JOURNAL OF MULTIVARIATE ANALYSIS 44, 23-46 (1993)

# Nearest Neighbor Estimators for Random Fields

LANH TAT TRAN AND SIDNEY YAKOWITZ*

*Department of Mathematics, Indiana University, Bloomington, Indiana 47405, and Systems and Industrial Engineering Department, University of Arizona, Tucson, Arizona 85721*

Generalizing the random sequence case, this study defines a $k-NN$ density estimator for random variables with multidimensional lattice points serving as index values. The central result is that under random field stationary and mixing assumptions, as well as standard smoothness postulates, our $k-NN$ estimate is found to be asymptotically normal. This result readily extends to $NN$-type estimates for jointly distributed random variables. For illustration, a simplified version of the $k-NN$ estimator is applied to obtain the density estimate for a soil-moisture data set selected from the geostatistical literature. 

## 1. INTRODUCTION

The nonparametric estimation of a probability density $f(x)$ is an interesting problem in statistical inference and has already played an important role in communication and pattern recognition. The literature dealing with density estimation when the observations are independent is extensive. The reader is referred to Izenman [11] and Wegman [36] for a review. 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 in order to provide alternative methodology beyond parametric second-order methods such as surveyed in (Ripley [30, Chap. 4]) or multivariate interpolation algorithms (e.g., Franke [5]) for situations in which parametric families cannot be adopted with confidence. A comparative application to soil moisture data is included in our Section 5.

Let $Z^N$, $N \geqslant 1$, denote the integer lattice points in the $N$-dimensional

Received March 18, 1991; revised May 21, 1992.
AMS 1980 subject classifications: primary 62G05, secondary 60J25.
Key words and phrases: random field, nearest neighbor estimator, mixing, kriging.
* This author's research supported, in part, by NIH Grant No. AI29426-01.



0047-259X/93 $5.00

Euclidean space. Let $\{X_{\mathbf{n}}\}$ be a strictly stationary random field 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 and written as $\mathbf{n} = \langle n_1, ..., n_N \rangle$. Let $S$ and $S'$ be two finite sets of sites. 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 $\sigma$-fields generated by the random variables $X_{\mathbf{n}}$ with $\mathbf{n}$ ranging over $S$ and $S'$, respectively. Let $\hat{d}(S, S')$ be the Euclidean distance between $S$ and $S'$. 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 $S, S' \subset Z^N$,

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

where $\operatorname{Card}(S)$ denotes the cardinality of $S$. Here $\hat{f}$ is a symmetric positive function nondecreasing in each variable. Throughout the paper, assume that $\hat{f}$ satisfies either

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

or

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

for some $\tilde{k} > 1$ and some $C > 0$. If $\hat{f} \equiv 1$, then $X_{\mathbf{n}}$ is called strongly mixing. In case $N = 1$, many stochastic processes and time series are known to be strongly mixing. Withers [37] has obtained various conditions for linear processes to be strongly mixing. Under certain weak assumptions autoregressive and more generally bilinear time series models are strongly mixing with exponential mixing rates. See Pham and Tran [29] and Pham [28]. Guyon [8] has shown that results of Withers extend to random fields $X_{\mathbf{n}} = \sum_{Z^N} g_{\mathbf{j}} Z_{\mathbf{n}-\mathbf{j}}$ with the $g_{\mathbf{j}}$'s and $Z_{\mathbf{j}}$'s satisfying certain conditions. Here $Z_{\mathbf{j}}$'s are independent rv's. Conditions (1.2) and (1.3) are the same as the mixing conditions used by Neaderhouser [26] and Takahata [34], respectively, and are weaker than the uniform mixing condition used by Nahapetian [24]. They are satisfied by many spatial models. Examples can be found in Neaderhouser [26], Rosenblatt [31], and Guyon [8]. For relevant works on random fields, see, e.g., Neaderhouser [26], Bolthausen [2], Guyon and Richardson [7], and Guyon [8].

Let $I_{\mathbf{n}}$ be a rectangular region defined by $I_{\mathbf{n}} = \{\mathbf{i} : \mathbf{i} \in Z^N, 1 \leqslant i_k \leqslant n_k, k = 1, ..., N\}$. Assume that we observe $\{X_{\mathbf{n}}\}$ on $I_{\mathbf{n}}$. Suppose $X_{\mathbf{n}}$ takes values in $R^d$ and has density $f$. The letter $C$ will be used to denote constants whose values are unimportant. We write $\mathbf{n} \to \infty$ if $\min\{n_k\} \to \infty$ and $|n_j/n_k| < C$ for some $0 < C < \infty$, $1 \leqslant j$, $k \leqslant N$. Let $\hat{\mathbf{n}} = n_1 \cdots n_N$. All limits are taken as $\mathbf{n} \to \infty$ unless indicated otherwise. For a site $\mathbf{i}$, we denote $\|\mathbf{i}\| = (i_1^2 + \cdots + i_N^2)^{1/2}$. Throughout the paper, $x$ denotes a fixed point of $R^d$.

Let $\Gamma_{\mathbf{n}}(x)$ be the Euclidean distance between $x$ and its $k$th nearest neighbor among the $X_{\mathbf{i}}$'s, where $k = k(\hat{\mathbf{n}})$ is a sequence of positive integers satisfying

$$k(\hat{\mathbf{n}}) \uparrow \infty, \qquad k(\hat{\mathbf{n}})\,\hat{\mathbf{n}}^{-1} \to 0.$$

The nearest neighbor density estimator $f_{\mathbf{n}}(x)$ of $f(x)$ is defined as

$$f_{\mathbf{n}}(x) = \frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}} V_{\Gamma_{\mathbf{n}}}(x)}, \tag{1.4}$$

where $V_{\Gamma_{\mathbf{n}}}(x)$ is the volume of the sphere centered at $x$ with radius $\Gamma_{\mathbf{n}} \equiv \Gamma_{\mathbf{n}}(x)$. The $k-NN$ density estimator was proposed by Loftsgaarden and Quesenberry [16] in the independent case. Consistency and asymptotic normality of $f_n$ have been obtained under various conditions. See, e.g., Devroye and Wagner [4] and Moore and Yackel [23]. Further discussion on the $k-NN$ method and its advantages can be found in Mack and Rosenblatt [19] and Mack [17].

Density estimation under dependence has received considerable attention. See, for example, Roussas [32, 33], Masry [21, 22], Ioannides and Roussas [10], and the references therein. Nearest neighbor methods have been investigated by Yakowitz [38], Boente and Fraiman [1], and Nguyen and Tran [27]. The reader is referred to Moore and Yackel [23], Mack and Rosenblatt [19], Tran [35], and Yakowitz [38] for background material.

Kernel density estimation on random fields has recently been investigated by Tran [35]. Nearest neighbor models have been employed widely to model stochastic images, particularly for image restoration and coding. For an account of this information, see Levy [15] and Jain [12]. More motivation for the $NN$ method is given in Section 5.

Our main result (see Theorem 3.1) gives conditions under which $f_{\mathbf{n}}$ is asymptotically normal. Much work remains to be done in density estimation on random fields. We only consider the nearest neighbor density estimator with uniform weight function here. Due to complications caused by spatial dependence, the analog of our main result for general weight functions has so far eluded us.

## 2. Preliminaries

The following lemma can be found in Ibragimov and Linnik [9] or Deo [3].

LEMMA 2.1. (i) *Suppose* (1.1) *holds. Let* $\mathscr{L}_r(\mathscr{F})$ *denote the class of* $\mathscr{F}$*-measurable rv's* $X$ *satisfying* $\|X\|_r = (E|X|^r)^{1/r} < \infty$. *Let* $X \in \mathscr{L}_r(\mathscr{B}(S))$ *and* $Y \in \mathscr{L}_s(\mathscr{B}(S'))$. *Suppose* $1 \leqslant r, s, h < \infty$, *and* $r^{-1} + s^{-1} + h^{-1} = 1$, *then*

$$|EXY\text{-}EXEY| \leqslant C\,\|X\|_r\,\|Y\|_s\,\{\hat{f}(\mathrm{Card}(S), \mathrm{Card}(S'))\,\varphi(\hat{d}(S, S'))\}^{1/h}. \tag{2.1}$$

(ii) *For rv's bounded with probability* 1, *the right-hand side of* (2.1) *can be replaced by* $C\hat{f}(\mathrm{Card}(S), \mathrm{Card}(S'))\,\varphi(\hat{d}(S, S'))$.

We will need the following lemma from Nakhapeytyan [25]:

LEMMA 2.2. *Let* $(\xi_1, ..., \xi_n)$ *be a random vector such that* $|C\xi_i| \leqslant 1$, $i = 1, ..., n$. *Then*

$$\Big|E\prod_{s=1}^{n} \xi_s - \prod_{s=1}^{n} E\xi_s\Big| \leqslant \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Big|E(\xi_i - 1)(\xi_j - 1) \times \prod_{s=j+1}^{n} \xi_s - E(\xi_i - 1)\,E(\xi_j - 1) \prod_{s=j+1}^{n} \xi_s\Big|.$$

*Assumption* 1. The density $f$ is continuous on $R^d$ and locally Lipschitz of order $\rho$ at $x$; that is, $|f(y) - f(x)| \leqslant C\,\|x - y\|^{\rho}$ for some $\rho > 0$ and all $y \in R^d$. In addition, $f(x) > 0$.

*Assumption* 2. There exists a constant $C$ and some neighborhood of $(x, x)$ such that $|f_{\mathbf{i},\mathbf{j}}(u, v) - f(u)f(v)| < C$ for all $\mathbf{i} \neq \mathbf{j}$ and for all $(u, v)$ in this neighborhood. Here, $f_{\mathbf{i},\mathbf{j}}$ denotes the joint density of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$.

