y(\epsilon)} g(x)dx \geq (A - C\epsilon)\lambda_k \{B_y(\epsilon)\},$$

where λ_k denotes the Lebesgue measure on $\mathbb{R}^k$. When $A \to +\infty$, the right hand side of the inequality tends to infinity, which is in contradiction with g integrable. □

Lemma 3.2. *Assume that Assumptions 1-2 hold. If $b_{\mathbf{n}} \to 0$ quickly enough so that (2.8) holds, then*

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

Proof. A simple computation shows that (2.8) implies that $b_{\mathbf{n}} = o(\Psi_{\mathbf{n}})$. By Assumptions 1 and 2,

$$|Ef_{\mathbf{n}}(x) - f(x)| = \left| (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \sum_{\mathbf{j}\in I_{\mathbf{n}}} \int_{R^{\tilde{d}}} K\{(x-y)/b_{\mathbf{n}}\} f(y)dy - f(x) \right|$$

$$= \left| \int_{R^{\tilde{d}}} K(z)\{f(x - b_{\mathbf{n}}z) - f(x)\} dz \right| \leq Cb_{\mathbf{n}} \int \|z\| K(z)dz \leq Cb_{\mathbf{n}} = o(\Psi_{\mathbf{n}}).$$

□

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

Lemma 3.3. *(i) Suppose (2.1) holds. Denote by $\mathcal{L}_r(\mathcal{F})$ the class of $\mathcal{F}$-measurable r.v.'s X satisfying $\|X\|_r = (E|X|^r)^{1/r} < \infty$. Suppose $X \in \mathcal{L}_r\{\mathcal{B}(S)\}$ and $Y \in \mathcal{L}_s\{\mathcal{B}(S')\}$. Assume also that $1 \leq r, s, t < \infty$ and $r^{-1} + s^{-1} + t^{-1} = 1$. Then*

(3.1)

$$|cov(X,Y)| \leq C\|X\|_r \|Y\|_s \left[h\left\{ Card(S), Card(S')\right\} \kappa \left\{ dist(S,S')\right\}\right]^{1/t}.$$

(ii) For r.v.'s bounded with probability 1, the right hand side of (3.1) can be replaced by

$$C\, h\left\{ Card(S), Card(S')\right\} \kappa\left\{ dist(S,S')\right\}.$$

The following lemma can be found in Carbon, Tran and Wu (1997).

Lemma 3.4. *If the polynomial mixing rate (2.5) holds for some* $\theta > 2N$*, or if the exponential mixing rate (2.6) holds, then*

$$\sum_{i=1}^{\infty} i^{N-1} \{\kappa(i)\}^a < \infty \tag{3.2}$$

for some $0 < a < 1/2$.

Denote by $K_{\mathbf{n}}(x)$ the averaging kernel $K_{\mathbf{n}}(x) = (1/b_{\mathbf{n}}^{\tilde{d}})K(x/b_{\mathbf{n}})$. Then the kernel density estimator writes

$$f_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{j} \in I_{\mathbf{n}}} K_{\mathbf{n}}(x - X_{(\mathbf{j})})$$

and the centered estimator $f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)$ is the average of the random variables

$$\Delta_{\mathbf{i}}(x) = K_{\mathbf{n}}(x - X_{(\mathbf{i})}) - EK_{\mathbf{n}}(x - X_{(\mathbf{i})}). \tag{3.3}$$

The variance of $f_{\mathbf{n}}(x)$ is the sum of

$$I_{\mathbf{n}}(x) = \sum_{(\mathbf{i},\mathbf{j}) \in S} |E\Delta_{\mathbf{i}}(x)\Delta_{\mathbf{j}}(x)| \quad \text{and} \quad R_{\mathbf{n}}(x) = \sum_{(\mathbf{i},\mathbf{j}) \in S^*} |E\Delta_{\mathbf{i}}(x)\Delta_{\mathbf{j}}(x)|$$

where $S = \{(\mathbf{i},\mathbf{j}) : \mathbf{i} \in I_{\mathbf{n}}, \mathbf{j} \in I_{\mathbf{n}}, t(\mathbf{i}) \cap t(\mathbf{j}) \neq \emptyset\}$ and $S^* = \{(\mathbf{i},\mathbf{j}) : \mathbf{i} \in I_{\mathbf{n}}, \mathbf{j} \in I_{\mathbf{n}}, t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset\}$.

Denote by $\tilde{r} = \max\{\|\mathbf{i} - \mathbf{j}\| : \ \mathbf{i}, \mathbf{j} \in t(\mathbf{i})\}$ the diameter of $t(\mathbf{i})$.

Lemma 3.5. *Assume that Assumptions 1–3 hold. If the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds, and the polynomial mixing rate (2.5) holds for some* $\theta > 2N$ *or the exponential mixing rate (2.6) holds, then*

$$\limsup \hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} \big(I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x)\big) < C,$$

where C *is a constant independent of* x.

Proof. Since Card $I_{\mathbf{n}}(x) = \hat{\mathbf{n}}$ and

$$\text{Card } \{\mathbf{j} : \ t(\mathbf{i}) \cap t(\mathbf{j}) \neq \emptyset\} \leq \sum_{\mathbf{k} \in t(\mathbf{i})} \text{Card } \{\mathbf{j} : \ \mathbf{k} \in t(\mathbf{j}) \neq \emptyset\} \leq \ell^2,$$

we have Card $S \leq \ell^2 \hat{\mathbf{n}}$ and $I_{\mathbf{n}}(x) \leq \ell^2 \hat{\mathbf{n}} E \Delta_{\mathbf{j}}^2(x)$. Thus we have

$$\hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} I_{\mathbf{n}}(x) \leq \ell^2 \int_{R^{\tilde{d}}} (1/b_{\mathbf{n}}^{\tilde{d}}) K^2 \left\{(x-u)/b_{\mathbf{n}}\right\} f(u) du$$

$$= \ell^2 \int_{R^{\tilde{d}}} K^2(v) f(x - b_{\mathbf{n}} v) dv.$$

Under Assumptions 1 and 2, we have

$$\left| \int_{R^{\tilde{d}}} K^2(v) \left\{ f(x - b_{\mathbf{n}} v) - f(x) \right\} dv \right| \tag{3.4}$$

$$\leq \int_{R^{\tilde{d}}} K^2(v) \left| f(x - b_{\mathbf{n}} v) - f(x) \right| dv \leq C b_{\mathbf{n}} = o(1).$$

Thus $\lim_{\mathbf{n}\to\infty} \int K^2(v) f(x - b_{\mathbf{n}} v) dv = f(x) \int K^2(v) dv$ and

$$\limsup_{\mathbf{n}\to\infty} \hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} I_{\mathbf{n}}(x) \leq \ell^2 f(x) \int K^2(u) du < \infty. \tag{3.5}$$

Lemma 3.1 and Assumption 2 show that f is bounded. Therefore the limit in (3.5) is uniformly bounded in x.

Recall that (3.2) holds. Define $c_{\mathbf{n}} = b_{\mathbf{n}}^{-\tilde{d}(1-\gamma)/\nu}$, where $\nu = -N - \epsilon + (1-\gamma) N a^{-1}$, with γ and ϵ being small positive numbers such that $a^{-1} - (N+\epsilon)(1-\gamma)^{-1} N^{-1} > 1$. This can be done since $0 < a < 1/2$. Therefore $\nu > N(1-\gamma)$. Define

$$S_1 = \{(\mathbf{i}, \mathbf{j}) : \|\mathbf{i} - \mathbf{j}\| \leq c_{\mathbf{n}} \text{ and } t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset\}, \tag{3.6}$$

$$S_2 = \{(\mathbf{i}, \mathbf{j}) : \|\mathbf{i} - \mathbf{j}\| > c_{\mathbf{n}} \text{ and } t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset\}.$$

We have

$$R_{\mathbf{n}}(x) \leq J_1 + J_2, \tag{3.7}$$

where

$$J_1 = \sum\sum_{(\mathbf{i},\mathbf{j})\in S_1} \int\int K_{\mathbf{n}}(x-u) K_{\mathbf{n}}(x-v) |f_{X_{(\mathbf{i})}, X_{(\mathbf{j})}}(u,v) - f(u) f(v)| du dv,$$

$$J_2 = \sum\sum_{(\mathbf{i},\mathbf{j})\in S_2} |\mathrm{cov}\{K_{\mathbf{n}}(x - X_{(\mathbf{i})}), K_{\mathbf{n}}(x - X_{(\mathbf{j})})\}|.$$

By Assumptions 1 and 3,

$$J_1 \le C\left(\int |K(v)|dv\right)^2 \sum\sum_{S_1} 1 \tag{3.8}$$

$$\le C\hat{\mathbf{n}}c_{\mathbf{n}}^N = C\hat{\mathbf{n}}b_{\mathbf{n}}^{-N\tilde{d}(1-\gamma)/\nu} = o(\hat{\mathbf{n}}b_{\mathbf{n}}^{-\tilde{d}})$$

since $\nu > N(1-\gamma)$ and $b_{\mathbf{n}} \to 0$.

