C. R. Acad. Sci. Paris, Ser. I 353 (2015) 635–639

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

A new spatial regression estimator in the multivariate context

Un nouvel estimateur de la fonction de régression spatiale pour données multivariées

Sophie Dabo-Niang [a,c], Camille Ternynck [a], Anne-Francoise Yao [b]

[a] *Laboratoire LEM, Université Lille-3, Maison de la recherche, BP 60149, 59653 Villeneuve d'Ascq cedex, France*
[b] *Laboratoire de Mathématiques, Université Blaise-Pascal, UMR 6620, CNRS, Campus des Cézeaux, BP 80026, 63171 Aubière cedex, France*
[c] *MODAL team, INRIA Lille-Nord de France, France*

ARTICLE INFO

Article history:
Received 30 January 2014
Accepted after revision 3 April 2015
Available online 29 April 2015

Presented by Paul Deheuvels

ABSTRACT

In this note, we propose a nonparametric spatial estimator of the regression function $x \to r(x) := \mathbb{E}[Y_{\mathbf{i}}|X_{\mathbf{i}} = x]$, $x \in \mathbb{R}^d$, of a stationary $(d+1)$-dimensional spatial process $\{(Y_{\mathbf{i}}, X_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^N\}$, at a point located at some station $\mathbf{j}$. The proposed estimator depends on two kernels in order to control both the distance between observations and the spatial locations. Almost complete convergence and consistency in L^q norm ($q \in \mathbb{N}^*$) of the kernel estimate are obtained when the sample considered is an α-mixing sequence.

RÉSUMÉ

Dans cette note, nous proposons un estimateur non paramétrique spatial de la fonction de régression $x \to r(x) := \mathbb{E}[Y_{\mathbf{i}}|X_{\mathbf{i}} = x]$, $x \in \mathbb{R}^d$, d'un champ stationnaire $\{(Y_{\mathbf{i}}, X_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^N\}$ de dimension $(d+1)$, à un point localisé à un site donné $\mathbf{j}$. L'estimateur proposé est composé de deux noyaux permettant de contrôler à la fois la distance entre les observations et entre les sites. La convergence presque complète ainsi que la convergence en moyenne d'ordre q (norme L^q) ($q \in \mathbb{N}^*$) de l'estimateur à noyaux sont obtenus en considérant des processus α-mélangeants.

1. Introduction

During the first half of the twentieth century, spatial statistics were mainly studied in the scope of geostatistics through the parametric framework. However, a preselected parametric model might be too restricted or too low-dimensional to fit unexpected features. Consequently, nowadays, a dynamic concerns the deployment of nonparametric methods to spatial statistics. In this note, we are interested in the nonparametric spatial regression estimation, which has received a great deal of attentional from the scientific community. Firstly, Biau and Cadre [1] dealt with the kernel prediction of a strictly

E-mail addresses: sophie.dabo@univ-lille3.fr (S. Dabo-Niang), ternynck.camille@gmail.com (C. Ternynck), Anne-francoise.Yao@math.univ-bpclermont.fr (A.-F. Yao).

http://dx.doi.org/10.1016/j.crma.2015.04.004
1631-073X/

stationary random field indexed in $(\mathbb{N}^*)^N$. Later, Dabo-Niang and Yao [6] were interested in the kernel regression estimation and prediction of continuously indexed random fields. In [11], nonparametric kernel prediction was considered for spatial stochastic processes when a stochastic sampling design is assumed for the selection of random locations. A main difference between them is that the last is based on a kernel that controls the distance between sites contrary to the others, which deal with a kernel on the values of the field.

More recently, Wang et al. [16] proposed a local linear spatio-temporal prediction model, using a kernel weight function taking into account the distance between sites. The specificity of the prediction procedure of Wang et al. [16] is to be based on the assumption that the error term of the model is autocorrelated, contrary to the present work.

Our proposed regression estimator takes advantages of each estimator introduced previously. In fact, it depends on two kernels, one of which controls the distance between observations and the other controls the spatial dependence structure. The advantage of the proposed estimate is to take directly into account the spatial dependency in its form, which is particularly interesting in a prevision context. This idea has been presented in [4] in the context of density estimation and in [15] to deal with a regression problem for functional data. The new kernel spatial estimator of the regression function is presented in Section 2. Then, in Section 3, the almost complete convergence and consistency in L^q norm ($q \in \mathbb{N}^*$) of the kernel estimate are obtained when the sample considered is an α-mixing sequence.