*Assumption* 3. The sequence $k(\hat{\mathbf{n}})$ tends to infinity slowly enough that

$$k(\hat{\mathbf{n}}) = o(\hat{\mathbf{n}}^{2\rho/(d + 2\rho)}).$$

*Assumption* 4. There exist sequences of positive integers $p \equiv p(\mathbf{n})$ and $q \equiv q(\mathbf{n})$ tending to infinity in such a manner that

(i) $p^N = o((k(\hat{\mathbf{n}}))^{1/2})$ as $\mathbf{n} \to \infty$,

(ii) $\hat{\mathbf{n}} \sum_{i=1}^{\infty} i^{N-1} \varphi(iq) \to 0$.

*Assumption* 5. For some $\delta > 0$,

(i) $(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{-\delta/(2+\delta)} \sum_{i=q}^{\infty} i^{N-1} (\varphi(i))^{\delta/(2+\delta)} \to 0$,

(ii) $(q/p)(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{-\delta/(2+\delta)} \to 0$.

*Assumption* 6. There exists a sequence of positive numbers $c_{\mathbf{n}} \uparrow \infty$ such that

(i) $c_{\mathbf{n}}^N = o(\hat{\mathbf{n}}/k(\hat{\mathbf{n}}))$,

(ii) $\hat{\mathbf{n}}^{(2+2\delta)/(2+\delta)} (k(\hat{\mathbf{n}}))^{-\delta/(2+\delta)} p^{-N} \sum_{\|\mathbf{i}\| > c_{\mathbf{n}}} \{\varphi(\|\mathbf{i}\|)\}^{\delta/(2+\delta)} \to 0$.

Define $S(x, r)$ to be the sphere of radius $r$ centered at $x$ and let

$$H(r) = P\{\|X - x\| \leqslant r\} = \int_{S(x,r)} f(u)\, du. \tag{2.2}$$

The following result is crucial for the proof of Theorem 3.1 and is also of independent interest.

THEOREM 2.1. *Suppose Assumptions* 1–6 *hold. Then*

$$A(\mathbf{n}, 1) \equiv (k(\hat{\mathbf{n}}))^{1/2} \left( \frac{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{H(\Gamma_{\mathbf{n}})} - 1 \right) \xrightarrow{\mathscr{L}} N(0, 1),$$

*where* $N(0, 1)$ *denotes the standard normal random variable.*

LEMMA 2.3. *Suppose Assumptions* 1–6 *hold. Then* $f_{\mathbf{n}}(x)$ *converges to* $f(x)$ *in probability.*

Proofs of Theorem 2.1 and Lemma 2.3 are relegated to the Appendix.

## 3. ASYMPTOTIC NORMALITY OF $f_{\mathbf{n}}$

THEOREM 3.1. *Suppose Assumptions* 1–6 *hold. Then*

$$(k(\hat{\mathbf{n}}))^{1/2} (f_{\mathbf{n}}(x) - f(x)) \xrightarrow{\mathscr{L}} N(0, f^2(x)).$$

*Proof.* Theorem 2.1 implies that

$$\frac{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{H(\Gamma_{\mathbf{n}})} \to 1 \tag{3.1}$$

in probability. Thus $H(\Gamma_{\mathbf{n}}) \to 0$ by Assumption 3. Hence $\Gamma_{\mathbf{n}} \to 0$ in probability since $f(x)$ is positive and continuous at $x$. By (2.2) and the mean value theorem for integrals, there exists some $x_{\mathbf{n}} \in S(x, \Gamma_{\mathbf{n}})$ such that

$$H(\Gamma_{\mathbf{n}}) = f(x_{\mathbf{n}})\, V_{\Gamma_{\mathbf{n}}}(x).$$

Consequently,

$$f(x_{\mathbf{n}}) = \frac{H(\Gamma_{\mathbf{n}})\, k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{V_{\Gamma_{\mathbf{n}}}(x)\, k(\hat{\mathbf{n}})/\hat{\mathbf{n}}} = f_{\mathbf{n}}(x) \frac{H(\Gamma_{\mathbf{n}})}{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}.$$

A simple computation shows

$$\frac{(k(\hat{\mathbf{n}}))^{1/2}}{f(x)}(f_{\mathbf{n}}(x)-f(x))=A(\mathbf{n},1)+A(\mathbf{n},2), \tag{3.2}$$

where

$$A(\mathbf{n},2)\equiv(k(\hat{\mathbf{n}}))^{1/2}\left(\frac{f(x_{\mathbf{n}})}{f(x)}-1\right)\frac{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{H(\Gamma_{\mathbf{n}})}. \tag{3.3}$$

Employing Assumption 1 and the fact that $x_{\mathbf{n}} \in S(x, \Gamma_{\mathbf{n}})$, we then obtain

$$(k(\hat{\mathbf{n}}))^{1/2}\left(\frac{f(x_{\mathbf{n}})}{f(x)}-1\right)\leqslant C(k(\hat{\mathbf{n}}))^{1/2}\,\|x_{\mathbf{n}}-x\|^{\rho}\leqslant C(k(\hat{\mathbf{n}}))^{1/2}\,\Gamma_{\mathbf{n}}^{\rho}. \tag{3.4}$$

Let $c$ denote the volume of a sphere in $R^d$ with radius 1. Lemma 2.3 implies that $f_{\mathbf{n}}(x)\to f(x)$ in probability. However, $f(x)$ is positive by Assumption 1. Employing also Assumption 3, for any given $\varepsilon>0$,

$$\begin{aligned}P\{(k(\hat{\mathbf{n}}))^{1/2}\,\Gamma_{\mathbf{n}}^{\rho}>\varepsilon\}&=P\{(k(\hat{\mathbf{n}}))^{1/2}\,(\Gamma_{\mathbf{n}}^{d})^{\rho/d}>\varepsilon\}\\&=P\{(k(\hat{\mathbf{n}}))^{1/2}\,\{k(\hat{\mathbf{n}})/(\hat{\mathbf{n}}cf_{\mathbf{n}}(x))\}^{\rho/d}>\varepsilon\}\to 0.\end{aligned} \tag{3.5}$$

Combining (3.1), (3.3), (3.4), and (3.5),

$$A(\mathbf{n},2)\to 0 \tag{3.6}$$

in probability. The theorem follows from (3.2), Theorem 2.1, and (3.6).

EXAMPLE 3.1. Suppose $\varphi(x)=O(x^{-v})$ for some $v>N$ and Assumptions 1–2 hold. Choose $k(\hat{\mathbf{n}})\uparrow\infty$ such that

$$k(\hat{\mathbf{n}})=o(\hat{\mathbf{n}}^{2\rho/(d+2\rho)}). \tag{3.7}$$

Then Assumption 3 is satisfied. Let $\delta>0$. Choose

$$q\equiv(k(\hat{\mathbf{n}}))^{(2+\delta+2N\delta)/[2N(2+\delta)]}\,\hat{\mathbf{n}}^{-\delta/(2+\delta)}\,(\log\log k(\hat{\mathbf{n}}))^{-2/N}, \tag{3.8}$$

for some $\delta>2N/(v-N)$; that is, $N<\delta v/(2+\delta)$. Let

$$p=(k(\hat{\mathbf{n}}))^{1/(2N)}\,(\log\log k(\hat{\mathbf{n}}))^{-1/N}. \tag{3.9}$$

Then Assumption 4(i) is satisfied.

Assumption 4(ii) is satisfied if

$$\hat{\mathbf{n}}\sum_{i=1}^{\infty} i^{N-1}(iq)^{-v}=\hat{\mathbf{n}}q^{-v}\sum_{i=1}^{\infty} i^{N-1-v}\to 0 \tag{3.10}$$

or

$$\hat{\mathbf{n}} q^{-\nu} \to 0. \tag{3.11}$$

Assumption (5)(i) is satisfied if

$$\begin{aligned}(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{-\delta/(2+\delta)} \sum_{i=q}^{\infty} i^{N-1-(\delta\nu/(2+\delta))} \\ = (k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{-\delta/(2+\delta)} q^{N-(\delta\nu/(2+\delta))} \to 0,\end{aligned}$$

which is equivalent to

$$(\hat{\mathbf{n}}/k(\hat{\mathbf{n}}))^{-\delta} q^{(\delta\nu - 2N - N\delta)} \to \infty. \tag{3.12}$$

Clearly,

$$\begin{aligned} q &= p(k(\hat{\mathbf{n}}))^{\delta/(2+\delta)} \hat{\mathbf{n}}^{-\delta/(2+\delta)} (\log\log k(\hat{\mathbf{n}}))^{-1/N} \\ &= p(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{\delta/(2+\delta)} (\log\log k(\hat{\mathbf{n}}))^{-1/N}. \end{aligned} \tag{3.13}$$

Thus Assumption 5(ii) is satisfied. Choose

$$c_{\mathbf{n}}^{N} = \hat{\mathbf{n}}(k(\hat{\mathbf{n}}) \log\log \hat{\mathbf{n}})^{-1}. \tag{3.13}$$

Then Assumption 6(ii) becomes

$$\hat{\mathbf{n}}^{(2+2\delta)/(2+\delta)}(k(\hat{\mathbf{n}}))^{-\delta/(2+\delta)} p^{-N} \{c_{\mathbf{n}}^{N}\}^{N-(\nu\delta/(2+\delta))} \to 0$$

or

$$\hat{\mathbf{n}}^{(2+2\delta)}(k(\hat{\mathbf{n}}))^{-\delta} p^{-N(2+\delta)} \{c_{\mathbf{n}}^{N}\}^{N(2+\delta)-\nu\delta} \to 0. \tag{3.14}$$

Summarizing, we have

PROPOSITION 3.1. *Suppose Assumptions* 1–2 *hold and* $\varphi(x) = O(x^{-\nu})$ *for some* $\nu > N$ *as* $x \to \infty$. *Let* $k(\hat{\mathbf{n}}) \uparrow \infty$ *in such a manner that*

$$k(\hat{\mathbf{n}}) = o(\hat{\mathbf{n}}^{2\rho/(d+2\rho)}).$$

Suppose that $q \to \infty$, where $q$ is given in (3.8). In addition, assume that (3.11), (3.12), and (3.14) are satisfied. Then the conclusion of Theorem 3.1 holds. As a specific example, suppose $N = 2$ and Assumptions 1 and 2 hold with $\rho$ equal to 1. Choose $k(\hat{\mathbf{n}}) = \hat{\mathbf{n}}^{1/2}$ and $\delta = \frac{1}{4}$. Then

$$q = \hat{\mathbf{n}}^{5/72}(\log\log(\hat{\mathbf{n}}^{1/2}))^{-1}$$

and

$$p = \hat{\mathbf{n}}^{1/8}(\log\log(\hat{\mathbf{n}}^{1/2}))^{-1/2}.$$

A simple computation shows that (3.11), (3.12), and (3.14) hold if $v$ is greater than $\frac{72}{5}$, $\frac{126}{5}$, and $\frac{65}{2}$, respectively. Thus Proposition 1 holds if $v > \frac{65}{2}$.