Denote

$$\delta = 2(1-\gamma)/\gamma. \tag{3.9}$$

Then $\gamma = 2/(2+\delta)$ and $\delta/(2+\delta) = 1-\gamma$. Applying Lemma 3.3 with $r = s = 2+\delta$, $t = (2+\delta)/\delta$,

$$|\text{cov}\{K_{\mathbf{n}}(x - X_{(\mathbf{i})}), K_{\mathbf{n}}(x - X_{(\mathbf{j})})\}| \tag{3.10}$$

$$\le C\left(\int |K_{\mathbf{n}}(x-u)|^{2+\delta} f(u)du\right)^{\gamma} \{h(\ell,\ell)\kappa\left(\text{dist }\{t(\mathbf{j}), t(\mathbf{i})\}\right)\}^{1-\gamma}$$

$$\le C\left(\int |K_{\mathbf{n}}(x-u)|^{2+\delta} f(u)du\right)^{\gamma} \{\kappa\left(\max\{\|\mathbf{i}-\mathbf{j}\| - \tilde{r}, 0\}\right)\}^{1-\gamma}.$$

Employing (3.10) and using the convention $\kappa(u) = \kappa(0)$ for $u < 0$,

$$J_2 \le C\left(\int |K_{\mathbf{n}}(x-u)|^{2+\delta} f(u)du\right)^{\gamma} \sum\sum_{S_2} \{\kappa(\|\mathbf{j}-\mathbf{i}\| - \tilde{r})\}^{1-\gamma}. \tag{3.11}$$

Putting $\mathbf{k} = \mathbf{j} - \mathbf{i}$ we have

$$\sum\sum_{S_2} \{\kappa(\|\mathbf{j}-\mathbf{i}\| - \tilde{r})\}^{1-\gamma} \le C\hat{\mathbf{n}} \sum_{\|\mathbf{k}\| > c_{\mathbf{n}}} \{\kappa(\|\mathbf{k}\| - \tilde{r})\}^{1-\gamma}. \tag{3.12}$$

Combining (3.11), (3.12)

$$\hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} J_2 \le C b_{\mathbf{n}}^{-\tilde{d}(1-\gamma)} \left(\int (1/b_{\mathbf{n}}^{\tilde{d}}) |K\left\{(x-u)/b_{\mathbf{n}}\right\}|^{2+\delta} f(u)du\right)^{\gamma} \tag{3.13}$$

$$\times \sum_{\|\mathbf{i}\| > c_{\mathbf{n}}} \{\kappa(\|\mathbf{i}\| - \tilde{r})\}^{1-\gamma}.$$

With the arguments used to prove (3.4) it can be shown that

$$\lim_{\mathbf{n}\to\infty} \int (1/b_{\mathbf{n}}^{\tilde{d}}) |K\left\{(x-u)/b_{\mathbf{n}}\right\}|^{2+\delta} f(u)du = f(x) \int |K(v)|^{2+\delta} dv < \infty. \tag{3.14}$$

Since $\sum_{i=1}^{\infty} i^{N-1} \{\kappa(i)\}^a < \infty$ and $\kappa(\cdot)$ is a nonincreasing function, we have $i^{N-1} \{\kappa(i)\}^a = o(1/i)$ as $i \to \infty$. Therefore $\kappa(x) = o(x^{-N/a})$ as $x \to \infty$, and $\|\mathbf{i}\|^\nu \{\kappa(\|\mathbf{i}\|)\}^{1-\gamma} = o(\|\mathbf{i}\|^{-N-\epsilon})$, since $\nu = -N - \epsilon + N(1-\gamma)a^{-1}$. Thus

$$\sum_{\|\mathbf{i}\|>0} \|\mathbf{i}\|^\nu \{\kappa(\|\mathbf{i}\|)\}^{1-\gamma} < \infty. \tag{3.15}$$

Also note that $c_\mathbf{n} - \tilde{r} > c_\mathbf{n}/2$ for sufficiently large $\hat{\mathbf{n}}$, since $c_\mathbf{n} \to \infty$. Using (3.13), (3.14), (3.15) and note that $b_\mathbf{n}^{-\tilde{d}(1-\gamma)} c_\mathbf{n}^{-\nu} = 1$, we get

$$\begin{aligned}\limsup \hat{\mathbf{n}}^{-1} b_\mathbf{n}^{\tilde{d}} J_2 &\le C \limsup b_\mathbf{n}^{-\tilde{d}(1-\gamma)} \sum_{\|\mathbf{i}\|>c_\mathbf{n}-\tilde{r}} \{\kappa(\|\mathbf{i}\|)\}^{1-\gamma} \\ &\le C \limsup b_\mathbf{n}^{-\tilde{d}(1-\gamma)} c_\mathbf{n}^{-\nu} \sum_{\|\mathbf{i}\|>c_\mathbf{n}/2} \|\mathbf{i}\|^\nu \{\kappa(\|\mathbf{i}\|)\}^{1-\gamma} \\ &\le C \limsup \sum_{\|\mathbf{i}\|>c_\mathbf{n}/2} \|\mathbf{i}\|^\nu \{\kappa(\|\mathbf{i}\|)\}^{1-\gamma},\end{aligned} \tag{3.16}$$

which is equal to zero since $c_\mathbf{n} \to \infty$. The conclusion follows from (3.5), (3.7), (3.8) and (3.16). □

The proof of the following lemma, which is based on Rio (1995), can be found in Carbon, Tran and Wu (1997, Lemma 4.5).

Lemma 3.6. *Suppose S_1, S_2, ..., S_r be sets containing m sites each with dist$(S_i, S_j) \ge \delta$ for all $i \ne j$ where $1 \le i \le r$ and $1 \le j \le r$. Suppose Y_1, Y_2, ..., Y_r is a sequence of real-valued r.v.'s measurable with respect to $\mathcal{B}(S_1)$, $\mathcal{B}(S_2)$, ..., $\mathcal{B}(S_r)$ respectively and Y_i takes values in $[a, b]$. Then there exists a sequence of independent r.v.'s Y_1^*, Y_2^*, ..., Y_r^* independent of Y_1, Y_2, ..., Y_r such that Y_i^* has the same distribution as Y_i and satisfies*

$$\sum_{i=1}^r E|Y_i - Y_i^*| \le 2r(b-a)h\{(r-1)m, m\}\kappa(\delta). \tag{3.17}$$

Choose

$$\tilde{\ell} = b_\mathbf{n}^{(\tilde{d}+1)} \Psi_\mathbf{n}. \tag{3.18}$$

Since D is compact, it can be covered with v cubes I_k having sides of length $\tilde{\ell}$ and center at x_k. Let C be a constant greater than the Lebesgue measure of $\cup_k I_k$. Taking disjoined cubes I_k, we have $v\tilde{\ell}^{\tilde{d}} \le C$. Thus

$$v \le C\big(b_\mathbf{n}^{(\tilde{d}+1)} \Psi_\mathbf{n}\big)^{-\tilde{d}}. \tag{3.19}$$

Define

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

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

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

Then

$$\sup_{x\in D} |f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)| \le Q_{1\mathbf{n}} + Q_{2\mathbf{n}} + Q_{3\mathbf{n}}. \tag{3.20}$$

In view of Lemma 3.2, the consistency results (2.15), (2.16) and (2.17) will be obtained by showing that, for $i = 1, 2$ and 3, $Q_{i\mathbf{n}} = O(\Psi_{\mathbf{n}})$ in probability and a.s. The following lemma deals with $Q_{1\mathbf{n}}$ and $Q_{2\mathbf{n}}$.