2. Kernel spatial estimator of the regression function

We consider a spatial process $(Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}}) \in \mathbb{R}^d \times \mathbb{R}, \mathbf{i} \in \mathbb{Z}^N)$ defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with same distribution as (X, Y) having unknown density $f_{X,Y}$ on $\mathbb{R}^{d+1}$. The density function of X on $\mathbb{R}^d$ is $f(\cdot)$. For the sake of simplicity, we will suppose that the variable Y is bounded. We are interested in the following regression model $Y_{\mathbf{i}} = r(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}$, where $r(\cdot) = \mathbb{E}(Y|X = x)$ is an unknown function, with real values, defined by $r(x) = \varphi(x)/f(x)$ where $\varphi(x) = \int y f_{XY}(x, y)\,\mathrm{d}y$, $x \in \mathbb{R}^d$, $(\varepsilon_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^N}$ is a centered spatial process independent of $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^N}$. The process is observed over the domain $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, \dots, i_N),\ 1 \le i_k \le n_k,\ k = 1, \dots, N\}$. We denote $\mathbf{n} = (n_1, \dots, n_N)$; let $\widehat{\mathbf{n}} := n_1 \times \dots \times n_N$ be the sample size. From now on, we assume for simplicity that $n_1 = n_2 = \dots = n_N = n$ (e.g., [7–9]) and write $\mathbf{n} \to \infty$ if $n \to \infty$, but the following results can be extended to a more general framework.

We are interested in the regression estimation of $r(\cdot)$, in particular the prediction of $Y_{\mathbf{j}}$ under the condition that $X_{\mathbf{j}} = x$ (as in [16]), which we denote in what follows $x_{\mathbf{j}}$; on the matter of the concerned location $\mathbf{j}$, see Remark 1. Considering normalized sites, the kernel estimator of $r(x_{\mathbf{j}})$, is defined as

$$r_{\mathbf{n}}(x_{\mathbf{j}}) = \frac{\varphi_{\mathbf{n}}(x_{\mathbf{j}})}{f_{\mathbf{n}}(x_{\mathbf{j}})} \quad \text{if } f_{\mathbf{n}}(x_{\mathbf{j}}) \neq 0; \qquad r_{\mathbf{n}}(x_{\mathbf{j}}) = \overline{Y} \quad \text{(empirical mean), otherwise,}$$

where

$$\varphi_{\mathbf{n}}(x_{\mathbf{j}}) = \frac{1}{a_{\mathbf{n},\mathbf{j}} b_{\mathbf{n}}^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(b_{\mathbf{n}}^{-1}(x_{\mathbf{j}} - X_{\mathbf{i}})\right) K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right),$$

$$f_{\mathbf{n}}(x_{\mathbf{j}}) = \frac{1}{a_{\mathbf{n},\mathbf{j}} b_{\mathbf{n}}^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} K_1\left(b_{\mathbf{n}}^{-1}(x_{\mathbf{j}} - X_{\mathbf{i}})\right) K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right),$$