## 4. Estimating Joint Densities

Let $\mathbf{j}_1, \mathbf{j}_2, ..., \mathbf{j}_m$ be $m$ distinct sites, where $m$ is a fixed positive integer greater than 1. Suppose $X_{\mathbf{j}_1}, ..., X_{\mathbf{j}_m}$ have joint density $f_{\mathbf{j}_1, ..., \mathbf{j}_m} \equiv g$, say. We will consider nearest neighbor estimates of $g$. Let

$$Y_{\mathbf{i}} = (X_{\mathbf{i}+\mathbf{j}_1}, ..., X_{\mathbf{i}+\mathbf{j}_m}). \tag{4.1}$$

Note that $Y_{\mathbf{i}}$ is strictly stationary with density $g$. The random field $Y_{\mathbf{i}}$ is strong mixing with values in $R^{d^*}$ with $d^* = md$. Let $y \in R^{d^*}$ and let $g_{\mathbf{n}}(y)$ denote the nearest neighbor density estimator of $g$ based on the $Y_{\mathbf{i}}$'s with $\mathbf{i}+\mathbf{j}_1 \in I_{\mathbf{n}}, ..., \mathbf{i}+\mathbf{j}_m \in I_{\mathbf{n}}$. By a slight variation of the proof of Theorem 3.1, we obtain:

Proposition 4.1. *Suppose $X_{\mathbf{n}}$ is strong mixing satisfying* (1.1)–(1.3) *and Assumptions* 1–6 *hold with $f$ and $d$ replaced by $g$ and $d^*$, respectively. Then*

$$(k(\hat{\mathbf{n}}))^{1/2}\,(g_{\mathbf{n}}(y) - g(y)) \xrightarrow{\mathscr{L}} (0, g^2(y)). \tag{4.2}$$

## 5. Motivations, Applications, and a Computational Illustration

As mentioned at the outset, this study is motivated by a desire to provide an alternative to the parametric second-order methodology now dominating the literature of spatial statistics. A previous foray into this task (Yakowitz and Szidarovsky [41]) was deficient for the random field problem in that it assumed that noises at different sites are independent. The present work overcomes this weakness, but still does not master the problem because the regression issue yet remains. The authors are presently at work on this gap.

Laslett *et al.* [14] have published an interesting set of hydrologic data which we anticipate will serve as the test bed in some of our forthcoming studies. Specifically, they have presented an 11 by 11 lattice of soil-surface pH measurements, which can be used to calibrate prediction and regression schemes. Within the same geographic region, they have obtained a second lattice which can serve to test proposed prediction methods. Rigorously justified $NN$ random-field prediction still remains a goal for us, but in Fig. 1, below, we have used the $k-NN$ density estimator in (1.4), with $k = 30$, to obtain a density estimate for this data set.

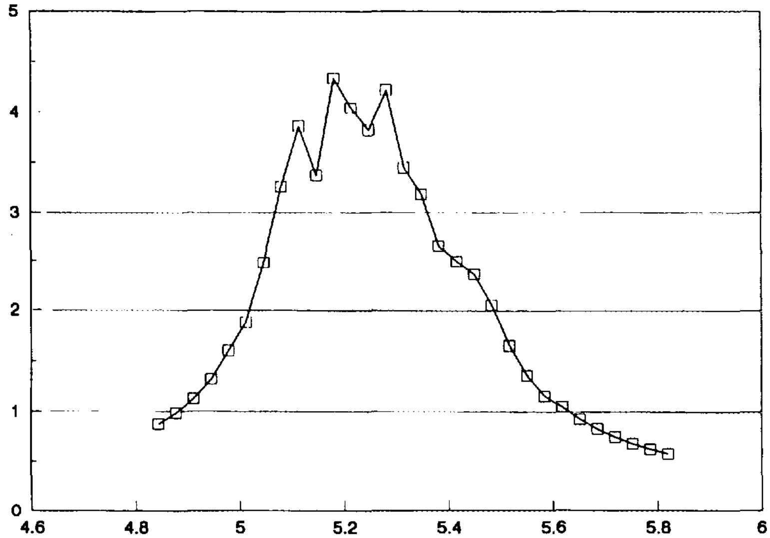


FIG. 1. A NN density estimator constructed from a random field observation.

Applied science and engineering interest in spatial methods includes topics of concern to your authors, most notably digital image analysis and spatial modelling of epidemics. We proceed to discuss points of contact between the present (and other on-going) studies with applications areas.

With respect to image analysis, as described in Geman and Geman [6], henceforth referred to as G&G, the shading of a pixel may, in the language of the present paper, be regarded as the observable $X(\mathbf{n})$ at a site $\mathbf{n}$. G&G add the constraint that the random field be homogeneous, isotropic Markov, which implies that the probability law for $X$ must be Gibbsian. An outgrowth of the theory is that associated with any site $\mathbf{n}$ is a vector $V(\mathbf{n})$ of $x$-values of fixed dimension and fixed ($x(\mathbf{n})$-deleted) "neighborhood" structure so that in distribution,

$$X(\mathbf{n}) \mid \{x(\mathbf{n}'), \mathbf{n}' \neq \mathbf{n}\} = \mathbf{X}(\mathbf{n}) \mid \mathbf{V}(\mathbf{n}). \tag{5.1}$$

The point is that the value at a site depends on other values in the field only through values at its neighbors. This important relation serves, in G&G, as a basis for Bayesian image analysis as well as a key construct for "Gibbs sampling," a simulation technique for computing the Bayes MAP image.

Statistical issues associated with this approach to image analysis have not been satisfactorily charted. Kashyap and Chellapa [13] and Section VII of G&G have some thoughts, but in the words of G&G, "A *general theory* [their italics] of interactive, self-adjusting models that is practical and mathematically coherent may lie far ahead." In particular, G&G do not address the problem of inferring the noise process from data, but just postulate that it is i.i.d. normal. Similarly, in situations in which learning images are available, inference of the local conditional density for (5.1) could give a rationale for the selection or the prior Gibbs law. Our thesis is that theory for spatial statistics in the face of dependency, such as explored in the present work, will play a central role in the statistical side of the image-processing endeavor. Thus, the joint-density construct of Section 4 can provide a nearest neighbor density estimator for the conditional density $f_{X(\mathbf{n})|V(\mathbf{n})}(x_{\mathbf{n}}|v)$ of the key MRF variable in (5.1). For if $g_{\mathbf{n}}(x_{\mathbf{n}}, v)$ is the multivariate NN estimator for the joint variable $(X_{\mathbf{n}}, V(\mathbf{n}))$ based on values over MRF-type neighborhood vectors without $x_{\mathbf{n}}$ deleted, then one can approximate

$$f_{X(\mathbf{n})|V(\mathbf{n})}(x_{\mathbf{n}}|v) \sim \frac{g_{\mathbf{n}}(x_{\mathbf{n}}, v)}{\hat{g}_{\mathbf{n}}(v)}, \tag{5.2}$$

where $\hat{g}_{\mathbf{n}}(v)$ is a density estimate of the vector $V$ obtained by applying the multivariate density method to deleted neighborhoods.

This conditional law itself is fundamental, but in turn, it can play a role in finding the potential $U(x)$ for the Gibbs invariant law that is appropriate for a clique class for the images of interest. Nearest neighbor methods for nonparametric inference, because they have intuitive content and are easy to visualize, have always been in favor in the engineering literature.

More recent developments are referenced in Marroquìn and Ramìrez [20] which puts the image processing plan into a dynamic MRF setting in which (5.1) is the vehicle for "asynchronous" evolution from one state to the next, and the Gibbs sampling idea is thereby tied to *random cellular automata* (RCAs). This perspective, and the claim that the present paper can provide a data-driven inference of the conditional density $f_{X(\mathbf{n})|V(\mathbf{n})}(x_{\mathbf{n}}|v)$, yield a bridge to work on spatial epidemics.

The name and concept "random cellular automata" is featured in the study by Yakowitz *et al.* [39] which shows that spatial RCAs proposed for epidemics by Durrett, Mollison, and others, can be made to coincide with the pervasive classical space-invariant "Susceptible, Infective, Removal" Markov epidemic model if the initial populations are strung out homogeneously on the lattice; but otherwise, the evolution is different. This analysis was intended to illustrate that the spatial nature of epidemics cannot safely be ignored. The subsequent investigation described in Yakowitz *et al.* [40] adds an inoculation band to the RCA model and poses and solves the optimal band placement problem. Toward relating these two works and other "contact" process models to practical problems, the statistical inference issue must be addressed.