Lemma 3.7. *Assume that Assumption 1 holds. Then $Q_{1\mathbf{n}} = O(\Psi_{\mathbf{n}})$ a.s. and $Q_{2\mathbf{n}} = O(\Psi_{\mathbf{n}})$.*

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

$$|f_{\mathbf{n}}(x) - f_{\mathbf{n}}(x_k)| \le Cb_{\mathbf{n}}^{-(\tilde{d}+1)}\|x - x_k\| \le Cb_{\mathbf{n}}^{-(\tilde{d}+1)}\tilde{\ell} = O(\Psi_{\mathbf{n}}) \quad \text{a.s.}$$

The lemma easily follows. □

To show that

$$Q_{3\mathbf{n}} = O(\Psi_{\mathbf{n}}) \quad \text{in probability (and a.s.)} \tag{3.21}$$

we need to introduce additional notations. Define

$$S_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in I_{\mathbf{n}}} \hat{\mathbf{n}}^{-1}\Delta_{\mathbf{i}}(x), \tag{3.22}$$

so that

$$S_{\mathbf{n}}(x) = f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x). \tag{3.23}$$

Showing (3.21) is equivalent to showing that

$$\max_{1\le k\le v} |S_{\mathbf{n}}(x_k)| = O(\Psi_{\mathbf{n}}) \quad \text{in probability (and a.s.).} \tag{3.24}$$

Now we will employ the blocking technique used in Carbon, Tran and Wu (1997). This technique is reminiscient of the blocking scheme in Tran

(1990) and Politis and Romano (1993). We will split the sum $S_{\mathbf{n}}(x)$ into 2^N sums of variables $Y_1, \dots, Y_r$ satisfying the assumptions of Lemma 3.6.

Without loss of generality assume that $n_i = 2pq_i$ for $1 \le i \le N$. 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

$$U(1, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k = 2j_k p+1 \\ k=1,\dots,N}}^{(2j_k+1)p} \hat{\mathbf{n}}^{-1} \Delta_{\mathbf{i}}(x), \tag{3.25}$$

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

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

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

and so on. The last two terms are

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

and

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

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

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

We then have

$$S_{\mathbf{n}}(x) = \sum_{i=1}^{2^N} T(\mathbf{n}, i, x). \tag{3.26}$$

To establish (3.24) by (3.26) it is sufficient to show that

$$\max_{1\le k\le v} |T(\mathbf{n}, i, x_k)| = O(\Psi_{\mathbf{n}}) \quad \text{in probability (and a.s.)} \tag{3.27}$$

for each $1 \le i \le 2^N$. Without loss of generality we will show (3.27) for $i = 1$. Now, $T(\mathbf{n}, 1, x)$ is the sum of

$$r = q_1 \times q_2 \times \cdots \times q_N \tag{3.28}$$

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 $X_{t_1(\mathbf{i})}, \ldots, X_{t_\ell(\mathbf{i})}$ with $t_1(\mathbf{i}), \ldots, t_\ell(\mathbf{i})$ belonging to the set of sites

$$\bigcup_{k'=1}^{\ell} \{t_{k'}(\mathbf{i}) : 2j_k p + 1 \le i_k \le (2j_k + 1)p, \quad k = 1, ..., N\}.$$

These sets of sites are separated by a distance of at least $p - \tilde{r}$, with $\tilde{r} = \max\{\|\mathbf{i} - \mathbf{j}\| : \ \mathbf{i}, \mathbf{j} \in t(\mathbf{i})\}$. 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 $Y_1, Y_2, ..., Y_r$ and $S_1, S_2, ..., S_r$. Approximate $Y_1, Y_2, ..., Y_r$ by the r.v.'s $Y_1^*, Y_2^*, ..., Y_r^*$ as was done in Lemma 3.6. In view of (3.3) we have $|\Delta_{\mathbf{i}}| < \tilde{K}/b_{\mathbf{n}}^{\tilde{d}}$, and in view of (3.25) we have

$$|Y_i| \le \max_{\mathbf{j}\in I_{\mathbf{n}}} |U(1, \mathbf{n}, \mathbf{j}, x)| < p^N (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \tilde{K}. \tag{3.29}$$

Denote $\epsilon_{\mathbf{n}} = \eta \Psi_{\mathbf{n}}$, where η is a constant to be chosen later. Set

$$\lambda_{\mathbf{n}} = (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}} \log \hat{\mathbf{n}})^{1/2}, \tag{3.30}$$

$$p = \left[\left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}}}{4\lambda_{\mathbf{n}} \tilde{K}} \right)^{1/N} \right] \sim \left(4\tilde{K}\right)^{-1/N} \left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}}}{\log \hat{\mathbf{n}}} \right)^{\frac{1}{2N}} \sim C \Psi_{\mathbf{n}}^{-\frac{1}{4N}}. \tag{3.31}$$

Note that the total number of sites involved in the Y_i's is less that $\tilde{\mathbf{n}} := \mathrm{Card}(\mathcal{I}_{\mathbf{n}})$. Define $\beta_{\mathbf{n}} = b_{\mathbf{n}}^{-\tilde{d}} h(\tilde{\mathbf{n}}, (p + \tilde{r})^N) \kappa(p - \tilde{r}) \Psi_{\mathbf{n}}^{-1}$.

We have seen that to obtain the consistency results of Theorem 2.1 it is sufficient to show (3.27). This will be done in the next section by means of the following lemma.

Lemma 3.8. *Assume that Assumptions 1–3 hold, that the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds, and the polynomial* *mixing rate (2.5) holds for some $\theta > 2N$ or the exponential mixing rate (2.6) holds. For any arbitrarily large positive constant A, there exist positive constants C and η such that*

$$P\left(\max_{1\le k\le v} |T(\mathbf{n},1,x_k)| > \epsilon_{\mathbf{n}}\right) \le Cv\big(\hat{\mathbf{n}}^{-A} + \beta_{\mathbf{n}}\big).$$

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

$$P\left(|T(\mathbf{n},1,x)| > \epsilon_{\mathbf{n}}\right) \le P\left(|\sum_{i=1}^{r} Y_i^*| > \epsilon_{\mathbf{n}}/2\right) + P\left(\sum_{i=1}^{r} |Y_i - Y_i^*| > \epsilon_{\mathbf{n}}/2\right). \tag{3.32}$$

We now proceed to obtain bounds for the two terms on the right hand side of (3.32).

Note that the Y_i's are measurable with respect to sets of sites which are separated by a distance of at least $p-\tilde{r}$, and contain less than $(p+\tilde{r})^N$ sites. By the Markov inequality and using (3.17) and (3.29),

$$P\left(\sum_{i=1}^{r} |Y_i - Y_i^*| > \epsilon_{\mathbf{n}}\right) \le Crp^N(\hat{\mathbf{n}}b_{\mathbf{n}}^{\tilde{d}})^{-1}h(\tilde{\mathbf{n}},(p+\tilde{r})^N)\kappa(p-\tilde{r})\epsilon_{\mathbf{n}}^{-1} \le C\beta_{\mathbf{n}}. \tag{3.33}$$

A simple computation yields, $\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} = \eta\log\hat{\mathbf{n}}$, and by Lemma 3.5

$$\lambda_{\mathbf{n}}^2 \sum_{i=0}^{r} E(Y_i^*)^2 \le C\hat{\mathbf{n}}^{-1}b_{\mathbf{n}}^{\tilde{d}}\big(I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x)\big)\log\hat{\mathbf{n}} < C\log\hat{\mathbf{n}}.$$

Using (3.29), we have $|\lambda_{\mathbf{n}}Y_i^*| < 1/2$. Applying Berstein's inequality,

$$P\left(|\sum_{i=0}^{r} Y_i^*| > \epsilon_{\mathbf{n}}\right) \le 2\exp\left(-\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} + \lambda_{\mathbf{n}}^2\sum_{i=0}^{r} E(Y_i^*)^2\right) \le 2\exp\{(-\eta + C)\log\hat{\mathbf{n}}\} \le \hat{\mathbf{n}}^{-A}, \tag{3.34}$$

for sufficiently large $\hat{\mathbf{n}}$, when η is a given large number, depending on A.

Combining (3.32), (3.33) and (3.34),

$$P\left(\sup_{x\in D} |T(\mathbf{n},1,x)| > \epsilon_{\mathbf{n}}\right) \le Cv\left(\hat{\mathbf{n}}^{-A} + \beta_{\mathbf{n}}\right).$$

□

4 Proof of Theorem 2.1

Since

$$\sup_{x\in D} |f_{\mathbf{n}}(x) - f(x)| \le \sup_{x\in D} |f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)| + \sup_{x\in D} |Ef_{\mathbf{n}}(x) - f(x)|,$$

Lemma 3.2 shows that it suffices to show the consistency of $\sup_{x\in D} |f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x)|$. From (3.20) and Lemma 3.7, it suffices to show the consistency of $Q_{3\mathbf{n}}$.

(i) Assume that the polynomial rate of convergence (2.5) holds and suppose that (2.9) or (2.10) holds. In view of (3.27) and Lemma 3.8, we will show that $v\hat{\mathbf{n}}^{-a} \to 0$ and $v\beta_{\mathbf{n}} \to 0$ in order to establish (2.15).

Using (3.19) and (2.4),

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

First note that (2.11) is implied by (2.9) with $\theta > N(\tilde{d}+3)$, and is also implied by (2.10) with $\theta > N(\tilde{d}+1+2\tilde{k})$. Relation (2.11) implies $\hat{\mathbf{n}} \ge Cb_{\mathbf{n}}^{-\tilde{d}}$. Thus $b_{\mathbf{n}}^{-\tilde{d}\{(\tilde{d}/2)+1\}} \le C\hat{\mathbf{n}}^{(\tilde{d}/2)+1}$ and $v \le C\hat{\mathbf{n}}^{\tilde{d}+1}(\log \hat{\mathbf{n}})^{-\tilde{d}/2}$. Therefore

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

which goes to 0 if $a > \tilde{d}+1$.

Using (3.19),

$$v\beta_{\mathbf{n}} \le Cb_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+2)} h(\tilde{\mathbf{n}}, (p+\tilde{r})^N)(p-\tilde{r})^{-\theta}\Psi_{\mathbf{n}}^{-(\tilde{d}+1)}. \tag{4.1}$$

Since $p \to \infty$ as $\mathbf{n} \to \infty$, and $\tilde{r}$ is fixed, it is clear that

$$p+\tilde{r} \sim p \quad \text{and} \quad \tilde{\mathbf{n}} \sim \hat{\mathbf{n}}. \tag{4.2}$$

In view of (2.4), (3.31), (4.2) and (4.1), the Neaderhouser condition (2.2) implies

$$\begin{aligned} v\beta_{\mathbf{n}} &\le Cb_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+2)}\hat{\mathbf{n}}\left(\frac{\hat{\mathbf{n}}b_{\mathbf{n}}^{\tilde{d}}}{\log \hat{\mathbf{n}}}\right)^{\frac{-\theta}{2N}} \left(\log \hat{\mathbf{n}}(\hat{\mathbf{n}}b_{\mathbf{n}}^{\tilde{d}})^{-1}\right)^{-\frac{\tilde{d}+1}{2}} \\ &= C\hat{\mathbf{n}}^{1-\frac{\theta}{2N}+\frac{\tilde{d}+1}{2}} b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+2)-\frac{\tilde{d}\theta}{2N}+\frac{\tilde{d}(\tilde{d}+1)}{2}} (\log \hat{\mathbf{n}})^{\frac{\theta}{2N}-\frac{\tilde{d}+1}{2}} \\ &= C\left(\hat{\mathbf{n}}b_{\mathbf{n}}^{\theta_1}(\log \hat{\mathbf{n}})^{\theta_2}\right)^{\frac{-\theta+N(\tilde{d}+3)}{2N}}. \end{aligned} \tag{4.3}$$

When $\theta > N(\tilde{d}+3)$ (2.9) is then equivalent to $v\beta_{\mathbf{n}} \to 0$. Analogously, the Takahata condition (2.3) implies

$$(4.4) \qquad v\beta_{\mathbf{n}} \le C b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+2)} \hat{\mathbf{n}}^{\tilde{k}} \left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}}}{\log \hat{\mathbf{n}}} \right)^{\frac{-\theta}{2N}} \left(\log \hat{\mathbf{n}} (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \right)^{-\frac{\tilde{d}+1}{2}}$$

$$= C \hat{\mathbf{n}}^{\tilde{k} - \frac{\theta}{2N} + \frac{\tilde{d}+1}{2}} b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+2) - \frac{\tilde{d}\theta}{2N} + \frac{\tilde{d}(\tilde{d}+1)}{2}} (\log \hat{\mathbf{n}})^{\frac{\theta}{2N} - \frac{\tilde{d}+1}{2}}$$

$$= C \left(\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_3} (\log \hat{\mathbf{n}})^{\theta_4} \right)^{\frac{-\theta + N(\tilde{d}+1+2\tilde{k})}{2N}}.$$

When $\theta > N(\tilde{d}+1+2\tilde{k})$ (2.10) is then equivalent to $v\beta_{\mathbf{n}} \to 0$. The proof of (2.15) is completed.

By the Fubini theorem it can be seen that $\sum \frac{1}{\hat{\mathbf{n}} g(\mathbf{n})} < \infty$, where the summation is over all $\mathbf{n}$ in Z^N. From (4.3) we see that the polynomial mixing rate (2.5) with $\theta > (\tilde{d}+5)N$, the Neaderhouser condition (2.2) and the bandwidth condition (2.13) imply $v\beta_{\mathbf{n}} \hat{\mathbf{n}} g(\mathbf{n}) \to 0$, which entails $\sum_{\mathbf{n} \in Z^N} v\beta_{\mathbf{n}} < \infty$. From (4.4), the same result is obtained under the polynomial mixing rate (2.5) with $\theta > (\tilde{d}+3+2\tilde{k})N$, the Takahata condition (2.3) and the bandwidth condition (2.14). Thus (2.16) follows by the Borel-Cantelli Lemma.

(ii) Now assume that the exponential rate of convergence (2.6) holds. Condition (2.12) implies that

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

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

$$p \sim (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}} / \log \hat{\mathbf{n}})^{1/2N} \ge (A/s) \log \hat{\mathbf{n}}.$$

Therefore

$$(4.5) \qquad \kappa(p) \le C \exp(-sp) \le C \exp\{-s(A/s) \log \hat{\mathbf{n}}\} = C \hat{\mathbf{n}}^{-A}.$$

If necessary, the constant C in inequality (4.5) can be increased so that the inequality holds for all $\hat{\mathbf{n}}$. Using (3.19) and (4.2), we have when the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds

$$(4.6) \qquad v\beta_{\mathbf{n}} \le C b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+2)} h(\tilde{\mathbf{n}}, (p+\tilde{r})^N) \hat{\mathbf{n}}^{-A} \Psi_{\mathbf{n}}^{-(\tilde{d}+1)}$$

$$\le C h(\tilde{\mathbf{n}}, (p+\tilde{r})^N) \hat{\mathbf{n}}^{-A+\frac{\tilde{d}+1}{2}} b_{\mathbf{n}}^{-\frac{\tilde{d}(\tilde{d}+3)}{2}}$$

$$\le C \hat{\mathbf{n}}^{-2} (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-\frac{\tilde{d}+3}{2}},$$

for sufficiently large A. Using (2.12), it is easy to show that $\sum_{\mathbf{n} \in Z^N} v\beta_{\mathbf{n}} < \infty$. Thus (2.17) follows by the Borel-Cantelli Lemma.