## APPENDIX

*Proof of Theorem* 2.1. Let $U_{\mathbf{i}} = H(\|X_{\mathbf{i}} - x\|)$. Then $H(\Gamma_{\mathbf{n}})$ is the $(k(\hat{\mathbf{n}}))$th order statistic of $\hat{\mathbf{n}}$ identically distributed uniform rv's on (0, 1). For any $a > 0$,

$$P_{\mathbf{n}}(a) \equiv P\left\{(k(\hat{\mathbf{n}}))^{1/2}\left(\frac{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{H(\Gamma_{\mathbf{n}})} - 1\right) \leqslant a\right\} = P\left\{H(\Gamma_n) \geqslant \frac{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{1 + a(k(\hat{\mathbf{n}}))^{1/2}}\right\}. \tag{A.1}$$

Let

$$\pi_{\hat{\mathbf{n}}} = \frac{k(\hat{\mathbf{n}})/\hat{\mathbf{n}}}{1 + a(k(\hat{\mathbf{n}}))^{1/2}}. \tag{A.2}$$

Define

$$\begin{aligned} \xi_{\hat{\mathbf{n}}\mathbf{i}} &= 1 - \pi_{\hat{\mathbf{n}}} \quad && \text{if} \quad U_{\mathbf{i}} < \pi_{\hat{\mathbf{n}}} \\ &= -\pi_{\hat{\mathbf{n}}} && \text{otherwise.} \end{aligned} \tag{A.3}$$

Clearly $H(\Gamma_{\mathbf{n}}) \geqslant \pi_{\hat{\mathbf{n}}}$ if and only if the number of $U_{\mathbf{i}}$'s, $\mathbf{i} \in I_{\mathbf{n}}$, with $U_{\mathbf{i}} < \pi_{\hat{\mathbf{n}}}$ is smaller than $k(\hat{\mathbf{n}})$. Thus

$$\begin{aligned} P_{\mathbf{n}}(a) &= P\left\{ \sum_{\substack{i_k = 1 \\ k = 1, \ldots, N}}^{n_k} (\xi_{\hat{\mathbf{n}}\mathbf{i}} + \pi_{\hat{\mathbf{n}}}) < k(\hat{\mathbf{n}}) \right\} \\ &= P\left\{ \frac{1}{\sigma_{\hat{\mathbf{n}}}} \sum_{\substack{i_k = 1 \\ k = 1, \ldots, N}}^{n_k} \xi_{\hat{\mathbf{n}}\mathbf{i}} < \frac{k(\hat{\mathbf{n}}) - \hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}}{\sigma_{\hat{\mathbf{n}}}} \right\}, \end{aligned} \tag{A.4}$$

where

$$\sigma_{\hat{\mathbf{n}}} \equiv (\hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}(1 - \pi_{\hat{\mathbf{n}}}))^{1/2}.$$

Assumption 3 and (A.2) imply that $\pi_{\hat{\mathbf{n}}} \to 0$ and hence $1 - \pi_{\hat{\mathbf{n}}} \to 1$. By (A.2), we have $k(\hat{\mathbf{n}})/(\hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}) \to 1$ since $k(\hat{\mathbf{n}}) \to \infty$ by Assumption 3. A simple computation then shows that

$$\frac{k(\hat{\mathbf{n}}) - \hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}}{\sigma_{\hat{\mathbf{n}}}} = a\left(\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}}\right)^{1/2} \times \left(\frac{1}{1 - \pi_{\hat{\mathbf{n}}}}\right)^{1/2} \times \frac{1}{1 + a(k(\hat{\mathbf{n}}))^{-1/2}} \to a.$$

The lemma can thus be completed by showing that

$$\frac{1}{\sigma_{\hat{\mathbf{n}}}} \sum_{\substack{i_k = 1 \\ k = 1, \ldots, N}}^{n_k} \xi_{\hat{\mathbf{n}}\mathbf{i}} \xrightarrow{\mathscr{L}} N(0, 1). \tag{A.5}$$

Let

$$S_{\mathbf{n}} = \sum_{\substack{i_k=1\\ k=1,\ldots,N}}^{n_k} \xi_{\hat{\mathbf{n}}\mathbf{i}}. \tag{A.6}$$

Assumption 4(i) and Assumption 3 imply that $p^N/\hat{\mathbf{n}} \to 0$ as $\hat{\mathbf{n}} \to \infty$. Since $n_j/n_k < C$ for some $0 < C < \infty$, $1 \leqslant j, k \leqslant N$, it follows that $p^N/\min\{n_1, \ldots, n_N\} \to 0$. Assumption 3 implies that $k(\hat{\mathbf{n}})/\hat{\mathbf{n}} \to 0$. Thus $q/p \to 0$ by Assumption 5(i). We write $\xi_{\hat{\mathbf{n}}\mathbf{i}}$ as $\xi_{\mathbf{i}}$ for simplicity. Assume for some integers $r_1, \ldots, r_N$ that we have $n_1 = r_1(p+q), \ldots, n_N = r_N(p+q)$. The rv's $\xi_{\mathbf{i}}$'s are now set into large blocks and small blocks. Let

$$U(1, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1\\ k=1,\ldots,N}}^{j_k(p+q)+p} \xi_{\mathbf{i}}, \tag{A.7}$$

$$U(2, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1\\ k=1,\ldots,N-1}}^{j_k(p+q)+p} \; \sum_{i_N = j_N(p+q)+p+1}^{(j_N+1)(p+q)} \xi_{\mathbf{i}}, \tag{A.8}$$

$$U(3, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1\\ k=1,\ldots,N-2}}^{j_k(p+q)+p} \; \sum_{i_{N-1} = j_{N-1}(p+q)+p+1}^{(j_{N-1}+1)(p+q)} \; \sum_{i_N = j_N(p+q)+1}^{j_N(p+q)+p} \xi_{\mathbf{i}}, \tag{A.9}$$

$$U(4, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+1\\ k=1,\ldots,N-2}}^{j_k(p+q)+p} \; \sum_{i_{N-1} = j_{N-1}(p+q)+p+1}^{(j_{N-1}+1)(p+q)} \; \sum_{i_N = j_N(p+q)+p+1}^{(j_N+1)(p+q)} \xi_{\mathbf{i}}, \tag{A.10}$$

and so on. Note that

$$U(2^{N-1}, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+p+1\\ k=1,\ldots,N-1}}^{(j_k+1)(p+q)} \; \sum_{i_N = j_N(p+q)+1}^{j_N(p+q)+p} \xi_{\mathbf{i}}. \tag{A.11}$$

Finally

$$U(2^N, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p+q)+p+1\\ k=1,\ldots,N}}^{(j_k+1)(p+q)} \xi_{\mathbf{i}}. \tag{A.12}$$

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

$$T(\mathbf{n}, i) = \sum_{\substack{j_k=0\\ k=1,\ldots,N}}^{r_k-1} U(i, \mathbf{n}, \mathbf{j}). \tag{A.13}$$

Clearly,

$$S_{\mathbf{n}} = \sum_{i=1}^{2^N} T(\mathbf{n}, i). \tag{A.14}$$

Note that $T(\mathbf{n}, 1)$ is the sum of the random variables $\xi_{\mathbf{i}}$ in large blocks. The $T(\mathbf{n}, i)$, $2 \leq i \leq 2^N$ are sums of random variables in small blocks. If it is not the case that $n_1 = r_1(p+q), ..., n_N = r_N(p+q)$ for some integers $r_1, ..., r_N$, then a term, say, $T(\mathbf{n}, 2^N+1)$, containing all the $\xi_{\mathbf{i}}$'s at the ends not included in the big or small blocks, can be added. This term will not change the proof much.

Clearly,

$$\frac{1}{\sigma_{\hat{\mathbf{n}}}} \sum_{\substack{i_k=1 \\ k=1, ..., N}}^{n_k} \xi_{\hat{\mathbf{n}}} = \frac{T(\mathbf{n}, 1)}{\sigma_{\hat{\mathbf{n}}}} + \sum_{i=2}^{2^N} \frac{T(\mathbf{n}, i)}{\sigma_{\hat{\mathbf{n}}}}. \tag{A.15}$$

Theorem 2.1 will follow from (A.5) and (A.15) if we can show that

$$\frac{T(\mathbf{n}, 1)}{\sigma_{\hat{\mathbf{n}}}} \xrightarrow{\mathscr{L}} N(0, 1), \tag{A.16}$$

$$\sum_{i=2}^{2^N} \frac{T(\mathbf{n}, i)}{\sigma_{\hat{\mathbf{n}}}} \to 0 \tag{A.17}$$

in probability.

The general approach is to show that

$$\sigma_{\hat{\mathbf{n}}}^{-2} \sum_{\substack{j_k=0 \\ k=1, ..., N}}^{r_k-1} E[U(1, \mathbf{n}, \mathbf{j})]^2 \to 1, \tag{A.18}$$

$$R \equiv \left| E\exp[iu(T(\mathbf{n}, 1)/\sigma_{\hat{\mathbf{n}}})] - \prod_{\substack{j_k=0 \\ k=1, ..., N}}^{r_k-1} E\exp[iu(U(1, \mathbf{n}, x, \mathbf{j})/\sigma_{\hat{\mathbf{n}}})]\right| \to 0, \tag{A.19}$$

$$\hat{R} \equiv \sigma_{\hat{\mathbf{n}}}^{-2} \sum_{\substack{j_k=0 \\ k=1, ..., N}}^{r_k-1} \int_{|U(1, \mathbf{n}, \mathbf{j})| \geq \varepsilon\sigma_{\hat{\mathbf{n}}}} [U(1, \mathbf{n}, \mathbf{j})]^2 \, dP \to 0, \tag{A.20}$$

$$E\left(\frac{\sum_{i=2}^{2^N} T(\mathbf{n}, i)}{\sigma_{\hat{\mathbf{n}}}}\right)^2 \to 0. \tag{A.21}$$

For the moment, assume that (A.18)–(A.21) hold. Denote

$$s_{\mathbf{n}}^2 \equiv \sum_{\substack{j_k=0 \\ k=1, ..., N}}^{r_k-1} E[U(1, \mathbf{n}, \mathbf{j})]^2. \tag{A.22}$$

Relations (A.18)–(A.19) imply that $T(\mathbf{n}, 1)/\sigma_{\hat{\mathbf{n}}}$ has the same limiting distribution as

$$s_{\mathbf{n}}^{-2} \sum_{\substack{j_k=0 \\ k=1,\ldots,N}}^{r_k-1} U^*(1, \mathbf{n}, \mathbf{j}),$$

where the $U^*(1, \mathbf{n}, \mathbf{j})$'s are independent random variables with $U^*(1, \mathbf{n}, \mathbf{j})$ having the same distribution as $U(1, \mathbf{n}, \mathbf{j})$. By (A.18) and (A.20),

$$s_{\hat{\mathbf{n}}}^{-2} \sum_{\substack{j_k=0 \\ k=1,\ldots,N}}^{r_k-1} \int_{|U^*(1,\mathbf{n},\mathbf{j})| \geq \varepsilon s_{\mathbf{n}}} [U^*(1, \mathbf{n}, \mathbf{j})]^2 \, dP \to 0, \tag{A.23}$$

which implies that

$$s_{\mathbf{n}}^{-2} \sum_{\substack{j_k=0 \\ k=1,\ldots,N}}^{r_k-1} U^*(1, \mathbf{n}, \mathbf{j}) \xrightarrow{\mathscr{L}} N(0, 1), \tag{A.24}$$

completing the proof of (A.16). The proof of (A.17) follows from (A.21). It remains to prove (A.18)–(A.21).

*Proof of* (A.18). Relation (A.18) is equivalent to

$$\sigma_{\hat{\mathbf{n}}}^{-2} \hat{\mathbf{r}} E[U(1, \mathbf{n}, \mathbf{0})]^2 \to 1, \tag{A.25}$$

with

$$\hat{\mathbf{r}} \equiv r_1 r_2 \cdots r_N; \tag{A.26}$$

that is,

$$[\hat{\mathbf{n}} \pi_{\hat{\mathbf{n}}} (1 - \pi_{\hat{\mathbf{n}}})]^{-1} \hat{\mathbf{r}} E[U(1, \mathbf{n}, \mathbf{0})]^2 \to 1, \tag{A.27}$$

or equivalently,

$$\frac{E[U(1, \mathbf{n}, \mathbf{0})]^2}{(\hat{\mathbf{n}}/\hat{\mathbf{r}}) \, \pi_{\hat{\mathbf{n}}} (1 - \pi_{\hat{\mathbf{n}}})} \to 1. \tag{A.28}$$

Using (A.6)

$$E[U(1, \mathbf{n}, \mathbf{0})]^2 = \operatorname{Var}\Bigg( \sum_{\substack{i_k=1 \\ k=1,\ldots,N}}^{p} \xi_{\mathbf{i}} \Bigg) = p^N \operatorname{Var} \xi_{\mathbf{i}} + Q, \tag{A.29}$$

where

$$Q \equiv \sum_{\substack{i_k=1 \\ k=1,\dots,N}}^{p} \sum_{\substack{j_k=1 \\ k=1,\dots,N}}^{p} \operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}}), \qquad i_k \neq j_k \text{ for some } 1 \leqslant k \leqslant N. \tag{A.30}$$

By (A.3)

$$\begin{aligned}\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}}) = {} & P\{H(\| X_{\mathbf{i}} - x \|) \leqslant \pi_{\hat{\mathbf{n}}}, H(\| X_{\mathbf{j}} - x \|) \leqslant \pi_{\hat{\mathbf{n}}}\} \\ & - P\{H(\| X_{\mathbf{i}} - x \|) \leqslant \pi_{\hat{\mathbf{n}}}\} P\{H(\| X_{\mathbf{j}} - x \|) \leqslant \pi_{\hat{\mathbf{n}}}\}. \end{aligned} \tag{A.31}$$

Since $f$ is positive and continuous at $x$, there exists an $r_0$ sufficiently small that

$$\inf_{x \in S(x, r_0)} f(x) > C_0 \tag{A.32}$$

for some constant $C_0 > 0$. For $r < r_0$,

$$H(r) = P\{\| X_1 - x \| \leqslant r\} = \int_{S(x,r)} f(x)\, dx > C_0 c r^d, \tag{A.33}$$

where $c$ is the colume of the sphere in $R^d$ with radius 1. Subsequently, for all $0 \leqslant a < H(r_0)$,

$$H^{-1}(a) \leqslant (a/(C_0 c))^{1/d}. \tag{A.34}$$

For $\hat{\mathbf{n}}$ sufficiently large that $\pi_{\hat{\mathbf{n}}} < H(r_0)$, we have by (A.31), (A.32), and (A.34),

$$\begin{aligned} & |\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}})| \\ & \leqslant \iint_{S(x, H^{-1}(\pi_{\hat{\mathbf{n}}})) \times S(x, H^{-1}(\pi_{\hat{\mathbf{n}}}))} | f_{\mathbf{i},\mathbf{j}}(u, v) - f(u) f(v)|\, du dv \\ & \leqslant \iint_{S(x, (\pi_{\hat{\mathbf{n}}}/(C_0 c))^{1/d}) \times S(x, (\pi_{\hat{\mathbf{n}}}/(C_0 c))^{1/d})} | f_{\mathbf{i},\mathbf{j}}(u, v) - f(u) f(v)|\, du\, dv. \end{aligned} \tag{A.35}$$

By Assumption 2, there exists an $r_1 > 0$ and a constant $C_1 > 0$ such that $| f_{\mathbf{i},\mathbf{j}}(u, v) - f(u) f(v)| < C_1$ for all $\mathbf{i} \neq \mathbf{j}$ and for $(u, v) \in S(x, r_1) \times S(x, r_1)$. By (A.35), for sufficiently large $\hat{\mathbf{n}}$ that $\pi_{\hat{\mathbf{n}}} < H(r_0)$ and, in addition, $(\pi_{\hat{\mathbf{n}}}/((C_0 c))^{1/d}) < r_1$,

$$|\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}})| \leqslant \iint_{S(x, (\pi_{\hat{\mathbf{n}}}/(C_0 c))^{1/d}) \times S(x, (\pi_{\hat{\mathbf{n}}}/(C_0 c))^{1/d})} C_1\, du\, dv \leqslant C \pi_{\hat{\mathbf{n}}}^2. \tag{A.36}$$

Let $\gamma = 2/(2+\delta)$, where $\delta$ is the constant of Assumption 5. By Lemma 2.1, for $\mathbf{i} \neq \mathbf{j}$,

$$|\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}})| \leqslant C(E \|\xi_{\mathbf{i}}\|^{2+\delta})^{2/(2+\delta)} \{\hat{f}(1,1)\, \varphi(\|\mathbf{j}-\mathbf{i}\|)\}^{\delta/(2+\delta)}$$
$$\leqslant C(E\|\xi_{\mathbf{i}}\|^{2/\gamma})^{\gamma} \{\varphi(\|\mathbf{j}-\mathbf{i}\|)\}^{1-\gamma} \tag{A.37}$$

for $\delta > 0$. Employing (A.3),

$$E\|\xi_{\mathbf{i}}\|^{2/\gamma} = (1-\pi_{\hat{\mathbf{n}}})^{2/\gamma}\, \pi_{\hat{\mathbf{n}}} + \pi_{\hat{\mathbf{n}}}^{2/\gamma}(1-\pi_{\hat{\mathbf{n}}}) \leqslant C\pi_{\hat{\mathbf{n}}}. \tag{A.38}$$

Combining (A.37) and (A.38),

$$|\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}})| \leqslant C\pi_{\hat{\mathbf{n}}}^{\gamma} \{\varphi(\|\mathbf{j}-\mathbf{i}\|)\}^{1-\gamma}. \tag{A.39}$$

Define

$$S_1 = \{(\mathbf{i},\mathbf{j}) : \mathbf{i} \neq \mathbf{j},\ 1 \leqslant i_k, j_k \leqslant p,\ k = 1, ..., N,\ \hat{d}(\mathbf{i},\mathbf{j}) \leqslant c_{\mathbf{n}}\} \tag{A.40}$$

and

$$S_2 = \{(\mathbf{i},\mathbf{j}) : \mathbf{i} \neq \mathbf{j},\ 1 \leqslant i_k, j_k \leqslant p,\ k = 1, ..., N,\ \hat{d}(\mathbf{i},\mathbf{j}) > c_{\mathbf{n}}\}, \tag{A.41}$$

where $c_{\mathbf{n}}$ is the sequence of positive numbers defined in Assumption 6.