5 Additional lemmas and proof of Theorem 2.2

Choose $\epsilon > 0$, $s > 0$ and denote $\tilde{\ell}(\mathbf{n}, \epsilon) \equiv \hat{\mathbf{n}} g(\mathbf{n})$, where $g(\mathbf{n})$ is defined by (2.7), and consider $B_{\mathbf{n}} = \left(\tilde{\ell}(\mathbf{n}, \epsilon)\right)^{1/s}$. Denote by $\psi_{\mathbf{n}}^B$ the truncated estimate

$$\psi_{\mathbf{n}}^B(x) = (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \sum_{\mathbf{j} \in I_{\mathbf{n}}} \varphi(X_{\mathbf{j}}) K\left\{(x - X_{(\mathbf{j})})/b_{\mathbf{n}}\right\} I\{|\varphi(X_{\mathbf{j}})| < B_{\mathbf{n}}\},$$

where I denotes the indicator function.

Lemma 5.1. *If $E|\varphi(X_{\mathbf{n}})|^s < \infty$ then almost surely*

$$\sup_{x \in D} |\psi_{\mathbf{n}}(x) - \psi_{\mathbf{n}}^B(x)| = 0$$

for sufficiently large $\mathbf{n}$.

Proof. By the Markov inequality, $P(|\varphi(X_{\mathbf{n}})| > B_{\mathbf{n}}) \leq B_{\mathbf{n}}^{-s} E|\varphi(X_{\mathbf{n}})|^s$. Thus

$$\sum_{\mathbf{n} \in Z^N} P(|\varphi(X_{\mathbf{n}})| > B_{\mathbf{n}}) \leq C \sum_{\mathbf{n} \in Z^N} \{\tilde{\ell}(\mathbf{n}, \epsilon)\}^{-1} < \infty.$$

The Borel-Cantelli lemma entails that almost surely $|\varphi(X_{\mathbf{n}})| \leq B_{\mathbf{n}}$ for sufficiently large $\mathbf{n}$. Since $B_{\mathbf{n}} \to \infty$ as $\mathbf{n} \to \infty$, almost surely for sufficiently large $\mathbf{n}$ we have $|\varphi(X_{\mathbf{j}})| \leq B_{\mathbf{n}}$ for all $\mathbf{j} \in I_{\mathbf{n}}$. The conclusion follows. $\square$

Define

$$\begin{aligned} \Delta_{\mathbf{i}}^*(x) &= \varphi(X_{\mathbf{i}}) K_{\mathbf{n}}(x - X_{(\mathbf{i})}) I\{|\varphi(X_{\mathbf{i}})| < B_{\mathbf{n}}\} \\ &- E\left\{\varphi(X_{\mathbf{i}}) K\mathbf{n}(x - X_{(\mathbf{i})}) I\{|\varphi(X_{\mathbf{i}})| < B_{\mathbf{n}}\}\right\}, \end{aligned}$$

$$S_1^* = \{(\mathbf{i}, \mathbf{j}) \in I_{\mathbf{n}} \times I_{\mathbf{n}} : t(\mathbf{i}) \cap t(\mathbf{j}) \neq \emptyset \text{ or } \mathbf{i} \in t(\mathbf{j}) \text{ or } \mathbf{j} \in t(\mathbf{i})\},$$

$$S_2^* = \{(\mathbf{i}, \mathbf{j}) \in I_{\mathbf{n}} \times I_{\mathbf{n}} : t(\mathbf{i}) \cap t(\mathbf{j}) = \emptyset \text{ and } \mathbf{i} \notin t(\mathbf{j}) \text{ and } \mathbf{j} \notin t(\mathbf{i})\},$$

$$I_{\mathbf{n}}^*(x) = \sum\sum\nolimits_{(\mathbf{i}, \mathbf{j}) \in S_1^*} |\text{cov}\{\Delta_{\mathbf{i}}^*(x), \Delta_{\mathbf{j}}^*(x)\}|$$

and

$$R_{\mathbf{n}}^*(x) = \sum\sum\nolimits_{(\mathbf{i}, \mathbf{j}) \in S_2^*} |\text{cov}\{\Delta_{\mathbf{i}}^*(x), \Delta_{\mathbf{j}}^*(x)\}|$$

Lemma 5.2. *Assume that Assumptions 1–3 and 5–6 hold. If the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds, and the polynomial mixing rate (2.5) for some $\theta > 2N$ or the exponential mixing rate (2.6) holds, then*

$$\limsup \hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} \left(I_{\mathbf{n}}^*(x) + R_{\mathbf{n}}^*(x)\right) < C,$$

where C is a constant independent of x.

Proof. Arguing as in the proof of Lemma 3.5 and using Assumption 5,

$$\hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} I_{\mathbf{n}}^*(x) \leq C \int_{R^{\tilde{d}} \times \{v: |\varphi(v)| < B_{\mathbf{n}}\}} b_{\mathbf{n}}^{-\tilde{d}} K^2 \left\{(x-u)/b_{\mathbf{n}}\right\} \varphi^2(v) f_{X_{(\mathbf{i})}, X_{\mathbf{i}}}(u,v) du dv$$

$$\leq C \int_{R^{\tilde{d}}} b_{\mathbf{n}}^{-\tilde{d}} K^2 \left\{(x-u)/b_{\mathbf{n}}\right\} \sup_u \left(\int_{R^d} \varphi^2(v) f_{X_{(\mathbf{i})}, X_{\mathbf{i}}}(u,v) dv \right) du$$

$$\leq C \int_{R^{\tilde{d}}} K^2(v) dv.$$

The Cauchy-Schwarz inequality shows,

$$\int\int |\varphi(u)\varphi(v)| f_{X_{\mathbf{i}}, X_{\mathbf{j}}}(u,v) du dv \leq C,$$

since by Assumption 5, $E|\varphi(X_{\mathbf{i}})|^2 < \infty$. By the previous inequality and Assumption 6, for $(\mathbf{i}, \mathbf{j}) \in S_2^*$,

$$\left| E\Delta_{\mathbf{i}}^*(x) \Delta_{\mathbf{j}}^*(x) \right|$$

$$\leq \int \left(b_{\mathbf{n}}^{\tilde{d}}\right)^{-2} \left|\varphi(u) K\left\{(x-y)/b_{\mathbf{n}}\right\} \varphi(v) K\left\{(x-z)/b_{\mathbf{n}}\right\}\right|$$

$$\times f_{X_{(\mathbf{i})}, X_{(\mathbf{j})} | X_{\mathbf{i}}, X_{\mathbf{j}}}(y, z | u, v)\, f_{X_{\mathbf{i}}, X_{\mathbf{j}}}(u,v) dy dz du dv$$

$$\leq C \int\int \left(b_{\mathbf{n}}^{\tilde{d}}\right)^{-2} K\left\{(x-y)/b_{\mathbf{n}}\right\} K\left\{(x-z)/b_{\mathbf{n}}\right\} dy dz \leq C \left(\int K(t) dt \right)^2 \leq C.$$

By a similar argument, we have $|E\Delta_{\mathbf{i}}^*(x)| < C$. Therefore

$$|\mathrm{cov}\{\Delta_{\mathbf{i}}^*(x), \Delta_{\mathbf{j}}^*(x)\}| \leq C, \quad \forall (\mathbf{i}, \mathbf{j}) \in S_2^*. \tag{5.1}$$

Consider $a, \gamma, \epsilon, \delta, c_{\mathbf{n}}$, S_1, S_2 defined in the proof of Lemma 3.5 and define

$$J_1^* = \sum\sum_{(\mathbf{i},\mathbf{j}) \in S_1 \cap S_2^*} |\mathrm{cov}\{\Delta_{\mathbf{i}}^*(x), \Delta_{\mathbf{j}}^*(x)\}|,$$

$$J_2^* = \sum\sum_{(\mathbf{i},\mathbf{j}) \in S_2 \cap S_2^*} |\mathrm{cov}\{\Delta_{\mathbf{i}}^*(x), \Delta_{\mathbf{j}}^*(x)\}|.$$

By (5.1) ,

$$J_1^* \leq C \sum\sum_{S_1} 1 \leq C \hat{\mathbf{n}} c_{\mathbf{n}}^N = o(\hat{\mathbf{n}} b_{\mathbf{n}}^{-\tilde{d}}).$$

Applying Lemma 3.3 as in the proof of Lemma 3.5,

$$
\begin{aligned}
(5.2) \qquad & |\mathrm{cov}\{\Delta_{\mathbf{i}}^*(x), \Delta_{\mathbf{j}}^*(x)\}| \\
&\le C\left(\int |\varphi(v)K_{\mathbf{n}}(x-u)|^{2+\delta} f_{X_{(\mathbf{i})},X_{\mathbf{i}}}(u,v)dudv\right)^{\gamma} \\
&\times \{h(\ell+1,\ell+1)\kappa\,(\text{dist } \{\{t(\mathbf{j}),\mathbf{j}\},\{t(\mathbf{i}),\mathbf{i}\}\})\}^{1-\gamma} \\
&\le C\left(\int |\varphi(v)K_{\mathbf{n}}(x-u)|^{2+\delta} f_{X_{(\mathbf{i})},X_{\mathbf{i}}}(u,v)dudv\right)^{\gamma} \\
&\times \{\kappa\,(\max\{\|\mathbf{i}-\mathbf{j}\| - \tilde{r}^*, 0\})\}^{1-\gamma}.
\end{aligned}
$$

Note that if $2/\gamma > s_0$, we can always found a sufficiently small ϵ such that $a^{-1} - (N+\epsilon)(1-\gamma)^{-1}N^{-1} > 1$. Therefore choose γ such that

$$
(5.3) \qquad s_0 < \frac{2}{\gamma} < s.
$$

By Assumptions 5,

$$
\begin{aligned}
(5.4) \qquad & \sup_u \int |\varphi(v)|^{2+\delta} f_{X_{(\mathbf{i})},X_{\mathbf{i}}}(u,v)dv \\
&= \sup_u \int |\varphi(v)|^{2/\gamma} f_{X_{(\mathbf{i})},X_{\mathbf{i}}}(u,v)dv < \infty.
\end{aligned}
$$

Using (5.4) and Assumptions 6,

$$
\begin{aligned}
(5.5) \qquad & \int |\varphi(v)K_{\mathbf{n}}(x-u)|^{2+\delta} f_{X_{(\mathbf{i})},X_{\mathbf{i}}}(u,v)dudv \\
&\le C\sup_u \int |\varphi(v)|^{2+\delta} f_{X_{(\mathbf{i})},X_{\mathbf{i}}}(u,v)dv \int |K_{\mathbf{n}}(x-u)|^{2+\delta}du \\
&\le Cb_{\mathbf{n}}^{-\tilde{d}(1+\delta)} \int |K(t)|^{2+\delta}dt.
\end{aligned}
$$

Using (5.2) and (5.5)

$$
\begin{aligned}
(5.6) \qquad \hat{\mathbf{n}}^{-1} b_{\mathbf{n}}^{\tilde{d}} J_2^* &\le C b_{\mathbf{n}}^{\tilde{d}(1-(1+\delta)\gamma)} \sum_{\|\mathbf{i}\|>c_{\mathbf{n}}} \{\kappa(\|\mathbf{i}\| - \tilde{r}^*)\}^{1-\gamma} \\
&= C b_{\mathbf{n}}^{-\tilde{d}(1-\gamma)} \sum_{\|\mathbf{i}\|>c_{\mathbf{n}}} \{\kappa(\|\mathbf{i}\| - \tilde{r}^*)\}^{1-\gamma}.
\end{aligned}
$$

The conclusion follows from the arguments used to obtain (3.16). □

Choose

$$\ell^* = b_{\mathbf{n}}^{(\tilde{d}+1)} \Psi_{\mathbf{n}} B_{\mathbf{n}}^{-1}. \tag{5.7}$$

and supposed that the compact set D is covered with v^* cubes I_k^* having sides of length ℓ^* and center at x_k^*. We have

$$\sup_{x\in D} |\psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}^B(x)| \le Q_{1\mathbf{n}}^* + Q_{2\mathbf{n}}^* + Q_{3\mathbf{n}}^* \tag{5.8}$$

where

$$Q_{1\mathbf{n}}^* = \max_{1\le k\le v^*} \sup_{x\in I_k^*} |\psi_{\mathbf{n}}^B(x) - \psi_{\mathbf{n}}^B(x_k^*)|,$$

$$Q_{2\mathbf{n}}^* = \max_{1\le k\le v^*} \sup_{x\in I_k^*} |E\psi_{\mathbf{n}}^B(x_k^*) - E\psi_{\mathbf{n}}^B(x)|,$$

$$Q_{3\mathbf{n}}^* = \max_{1\le k\le v^*} |\psi_{\mathbf{n}}^B(x_k^*) - E\psi_{\mathbf{n}}^B(x_k^*)|.$$

Lemma 5.3. *Assume that Assumption 1 holds. Then* $Q_{1\mathbf{n}}^* = O(\Psi_{\mathbf{n}})$ *a.s. and* $Q_{2\mathbf{n}}^* = O(\Psi_{\mathbf{n}})$.

Proof. By Assumption 1, for all $x \in I_k$,

$$|\psi_{\mathbf{n}}^B(x) - \psi_{\mathbf{n}}^B(x_k)| \le C b_{\mathbf{n}}^{-(\tilde{d}+1)} B_{\mathbf{n}} \|x - x_k\| \le C b_{\mathbf{n}}^{-(\tilde{d}+1)} B_{\mathbf{n}} \ell^* = O(\Psi_{\mathbf{n}}) \quad \text{a.s.}$$

□

Define

$$S_{\mathbf{n}}^*(x) = \sum_{\mathbf{i}\in I_{\mathbf{n}}} \hat{\mathbf{n}}^{-1} \Delta_{\mathbf{i}}^*(x) = \psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}^B(x).$$

We have

$$Q_{3\mathbf{n}}^* = \max_{1\le k\le v^*} |S_{\mathbf{n}}^*(x_k^*)|. \tag{5.9}$$

Define also $U^*(i, \mathbf{n}, \mathbf{j}, x)$ and $T^*(\mathbf{n}, i, x)$ to be the same as $U(i, \mathbf{n}, \mathbf{j}, x)$ and $T(\mathbf{n}, i, x)$ in Section 3 except with Δ_j replaced by Δ_j^*. Using (5.9), arguing that $S_{\mathbf{n}}^*(x_k^*)$ is a finite sum of the $T^*(\mathbf{n}, i, x_k^*)$'s, and that the random field is stationary,

$$Q_{3\mathbf{n}}^* = O(\Psi_{\mathbf{n}}) \quad \text{if and only if} \quad |T^*(\mathbf{n}, 1, x_k^*)| = O(\Psi_{\mathbf{n}}). \tag{5.10}$$

Now, $T^*(\mathbf{n}, 1, x)$ is the sum of $r = q_1 \times q_2 \times \cdots \times q_N$ of the $U^*(1, \mathbf{n}, \mathbf{j}, x)$'s. Denote $t_0(\mathbf{i}) = \mathbf{i}$. Note that $U^*(1, \mathbf{n}, \mathbf{j}, x)$ is measurable with the σ-field generated by $X_{t_0(\mathbf{i})}, X_{t_1(\mathbf{i})}, \dots, X_{t_\ell(\mathbf{i})}$ with $t_0(\mathbf{i}), t_1(\mathbf{i}), \dots, t_\ell(\mathbf{i})$ belonging to the set of sites

$$\bigcup_{k'=0}^{\ell} \{t_{k'}(\mathbf{i}) : 2j_k p + 1 \le i_k \le (2j_k+1)p, \quad k = 1, ..., N\}.$$

These sets of sites are separated by a distance of at least $p - \tilde{r}^*$. 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 $Y_1, Y_2, ..., Y_r$ and $S_1, S_2, ..., S_r$. Approximate $Y_1, Y_2, ..., Y_r$ by the r.v.'s $Y_1^*, Y_2^*, ..., Y_r^*$ as was done in Lemma 3.6. As in Section 3, denote $\epsilon_{\mathbf{n}} = \eta \Psi_{\mathbf{n}}$. Set

$$p = p^* = \left[\left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}}}{4 B_{\mathbf{n}} \lambda_{\mathbf{n}} \tilde{K}} \right)^{1/N} \right] \sim \left(4\tilde{K}\right)^{-1/N} B_{\mathbf{n}}^{-\frac{1}{N}} \Psi_{\mathbf{n}}^{-\frac{1}{4N}}, \tag{5.11}$$

where $\lambda_{\mathbf{n}}$ is defined by (3.30). Define $\beta_{\mathbf{n}}^* = b_{\mathbf{n}}^{-\tilde{d}} h(\tilde{\mathbf{n}}^*, (p+\tilde{r}^*)^N) \kappa(p - \tilde{r}^*) \Psi_{\mathbf{n}}^{-1}$.