By (A.31)

$$|Q| \leqslant Q_1 + Q_2, \tag{A.42}$$

where

$$Q_1 \equiv \sum_{S_1}\sum |\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}})| \tag{A.43}$$

and

$$Q_2 \equiv \sum_{S_2}\sum |\operatorname{Cov}(\xi_{\mathbf{i}}, \xi_{\mathbf{j}})|. \tag{A.44}$$

By (A.36),

$$Q_1 \leqslant C\pi_{\hat{\mathbf{n}}}^2 \sum_{S_1}\sum 1 \leqslant C\pi_{\hat{\mathbf{n}}}^2 p^N c_{\mathbf{n}}^N = Cp^N \pi_{\hat{\mathbf{n}}} \pi_{\hat{\mathbf{n}}} c_{\mathbf{n}}^N. \tag{A.45}$$

By Assumption 6(i),

$$(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})\, c_{\mathbf{n}}^N \to 0,$$

which by (A.2) implies that

$$\pi_{\hat{\mathbf{n}}} c_{\mathbf{n}}^N \to 0. \tag{A.46}$$

By (A.45) and (A.46),

$$Q_1 = o(p^N \pi_{\hat{\mathbf{n}}}). \tag{A.47}$$

Relation (A.2) gives $\pi_{\hat{\mathbf{n}}} \leqslant Ck(\hat{\mathbf{n}})/\hat{\mathbf{n}}$. Thus,

$$\hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}^{\gamma} \leqslant Cp^N \pi_{\hat{\mathbf{n}}} \hat{\mathbf{n}}(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{-1+\gamma} p^{-N} = Cp^N \pi_{\hat{\mathbf{n}}} \hat{\mathbf{n}}^{2-\gamma}(k(\hat{\mathbf{n}}))^{-1+\gamma} p^{-N}. \tag{A.48}$$

By (A.48) and Assumption 6(ii),

$$Q_2 \leqslant C\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{S_2}\sum \{\varphi(\|\mathbf{j}-\mathbf{i}\|)\}^{1-\gamma}$$

$$\leqslant C\hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{\|i\|>c_{\mathbf{n}}} \{\varphi(\|\mathbf{i}\|)\}^{1-\gamma} = o(p^N \pi_{\hat{\mathbf{n}}}). \tag{A.49}$$

*Proof of* (A.19). Denote

$$V(1, \mathbf{n}, \mathbf{j}) = U(1, \mathbf{n}, \mathbf{j})/\sigma_{\hat{\mathbf{n}}}.$$

Enumerate the rv's $V(1, \mathbf{n}, \mathbf{j})$ in an arbitrary manner and refer to them as $V_1, \ldots, V_M$. Note that $M = \prod_{k=1}^{N} r_k = \hat{\mathbf{n}}(p+q)^{-N} \leqslant \hat{\mathbf{n}}p^{-N}$. Let

$$I(1, \mathbf{n}, \mathbf{j}) = \{\mathbf{i} : j_k(p+q)+1 \leqslant i_k \leqslant j_k(p+q)+p\}. \tag{A.50}$$

Distinct sets of sites $I(1, \mathbf{n}, \mathbf{j})$ are far apart by a distance of at least $q$. Clearly $I(1, \mathbf{n}, \mathbf{j})$ contains $p^N$ sites and $I(1, \mathbf{n}, \mathbf{j})$ is the set of sites involved with $V(1, \mathbf{n}, \mathbf{j})$. Lemma 2.2 shows

$$R \leqslant \sum_{k=1}^{M-1} \sum_{j=k+1}^{M} \Bigg| E(\exp[iuV_k]-1)(\exp[iuV_j]-1) \prod_{s=j+1}^{M} \exp[iuV_s]$$
$$- E(\exp[iuV_k]-1)\, E(\exp[iuV_j]-1) \prod_{s=j+1}^{M} \exp[iuV_s]\Bigg|. \tag{A.51}$$

Let $\tilde{I}_j$ be the sets of sites involved with $V_j$. An application of Lemma 2.1(ii) gives

$$|E(\exp[iuV_k]-1)(\exp[iuV_j]-1) - E(\exp[iuV_k]-1)\, E(\exp[iuV_j]-1)|$$
$$\leqslant C\varphi(\hat{d}(\tilde{I}_j, \tilde{I}_k))\, p^N.$$

Thus

$$R \leqslant Cp^N \sum_{k=1}^{M-1} \sum_{j=k+1}^{M} \varphi(\hat{d}(\tilde{I}_j, \tilde{I}_k)) \leqslant Cp^N M \sum_{k=2}^{M} \varphi(\hat{d}(\tilde{I}_1, \tilde{I}_k))$$

$$\leqslant Cp^N M \sum_{i=1}^{\infty} \sum_{k:\, iq \leqslant \hat{d}(\tilde{I}_1, \tilde{I}_k) < (i+1)q} \varphi(\hat{d}(\tilde{I}_1, \tilde{I}_k))$$

$$\leqslant Cp^N M \sum_{i=1}^{\infty} i^{N-1}\varphi(iq) \leqslant C\hat{\mathbf{n}} \sum_{i=1}^{\infty} i^{N-1}\varphi(iq), \tag{A.52}$$

which tends to zero by Assumption 4(ii).

*Proof of* (A.20). Clearly,

$$|U(1, \mathbf{n}, \mathbf{j})| \leqslant p^N. \tag{A.53}$$

Using (A.2) and the definition of $\sigma_{\hat{\mathbf{n}}}$,

$$\sigma_{\hat{\mathbf{n}}}^2 / k(\hat{\mathbf{n}}) \to 1 \qquad \text{as} \quad \mathbf{n} \to \infty. \tag{A.54}$$

By (A.53), (A.54), and Assumption 4(i), for sufficiently large $\hat{\mathbf{n}}$,

$$P\{|U(1, \mathbf{n}, \mathbf{j})| \geqslant \varepsilon \sigma_{\hat{\mathbf{n}}}\} = 0,$$

from which (A.20) follows.

*Proof of* (A.21). It is enough to show that

$$\sigma_{\hat{\mathbf{n}}}^{-2} E[T(\mathbf{n}, i)]^2 \to 0 \qquad \text{for each} \quad 2 \leqslant i \leqslant 2^N.$$

Without loss of generality, consider $E[T(\mathbf{n}, 2)]^2$. Enumerate the rv's $U(2, \mathbf{n}, \mathbf{j})$ in an arbitrary manner and refer to them as $\hat{U}_1, \ldots, \hat{U}_M$. Now

$$E[T(\mathbf{n}, 2)]^2 = \sum_{i=0}^{M} \operatorname{var}(\hat{U}_i) + 2 \sum_{i=1}^{M} \sum_{\substack{j=1 \\ i>j}}^{M} \operatorname{cov}(\hat{U}_i, \hat{U}_j) \equiv A_1 + A_2. \tag{A.55}$$

Since $X_{\mathbf{n}}$ is stationary,

$$\operatorname{var}(\hat{U}_i) = \operatorname{var}\Bigg( \sum_{\substack{i_k=1 \\ k=1,\ldots,N-1}}^{p} \sum_{i_N=1}^{q} \xi_{\mathbf{i}} \Bigg)^2 = p^{N-1} q \operatorname{Var} \xi_{\mathbf{i}}$$

$$+ \sum_{\substack{j_k=1 \\ k=1,\ldots,N-1}}^{p} \sum_{j_N=1}^{q} \sum_{\substack{i_k=1 \\ k=1,\ldots,N-1}}^{p} \sum_{i_N=1}^{q} \operatorname{Cov}\{\xi_{\mathbf{i}}, \xi_{\mathbf{j}}\},$$

$$i_k \neq j_k \text{ for some } 1 \leqslant k \leqslant N.$$

Now

$$\operatorname{Var} \xi_{\mathbf{i}} \leqslant C \pi_{\hat{\mathbf{n}}}.$$

By (A.56) and (A.39),

$$\operatorname{var}(\hat{U}_i) \leqslant C p^{N-1} q \Bigg( \pi_{\hat{\mathbf{n}}} + \sum_{\substack{i_k=1 \\ k=1,\ldots,N-1}}^{p} \sum_{i_N=1}^{q} \pi_{\hat{\mathbf{n}}}^{\gamma} \{\varphi(\|\mathbf{i}\|)\}^{1-\gamma} \Bigg)$$

$$\leqslant C p^{N-1} q \pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{\substack{i_k=1 \\ k=1,\ldots,N-1}}^{p} \sum_{i_N=1}^{q} \{\varphi(\|\mathbf{i}\|)\}^{1-\gamma}. \tag{A.57}$$

Using the definition of $A_1$ and (A.57),

$$A_1 \leqslant CMp^{N-1} q\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^{1-\gamma}. \tag{A.58}$$

Let

$$\begin{aligned} I(2, n, x, \mathbf{j}) = \{\mathbf{i} : j_k(p+q)+1 \leqslant i_k \leqslant j_k(p+q)+p,\ 1 \leqslant k \leqslant N-1, \\ j_N(p+q)+p+1 \leqslant i_N \leqslant (j_N+1)(p+q)\}. \end{aligned}$$

Then $U(2, \mathbf{n}, x, \mathbf{j})$ is the sum of $\xi_{\mathbf{i}}$ with sites in $I(2, n, x, \mathbf{j})$. Since $p > q$, if $\mathbf{j}$ and $\mathbf{j}'$ belong to two distinct sets $I(2, \mathbf{n}, x, \mathbf{j})$ and $I(2, \mathbf{n}, x, \mathbf{j}')$, then $j_k \neq j'_k$ for some $1 \leqslant k \leqslant N$ and $\|\mathbf{j}-\mathbf{j}'\| > q$. We obtain

$$\begin{aligned} |A_2| &\leqslant C \sum_{\substack{j_k=1 \\ k=1,\dots,N}}^{n_k} \sum_{\substack{i_k=1 \\ k=1,\dots,N}}^{n_k} |\mathrm{Cov}\{\xi_{\mathbf{i}}, \xi_{\mathbf{j}}\}|, \qquad \|\mathbf{i}-\mathbf{j}\| > q \\ &\leqslant C\hat{\mathbf{n}} \sum_{\substack{i_k=1 \\ k=1,\dots,N \\ \|\mathbf{i}\|>q}}^{n_k} \pi_{\hat{\mathbf{n}}}^{\gamma} \{\varphi(\|\mathbf{i}\|)\}^{1-\gamma} \leqslant C\hat{\mathbf{n}}\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{i=q}^{\infty} i^{N-1}(\varphi(i))^{1-\gamma}. \end{aligned} \tag{A.59}$$

By (4.55) and (A.59),

$$\begin{aligned} \sigma_{\hat{\mathbf{n}}}^{-2} E[T(\mathbf{n}, x, 2)]^2 \leqslant{}& CMp^{N-1} q\sigma_{\hat{\mathbf{n}}}^{-2}\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^{1-\gamma} \\ &+ C\hat{\mathbf{n}}\sigma_{\hat{\mathbf{n}}}^{-2}\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{i=q}^{\infty} i^{N-1}(\varphi(i))^{1-\gamma}. \end{aligned} \tag{A.60}$$

Next

$$\begin{aligned} Mp^{N-1} q\sigma_{\hat{\mathbf{n}}}^{-2}\pi_{\hat{\mathbf{n}}}^{\gamma} &= \hat{\mathbf{n}}(p+q)^{-N} p^{N-1} q\sigma_{\hat{\mathbf{n}}}^{-2}\pi_{\hat{\mathbf{n}}}^{\gamma} \leqslant \hat{\mathbf{n}}(q/p)\,\sigma_{\hat{\mathbf{n}}}^{-2}\pi_{\hat{\mathbf{n}}}^{\gamma} \\ &\leqslant C\hat{\mathbf{n}}(q/p)(\hat{\mathbf{n}}\pi_{\mathbf{n}})^{-1}\pi_{\hat{\mathbf{n}}}^{\gamma} \leqslant C(q/p)\,\pi_{\hat{\mathbf{n}}}^{\gamma-1} \\ &\leqslant C(q/p)(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{\gamma-1}. \end{aligned} \tag{A.61}$$

Finally,

$$\hat{\mathbf{n}}\sigma_{\hat{\mathbf{n}}}^{-2}\pi_{\hat{\mathbf{n}}}^{\gamma} \sum_{i=q}^{\infty} i^{N-1}(\varphi(i))^{1-\gamma} \leqslant C(k(\hat{\mathbf{n}})/\hat{\mathbf{n}})^{\gamma-1} \sum_{i=q}^{\infty} i^{N-1}(\varphi(i))^{1-\gamma} \tag{A.62}$$

tends to zero by Assumption 5(i).

*Proof of Lemma* 2.3. Observe that

$$\{|f_{\mathbf{n}}(x)-f(x)|>\varepsilon\}=\left\{\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}V_{\Gamma_{\mathbf{n}}(x)}}<f(x)-\varepsilon\right\}\cup\left\{\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}V_{\Gamma_{\mathbf{n}}(x)}}>f(x)+\varepsilon\right\}. \tag{A.63}$$

By (A.64), if $\varepsilon>f(x)$, then

$$\{|f_{\mathbf{n}}(x)-f(x)|>\varepsilon\}=\left\{V_{\Gamma_{\mathbf{n}}}(x)<\frac{(k(\hat{\mathbf{n}}))}{\hat{\mathbf{n}}(f(x)+\varepsilon)}\right\} \tag{A.64}$$

and, if $0<\varepsilon\leqslant f(x)$, then

$$\{|f_{\mathbf{n}}(x)-f(x)|>\varepsilon\}$$
$$=\left\{V_{\Gamma_{\mathbf{n}}}(x)>\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(f(x)-\varepsilon)}\right\}\cup\left\{V_{\Gamma_{\mathbf{n}}}(x)<\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(f(x)+\varepsilon)}\right\}. \tag{A.65}$$

Thus the proof of the lemma is completed if we show that

(i) For any $\varepsilon>0$ such that $0<\varepsilon\leqslant f(x)$,

$$P\left\{V_{\Gamma_{\mathbf{n}}}(x)>\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(f(x)-\varepsilon)}\right\}\to 0; \tag{A.66}$$

(ii) For any $\varepsilon>0$,

$$P\left\{V_{\Gamma_{\mathbf{n}}}(x)<\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(f(x)+\varepsilon)}\right\}\to 0. \tag{A.67}$$

Consider case (i). We have

$$\left\{V_{\Gamma_{\mathbf{n}}}(x)>\frac{k(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(f(x)-\varepsilon)}\right\}, \tag{A.68}$$

where $c$ is the volume of the sphere in $R^d$ with radius 1. Let

$$\delta_{\hat{\mathbf{n}}}=\left(\frac{(k(\hat{\mathbf{n}}))}{c\hat{\mathbf{n}}(f(x)-\varepsilon)}\right)^{1/d} \tag{A.69}$$

and

$$Y_{\mathbf{in}}=I(\|X_{\mathbf{i}}-x\|>\delta_{\hat{\mathbf{n}}}). \tag{A.70}$$

Denote

$$p_{\hat{\mathbf{n}}}\equiv P\{\|X_{\mathbf{i}}-x\|\leqslant\delta_{\hat{\mathbf{n}}}\}. \tag{A.71}$$

Recall that $c$ denotes the volume of a sphere in $R^d$ with radius 1. Let $\{a_{\mathbf{n}}\}$ and $\{b_{\mathbf{n}}\}$ be sequences of positive numbers. We write $a_{\mathbf{n}} \sim b_{\mathbf{n}}$ to denote that $a_{\mathbf{n}}/b_{\mathbf{n}}$ is bounded away from zero and infinity as $\mathbf{n} \to \infty$. Assumption 1 and (A.69) imply that

$$p_{\hat{\mathbf{n}}} = \int_{S(x,\,\delta_{\hat{\mathbf{n}}})} f(u)\,du = \int_{S(x,\,\delta_{\hat{\mathbf{n}}})} (f(x) - f(u) + f(x))\,du$$

$$= cf(x)\,\delta_{\hat{\mathbf{n}}}^d + O(\delta_{\hat{\mathbf{n}}}^{\delta+d}) \sim \delta_{\hat{\mathbf{n}}}^d \sim k(\hat{\mathbf{n}})/\hat{\mathbf{n}}. \tag{A.72}$$

By (A.69), (A.72), and since $\varepsilon < f(x)$, for sufficiently large $\hat{\mathbf{n}}$,

$$\hat{\mathbf{n}}p_{\hat{\mathbf{n}}} - k(\hat{\mathbf{n}}) = \hat{\mathbf{n}}cf(x)\,\delta_{\hat{\mathbf{n}}}^d + \hat{\mathbf{n}}O(\delta_{\hat{\mathbf{n}}}^{\rho+d}) - k(\hat{\mathbf{n}})$$

$$= \frac{f(x)(k(\hat{\mathbf{n}}))}{(f(x) - \varepsilon)} + o(k(\hat{\mathbf{n}})) - k(\hat{\mathbf{n}}) > k(\hat{\mathbf{n}}). \tag{A.73}$$

Write $Y_{\mathbf{i}\mathbf{n}}$ as $Y_{\mathbf{i}}$ for brevity. Claim that

$$\hat{V} \equiv \operatorname{Var}\Bigg( \sum_{\substack{i_k = 1 \\ k = 1, \ldots, N}}^{n_k} Y_{\mathbf{i}} \Bigg) \sim \hat{\mathbf{n}}p_{\hat{\mathbf{n}}}(1 - p_{\hat{\mathbf{n}}}). \tag{A.74}$$

For the moment, let us assume that (A.74) holds. By Chebyshev's inequality and (A.72)–(A.74), for sufficiently large $\hat{\mathbf{n}}$,

$$P\{\Gamma_{\mathbf{n}} > \delta_{\hat{\mathbf{n}}}\} \leqslant P\Bigg\{ \sum_{\substack{i_k = 1 \\ k = 1, \ldots, N}}^{n_k} Y_{\mathbf{i}} > \hat{\mathbf{n}} - k(\hat{\mathbf{n}}) \Bigg\}$$

$$\leqslant P\Bigg\{ \sum_{\substack{i_k = 1 \\ k = 1, \ldots, N}}^{n_k} Y_{\mathbf{i}} - \hat{\mathbf{n}}(1 - p_{\mathbf{n}}) > \hat{\mathbf{n}}p_{\mathbf{n}} - k(\hat{\mathbf{n}}) \Bigg\},$$

$$\leqslant (\hat{\mathbf{n}}p_{\hat{\mathbf{n}}} - k(\hat{\mathbf{n}}))^{-2}\,\hat{V} \leqslant C(\hat{\mathbf{n}}p_{\hat{\mathbf{n}}} - k(\hat{\mathbf{n}}))^{-2}\,\hat{\mathbf{n}}p_{\hat{\mathbf{n}}}$$

$$\leqslant C(k(\hat{\mathbf{n}}))^{-2}\,k(\hat{\mathbf{n}}) = C(k(\hat{\mathbf{n}}))^{-1}, \tag{A.75}$$

which tends to zero since $k(\hat{\mathbf{n}}) \to \infty$ by Assumption 3. The proof of (A.66) follows from (A.68) and (A.75).