Lemma 5.4. *Assume that Assumptions 1–3 and 5–6 hold, that the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds, and the polynomial mixing rate (2.5) for some $\theta > 2N$ or the exponential mixing rate (2.6) holds. For any arbitrarily large positive constant A, there exist two positive constants C and η such that*

$$P\left(\max_{1 \le k \le v^*} |T^*(\mathbf{n}, 1, x_k^*)| > \epsilon_{\mathbf{n}} \right) \le C v^* \left(\hat{\mathbf{n}}^{-A} + \beta_{\mathbf{n}}^* \right).$$

Proof. By already given arguments

$$|Y_i| < p^N B_{\mathbf{n}} (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \tilde{K}. \tag{5.12}$$

We have $\hat{\mathbf{n}} = \prod_{i=1}^N 2pq_i = 2^N p^N r$. The sets of sites with respect to which the Y_i's are measurable are separated by a distance of at least $p - \tilde{r}^*$, and contain less than $(p+\tilde{r}^*)^N$ sites. Therefore, (3.17), (5.12) and the Markov inequality give,

$$P\left(\sum_{i=1}^{r} |Y_i - Y_i^*| > \epsilon_{\mathbf{n}} \right) \tag{5.13}$$

$$\le 2rp^N B_{\mathbf{n}} (\hat{\mathbf{n}}^* b_{\mathbf{n}}^{\tilde{d}})^{-1} h(\tilde{\mathbf{n}}, (p+\tilde{r}^*)^N) \kappa(p - \tilde{r}^*) \epsilon_{\mathbf{n}}^{-1} \le C\beta_{\mathbf{n}}^*.$$

Using (5.12) and (5.11), $|\lambda_{\mathbf{n}} Y_i^*| < 1/2$. Employing Lemma 5.2 instead of Lemma 3.5, we obtain,

$$P\left(|\sum_{i=0}^{r} Y_i^*| > \epsilon_{\mathbf{n}}\right) \le C\hat{\mathbf{n}}^{-A},$$

and the conclusion follows as in the proof of Lemma 3.8. □

Lemma 5.5. *Assume that Assumptions 1–7 hold.*
(i) Suppose the polynomial mixing rate (2.5) holds. Let s satisfying Assumption 5. If $\theta > N\{3s + \tilde{d}(s+2)\}/(s-1)$*, the Neaderhouser condition (2.2) holds, and (2.19) holds, or if* $\theta > N\{(2\tilde{k}+1)s + \tilde{d}(s+2)\}/(s-1)$*, the Takahata condition (2.3) holds, and (2.20) holds, then*

$$\sup_{x\in D} |\psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}^B(x)| = O(\Psi_{\mathbf{n}}) \quad \textit{in probability.} \tag{5.14}$$

Suppose $\theta > N\{5s + \tilde{d}(s+2)\}/(s-1)$*, the Neaderhouser condition (2.2), (2.21) hold. In addition, suppose* $\theta > N\{(2\tilde{k}+3)s + \tilde{d}(s+2)\}/(s-1)$*, the Takahata condition (2.3) holds, and (2.22) holds then*

$$\sup_{x\in D} |\psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}^B(x)| = O(\Psi_{\mathbf{n}}) \quad a.s. \tag{5.15}$$

(ii) Now assume that the exponential mixing rate (2.6) holds. If the Neaderhouser condition (2.2) or the Takahata condition (2.3) holds, and if (2.12) holds then

$$\sup_{x\in D} |\psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}^B(x)| = O(\Psi_{\mathbf{n}}) \quad a.s. \tag{5.16}$$

Proof. From (5.8), Lemma 5.3, (5.10) and Lemma 5.4, to prove (2.15), it suffices to show that for all $\varsigma > 0$, there exists an $\eta > 0$ such that $v^*\big(\hat{\mathbf{n}}^{-A} + \beta_{\mathbf{n}}^*\big) < \varsigma$ for all $\mathbf{n}$. Choose $\varsigma > 0$. Since A can be chosen arbitrarily large, it is easy to show, as in the proof of Theorem 2.1, that there exists $\eta > 0$ such that $v^*\hat{\mathbf{n}}^{-A} < \varsigma$ for all $\mathbf{n}$.
Assume that the polynomial mixing rate (2.5) holds. We have

$$v^* \le C(l^*)^{-\tilde{d}} \le C\hat{\mathbf{n}}^{\tilde{d}/2} b_{\mathbf{n}}^{-\tilde{d}\{(\tilde{d}/2)+1\}} (\log \hat{\mathbf{n}})^{-\tilde{d}/2} B_{\mathbf{n}}^{\tilde{d}},$$

and with $\tilde{\mathbf{n}}^* = \mathrm{Card}(I_{\mathbf{n}} \cup \mathcal{I}_{\mathbf{n}})$

$$v^*\beta_{\mathbf{n}}^* \le CB_{\mathbf{n}}^{\tilde{d}} \hat{\mathbf{n}}^{(\tilde{d}+1)/2} b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+3)/2} (\log \hat{\mathbf{n}})^{-(\tilde{d}+1)/2} h(\tilde{\mathbf{n}}^*, (p+\tilde{r}^*)^N)(p-\tilde{r}^*)^{-\theta}$$

$$\leq C\hat{\mathbf{n}}^{(\tilde{d}+1)/2+\tilde{d}/s} b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+3)/2} (\log \hat{\mathbf{n}})^{-(\tilde{d}+1)/2} \{g(\mathbf{n})\}^{\tilde{d}/s} h(\tilde{\mathbf{n}}^*, (p+\tilde{r}^*)^N)(p-\tilde{r}^*)^{-\theta}.$$

Using (5.11), the Neaderhouser condition (2.2) implies

$$v^*\beta_{\mathbf{n}}^* \leq C\hat{\mathbf{n}}^{(\tilde{d}+1)/2+\tilde{d}/s} b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+3)/2} (\log \hat{\mathbf{n}})^{-(\tilde{d}+1)/2} \{g(\mathbf{n})\}^{\tilde{d}/s} \hat{\mathbf{n}} \left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}}}{B_{\mathbf{n}} \log \hat{\mathbf{n}}} \right)^{-\frac{\theta}{2N}}$$

$$= C\hat{\mathbf{n}}^{1-\frac{\theta}{2N}+\frac{\theta}{2sN}+(\tilde{d}+1)/2+\tilde{d}/s} b_{\mathbf{n}}^{-\tilde{d}(\theta/N+\tilde{d}+3)/2} (\log \hat{\mathbf{n}})^{\frac{\theta}{2N}-(\tilde{d}+1)/2} \{g(\mathbf{n})\}^{\frac{\theta}{2sN}+\tilde{d}/s}$$

$$= C\left(\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_1^*} (\log \hat{\mathbf{n}})^{\theta_2^*} \{g(\mathbf{n})\}^{\theta_3^*} \right)^{\frac{\theta(1-s)+N\{3s+\tilde{d}(s+2)\}}{2sN}}.$$

Then when $\theta > N\{3s+\tilde{d}(s+2)\}/(s-1)$ the bandwidth condition (2.19) is equivalent to $v^*\beta_{\mathbf{n}}^* \to 0$. Analogously the Takahata condition (2.3) implies

$$v^*\beta_{\mathbf{n}}^* \leq C\hat{\mathbf{n}}^{(\tilde{d}+1)/2+\tilde{d}/s} b_{\mathbf{n}}^{-\tilde{d}(\tilde{d}+3)/2} (\log \hat{\mathbf{n}})^{-(\tilde{d}+1)/2} \{g(\mathbf{n})\}^{\tilde{d}/s} \hat{\mathbf{n}}^{\tilde{k}} \left(\frac{\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}}}{B_{\mathbf{n}} \log \hat{\mathbf{n}}} \right)^{\frac{-\theta}{2N}}$$

$$= C\hat{\mathbf{n}}^{\tilde{k}-\frac{\theta}{2N}+\frac{\theta}{2sN}+(\tilde{d}+1)/2+\tilde{d}/s} b_{\mathbf{n}}^{-\tilde{d}(\theta/N+\tilde{d}+3)/2} (\log \hat{\mathbf{n}})^{\frac{\theta}{2N}-(\tilde{d}+1)/2} \{g(\mathbf{n})\}^{\frac{\theta}{2sN}+\tilde{d}/s}$$

$$= C\left(\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_4^*} (\log \hat{\mathbf{n}})^{\theta_5^*} \{g(\mathbf{n})\}^{\theta_6^*} \right)^{\frac{\theta(1-s)+N\{2\tilde{k}s+s+\tilde{d}(s+2)\}}{2sN}}.$$

Then when $\theta > N\{(2\tilde{k}+1)s+\tilde{d}(s+2)\}/(s-1)$ the bandwidth condition (2.20) is equivalent to $v^*\beta_{\mathbf{n}}^* \to 0$. The proof of (5.14) is completed. The rest of the proof is obtained as in the proof of Theorem 2.1. □

Lemma 5.6. *Assume that Assumptions 1, 3 and 7 hold. If the bandwidth is such that (2.8) holds, then*

$$\sup_{x \in D} |E\psi_{\mathbf{n}}(x) - \psi(x)| = o(\Psi_{\mathbf{n}}).$$

Proof. Under Assumption 3 the conditional density $f_{X_{\mathbf{i}}|X_{(\mathbf{i})}}$ of $X_{\mathbf{i}}$ given $X_{(\mathbf{i})}$ exists and we have

$$\psi(x) = r(x)f(x) = \int_{R^d} \varphi(z) f_{X_{\mathbf{i}}|X_{(\mathbf{i})}}(z|x) f(x) dz = \int_{R^d} \varphi(z) f_{X_{\mathbf{i}},X_{(\mathbf{i})}}(z,x) dz.$$

Since (2.8) implies that $b_{\mathbf{n}} = o(\Psi_{\mathbf{n}})$, under Assumptions 1 and 7,

$$\begin{aligned}
& |E\psi_{\mathbf{n}}(x) - \psi(x)| \\
= & \left| (\hat{\mathbf{n}} b_{\mathbf{n}}^{\tilde{d}})^{-1} \sum_{\mathbf{j} \in I_{\mathbf{n}}} \int_{R^d} \int_{R^{\tilde{d}}} \varphi(z) K\left\{(x-y)/b_{\mathbf{n}}\right\} f_{X_{\mathbf{i}}, X_{(\mathbf{i})}}(z, y) dy dz - \psi(x) \right| \\
= & \left| (b_{\mathbf{n}}^{\tilde{d}})^{-1} \int_{R^d} \int_{R^{\tilde{d}}} \varphi(z) K\left\{(x-y)/b_{\mathbf{n}}\right\} f_{X_{\mathbf{i}}, X_{(\mathbf{i})}}(z, y) dy dz - \psi(x) \right| \\
= & \left| (b_{\mathbf{n}}^{\tilde{d}})^{-1} \int_{R^{\tilde{d}}} \psi(y) K\left\{(x-y)/b_{\mathbf{n}}\right\} dy - \psi(x) \right| \\
= & \left| \int_{R^{\tilde{d}}} K(t) \left\{\psi(x - b_{\mathbf{n}} t) - \psi(x)\right\} dt \right| \le C b_{\mathbf{n}} \int \|t\| K(t) dt \le C b_{\mathbf{n}} = o(\Psi_{\mathbf{n}}).
\end{aligned}$$

□

Lemma 5.7. *Assume that Assumptions 1, 3 and 5 hold. If the bandwidth satisfies (2.8), then*

$$\sup_{x \in D} E|\psi_{\mathbf{n}}(x) - \psi_{\mathbf{n}}^B(x)| = o(\Psi_{\mathbf{n}}).$$

Proof. From Assumptions 3, 1 and 5, we have

$$\begin{aligned}
& E|\psi_{\mathbf{n}}(x) - \psi_{\mathbf{n}}^B(x)| \\
\le & (b_{\mathbf{n}}^{\tilde{d}})^{-1} \int_{R^{\tilde{d}} \times \{|\varphi(z)| \ge B_{\mathbf{n}}\}} \left| \varphi(z) K\left\{(x-y)/b_{\mathbf{n}}\right\} f_{X_{\mathbf{i}}, X_{(\mathbf{i})}}(z, y) \right| dy dz \\
\le & (b_{\mathbf{n}}^{\tilde{d}})^{-1} \int_{R^{\tilde{d}}} |K\left\{(x-y)/b_{\mathbf{n}}\right\}| dy \sup_y \int_{\{|\varphi(z)| \ge B_{\mathbf{n}}\}} \left| \varphi(z) f_{X_{\mathbf{i}}, X_{(\mathbf{i})}}(z, y) \right| dz \\
= & \sup_y \int_{\{|\varphi(z)| \ge B_{\mathbf{n}}\}} \left| \varphi(z) f_{X_{\mathbf{i}}, X_{(\mathbf{i})}}(z, y) \right| dz \\
\le & B_{\mathbf{n}}^{1-s} \sup_y \int_{\{|\varphi(z)| \ge B_{\mathbf{n}}\}} |\varphi(z)|^s f_{X_{\mathbf{i}}, X_{(\mathbf{i})}}(z, y) dz \\
\le & B_{\mathbf{n}}^{1-s} C_s = O\left(\{\hat{\mathbf{n}} g(\mathbf{n})\}^{-1/2} \right) = o(\Psi_{\mathbf{n}}),
\end{aligned}$$

since $s \ge 2$ and (2.8) implies that $b_{\mathbf{n}} = o(1)$. □

Proof of Theorem 2.2. When $f_{\mathbf{n}}(x) \ne 0$,

$$r_{\mathbf{n}}(x) - r(x) = a_1 + a_2 + a_3 + a_4 + a_5 \tag{5.17}$$

with

$$\begin{aligned}
a_1 &= -r(x)\{f_{\mathbf{n}}(x) - f(x)\}/f_{\mathbf{n}}(x) \\
a_2 &= \{\psi_{\mathbf{n}}(x) - \psi_{\mathbf{n}}^B(x)\}/f_{\mathbf{n}}(x) \\
a_3 &= \{E\psi_{\mathbf{n}}(x) - \psi(x)\}/f_{\mathbf{n}}(x) \\
a_4 &= \{E\psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}(x)\}/f_{\mathbf{n}}(x) \\
a_5 &= \{\psi_{\mathbf{n}}^B(x) - E\psi_{\mathbf{n}}^B(x)\}/f_{\mathbf{n}}(x).
\end{aligned}$$

By Assumption 4, Theorem 2.1, Lemma 5.1, Lemma 5.5, Lemma 5.6 and Lemma 5.7 we show that a_1, a_2, a_3, a_4 and a_5 are $O(\Psi_{\mathbf{n}})$'s. □

Acknowlegement The authors are very grateful to the referee for his helpful remarks.

REFERENCES

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

Carbon, M., Tran, L. T. and Wu , B. (1997). Kernel density estimation for random fields. Statist. and Probab. Lett. 36, 115-125.

Cressie, N.A.C. (1991). Statistics for Spatial Data. New York: John Wiley.

Guyon, X. et 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. 6^{th} Franco-Belgian Meeting of Statisticians.

Guyon, X. (1995). Random Fields on a Network. New York: Springer Verlag.

Hallin, M., Lu Z., and Tran, L.T. (2004) Local linear spatial regression. Ann. Statist. 2469-2500.

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

Nahapetian, B. S. (1980). The central limit theorem for random fields with mixing conditions. Adv. in Probability 6, Multicomponent systems. ed. R. L. Dobrushin & Ya G. Sinai, 531-548.

Nahapetian, B. S. (1987). An approach to proving limit theorems for dependent random variables. Theory Prob. 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. and 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. Prob. 23, 1188-1203.

Ripley, B. (1981). Spatial Statistics. New York: Wiley.

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

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.