It remains to prove (A.74). Clearly,

$$\hat{V} = \hat{\mathbf{n}}p_{\mathbf{n}}(1 - p_{\hat{\mathbf{n}}}) + \hat{Q}, \tag{A.76}$$

where

$$\hat{Q} \equiv \sum_{\substack{i_k = 1 \\ i = 1, \ldots, N}}^{n_k} \; \sum_{\substack{j_k = 1 \\ k = 1, \ldots, N}}^{n_k} \operatorname{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}}), \qquad \mathbf{i} \neq \mathbf{j}.$$

Clearly, $(\delta_{\hat{\mathbf{n}}}^d/p_{\hat{\mathbf{n}}}) \to 1/(\mathrm{c}f(x))$ since $\delta_{\hat{\mathbf{n}}} \to 0$. Thus, for large $\hat{\mathbf{n}}$, so that $\delta_{\hat{\mathbf{n}}} < \varepsilon$, by Assumption 2,

$$\begin{aligned} &|\mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}})| \\ &\quad = |P\{\|X_{\mathbf{i}} - x\| \leqslant \delta_{\hat{\mathbf{n}}}, \|X_{\mathbf{j}} - x\| \leqslant \delta_{\hat{\mathbf{n}}}\} \\ &\qquad - P\{\|X_{\mathbf{i}} - x\| \leqslant \delta_{\hat{\mathbf{n}}}\} P\{\|X_{\mathbf{j}} - x\| \leqslant \delta_{\hat{\mathbf{n}}}\}| \\ &\quad \leqslant \iint_{S(x, \delta_{\hat{\mathbf{n}}}) \times S(x, \delta_{\hat{\mathbf{n}}})} |f_{\mathbf{i},\mathbf{j}}(u, v) - f(u) f(v)| \, du\, dv \leqslant C\delta_{\hat{\mathbf{n}}}^{2d} \leqslant Cp_{\hat{\mathbf{n}}}^2. \end{aligned} \tag{A.77}$$

By Lemma 2.1,

$$|\mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}})| \leqslant C(E\,\|1 - Y_{\mathbf{i}}\|^{2/\gamma})^{\gamma}\, \{\hat{f}(1, 1)\, \varphi(\|\mathbf{j} - \mathbf{i}\|)\}^{1-\gamma}. \tag{A.78}$$

Since $1 - Y_{\mathbf{i}} = I(\|X_{\mathbf{i}} - x\| \leqslant \delta_{\hat{\mathbf{n}}})$,

$$E\,\|1 - Y_{\mathbf{i}}\|^{2/\gamma} = p_{\hat{\mathbf{n}}}. \tag{A.79}$$

Employing (A.78) and (A.79),

$$|\mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}})| \leqslant Cp_{\hat{\mathbf{n}}}^{\gamma}\{\varphi(\|\mathbf{j} - \mathbf{i}\|)\}^{1-\gamma}. \tag{A.80}$$

Define

$$\hat{S}_1 = \{\mathbf{i}, \mathbf{j} \in I_{\mathbf{n}} \mid \hat{d}(\mathbf{i}, \mathbf{j}) \leqslant c_{\mathbf{n}}\}, \qquad \hat{S}_2 = \{\mathbf{i}, \mathbf{j} \in I_{\mathbf{n}} \mid \hat{d}(\mathbf{i}, \mathbf{j}) > c_{\mathbf{n}}\}, \tag{A.81}$$

where $c_{\mathbf{n}}$ is the sequence of positive numbers given in Assumption 6. Note that

$$|\hat{Q}| \leqslant Q_1 + Q_2, \tag{A.82}$$

where

$$\hat{Q}_1 \equiv \sum_{\hat{S}_1}\sum |\mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}})|, \qquad \hat{Q}_2 \equiv \sum_{\hat{S}_2}\sum |\mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}})|. \tag{A.83}$$

By (A.77) and Assumption 6(i),

$$\hat{Q}_1 \leqslant Cp_{\hat{\mathbf{n}}}^2 \sum_{\hat{S}_1}\sum 1 \leqslant Cp_{\hat{\mathbf{n}}}^2\, \hat{\mathbf{n}} c_{\mathbf{n}}^N = C\hat{\mathbf{n}} p_{\hat{\mathbf{n}}}\, p_{\hat{\mathbf{n}}}\, c_{\mathbf{n}}^N = o(\hat{\mathbf{n}} p_{\hat{\mathbf{n}}}). \tag{A.84}$$

By (A.80) and the definition of $\hat{Q}_2$ in (A.83),

$$\hat{Q}_2 \leqslant Cp_{\hat{\mathbf{n}}}^{\gamma} \sum_{\hat{S}_2}\sum \{\varphi(\|\mathbf{j} - \mathbf{i}\|)\}^{1-\gamma}.$$

Recall that $p_{\hat{\mathbf{n}}} \sim (k(\hat{\mathbf{n}})/\hat{\mathbf{n}})$ by (A.72). Using Assumption 6(ii) and following the same line of argument as in (A.48) and (A.49), we have

$$\hat{Q}_2 = o(p^N p_{\hat{\mathbf{n}}}) = o(\hat{\mathbf{n}} p_{\hat{\mathbf{n}}}); \tag{A.85}$$

the last equality follows since $p^N = o(\hat{\mathbf{n}})$ by Assumptions 3 and 4(i).

## Acknowledgment

The authors thank the referee for many valuable comments.

## References

[1] Boente, G., and Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *J. Multivariate Anal.* **25** 90–99.

[2] Bolthausen, E. (1982). On the central limit theorem for stationary random fields. *Ann. Probab.* **10** 1047–1050.

[3] Deo, C. M. (1973). A note on empirical processes of strong mixing sequences. *Ann. Probab.* **1** 870–875.

[4] Devroye, L. P., and Wagner, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.* **5** 536–540.

[5] Franke, R. (1982). Scattered data interpolation: Tests of some methods. *Math. Comp.* **38** 181–199.

[6] Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6** 721–740.

[7] 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.

[8] Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblable conditionelle: Étude asymptotique et application au cas Markovien. In *Proceedings, 6th Franco-Belgian Meeting of Statisticians.*

[9] Ibragimov, I. A., and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[10] Ioannides, D. A., and Roussas, G. G. (1987). Note on the uniform convergence of density estimates for mixing random variables. *Statist. Probab. Lett.* **5** 279–285.

[11] Izenman, A. J. (1991). Recent developments in nonparametric density estimation. *J. Amer. Statist. Assoc.* **86** 205–224.

[12] Jain, A. K. (1988). *Fundamentals of Digital Image Processing.* Prentice-Hall, Englewood Cliffs, NJ.

[13] Kashyap, R. L., and Chellappa, R. (1983). Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Trans. Inform. Theory* **IT-20** 60–72.

[14] Laslett, G. M., McBratney, A. B., Pahl, P. H., and Hutchinson, F. (1987). Comparison of several spatial prediction methods for soil pH. *J. Soil Sci.* **38** 325–341.

[15] Levy, B. C., Adams, M. B., and Willsky, A. S. (1990). Solution and linear estimation of 2-D nearest neighbor models. *Proc. IEEE* **78** 627–641.

[16] Loftsgaarden, D. O., and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate probability density function. *Ann. Math. Statist.* **28** 1049–1051.

[17] MACK, Y. P. (1980). Asymptotic normality of multivariate *k-NN* density estimates. *Sankhya Ser. A* **42** 53–63.

[18] MACK, Y. P. (1981). Local properties of *k-NN* regression estimates. *SIAM J. Algo. Discrete Methods* **2** 311–323.

[19] MACK, Y. P., AND ROSENBLATT, M. (1979). Multivariate *k*-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15.

[20] MARROQUÍN, J. L., AND RAMÍREZ, A. (1991) Stochastic cellular automata and Gibbsian invariant measures. *IEEE Trans. Inform. Theory* **IT–37** 541–551.

[21] MASRY, E. (1983). Probability density estimation from sampled data. *IEEE Trans. Inform. Theory* **IT-29** 696–709.

[22] MASRY, E. (1986). Recursive probability density estimation from weakly dependent processes. *IEEE Trans. Inform. Theory* **IT–32** 254–267.

[23] MOORE, D. S., AND YACKEL, T. W. (1977). Large sample properties of nearest neighbor density estimators. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. S. Moore, Eds.). Academic Press, New York.

[24] NAHAPETIAN, B. S. (1980). The central limit theorem for random fields with mixing conditions. In *Adv. in Probability* 6, *Multicomponent Random Systems* (R. L. Dobrushin and Ya. G. Sinai, Eds.), pp. 531–548. Dekker, New York.

[25] NAKHAPETYAN, B. S. (1987). An approach to proving limit theorems for dependent random variables. *Theory Probab. Appl.* **32** 535–539.

[26] NEADERHAUSER, C. C. (1980). Convergence of block spins defined on random fields. *J. Statist Phys.* **22** 673–684.

[27] NGUYEN, H. T., AND TRAN, L. T. (1989). Nearest neighbor density estimator under serial dependence. *Publ. Inst. Statist. Univ. Paris* **34**, No. 1 69–93.

[28] PHAM, T. D. (1986). The mixing properties of bilinear and generalized random coefficient autoregressive models. *Stochastic Process. Appl.* **23** 291–300.

[29] PHAM, D. T., AND TRAN, L. T. (1985). Some strong mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303.

[30] RIPLEY, B. (1981). *Spatial Statistics*. Wiley, New York.

[31] ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Borkhauser, Boston.

[32] ROUSSAS, G. G. (1969). Nonparametric estimation of the transition distribution of a Markov process. *Ann. Inst. Statist. Math.* **21** 73–87.

[33] ROUSSAS, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *J. Statist. Plann. Inference* **18** 135–149.

[34] TAKAHATA, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. *Z. Wahrsch. Verw. Gebiete* **62** 477–480.

[35] TRAN, L. T. (1990). Kernel density estimation on random fields. *J. Multivariate Anal.* **34** 37–53.

[36] WEGMAN, E. J. (1972). Nonparametric probability density estimation. I. A summary of available methods. *Technometrics* **14** 533–546.

[37] WITHERS, C. S. (1981. Conditions for linear processes to be strong mixing. *Z. Wahrsch. Verw. Gebiete* **57** 477–480.

[38] YAKOWITZ, S. (1987). Nearest-neighbor methods for time series analysis. *J. Time Ser. Anal.* **8** 235–247.

[39] YAKOWITZ, S., GANI, J., AND HAYES, R. (1990). Random cellular automata models for epidemiology. *Appl. Math. Comput.* **40** 41–54.

[40] YAKOWITZ, S., HAYES, R., AND GANI, J. (1992). Automatic learning for dynamic Markov fields, with application to epidemiology. *Oper. Res.* to appear.

[41] YAKOWITZ, S., AND SZIDAROVSKY, F. (1985). A comparison of kriging with nonparametric regression methods. *J. Multivariate Anal.* **16** 21–53.