with $a_{\mathbf{n},\mathbf{j}} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right)$. In addition, K_1 and K_2 are kernels respectively defined on $\mathbb{R}^d$ and $\mathbb{R}$, $b_{\mathbf{n}}$ and $\rho_{\mathbf{n}}$ are bandwidths tending to zero. Note that $K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right) = K_2\left(\frac{\|\mathbf{j}-\mathbf{i}\|}{n\rho_{\mathbf{n}}}\right)$, (where $\frac{\mathbf{i}}{\mathbf{n}} = (\frac{i_1}{n}, \frac{i_2}{n}, \dots, \frac{i_N}{n})$). For each site $\mathbf{j}$, the estimator $r_{\mathbf{n}}(x_{\mathbf{j}})$ is a function of the number $k_{\mathbf{n}} = k_{\mathbf{n},\mathbf{j}} = \sum_{\mathbf{i}} 1_{\|\mathbf{i}-\mathbf{j}\| \le d_{\mathbf{n}}}$ of neighbors sites $\mathbf{i}$, for which the distance between $\mathbf{i}$ and $\mathbf{j}$ is less or equal to distance $d_{\mathbf{n}} > 0$ such that $d_{\mathbf{n}} \to \infty$ as $\mathbf{n} \to \infty$. More precisely, in what follows, we assume that $k_{\mathbf{n}} = C_N d_{\mathbf{n}}^N(1 + o(1))$ as $d_{\mathbf{n}} \to \infty$ where C_N is a constant that depends on N. This is based on the problem of counting points with integer coordinates in the N-dimensional ball (see, e.g., [3]). In this work, $d_{\mathbf{n}}$ is chosen to be $n\rho_{\mathbf{n}}$ involving that $d_{\mathbf{n}}^N = \widehat{\mathbf{n}}\rho_{\mathbf{n}}^N$ and $k_{\mathbf{n}} = O(\widehat{\mathbf{n}}\rho_{\mathbf{n}}^N)$. We notice that the kernel K_2 is here to handle the nearness between locations.

Remark 1.

- As said above, we are particularly interested here in a spatial prediction methodology taking explicitly into account the spatial locations. Suppose one wants to predict $Y_{\mathbf{i}}$ in some unobserved location $\mathbf{j}$. More precisely, we suppose that the field $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^N}$ is observed on the set $\mathcal{O}_{\mathbf{n}}$ contained in $\mathcal{I}_{\mathbf{n}}$. The main purpose is to predict the unobserved value $Y_{\mathbf{j}}$ given $X_{\mathbf{j}}$ for a location $\mathbf{j} \in \mathcal{I}_{\mathbf{n}}$ but $\mathbf{j} \notin \mathcal{O}_{\mathbf{n}}$.
  To achieve the forecasting at the site $\mathbf{j}$, we propose to use the regression function estimator $r_{\mathbf{n}}(X_{\mathbf{j}})$. Then, the prediction of the value of the field $(Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^N}$ at the location $\mathbf{j} \notin \mathcal{O}_{\mathbf{n}}$ is written

$$\widehat{Y}_{\mathbf{j}} = r_{\mathbf{n}}(X_{\mathbf{j}}) = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{X_{\mathbf{j}}-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right) K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{X_{\mathbf{j}}-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right) K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right)}. \tag{1}$$

One can derive an asymptotic result such as almost complete convergence and consistency in L^q norm ($q \in \mathbb{N}^*$) for $\widehat{Y}_{\mathbf{j}}$ from the kernel regression estimate, given below.

– More generally, one can extend $\widehat{Y}_{\mathbf{j}}$ by considering $\widehat{Y}_{\mathbf{j}} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{X_{\mathbf{j}}-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right) K_2\left(\frac{\mathbf{i}-\mathbf{j}}{\rho_{\mathbf{n}}}\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{X_{\mathbf{j}}-X_{\mathbf{i}}}{b_{\mathbf{n}}}\right) K_2\left(\frac{\mathbf{i}-\mathbf{j}}{\rho_{\mathbf{n}}}\right)}$, where sites $\mathbf{i}$ and $\mathbf{j}$ are not normalized and $K_2(\cdot)$ is a kernel on $\mathbb{R}^N$.

3. Assumptions and results

To take into account the spatial dependency, we assume that the process $(Z_{\mathbf{i}})$ satisfies the following α-mixing condition: there exists a function $\varphi(x) \searrow 0$ as $x \to \infty$, such that $\alpha(\sigma(S), \sigma(S')) \le \psi(\text{Card}(S), \text{Card}(S'))\varphi(\text{dist}(S, S'))$, where S and S' are two finite sets of sites, Card(S) denotes the cardinality of S, $\sigma(S) = \{Z_{\mathbf{i}}, \mathbf{i} \in S\}$ denotes a σ-*field* generated by $Z_{\mathbf{i}}$, dist$(\cdot,\cdot)$ is the Euclidean distance, $\psi(\cdot)$ is a positive symmetric function nondecreasing in each variable. We will assume that $\varphi(i)$ tends to zero at a polynomial rate, i.e. $\varphi(i) \le Ci^{-\theta}$. Let $u_{\mathbf{n}} = \prod_{i=1}^{N} (\log n_i)(\log\log n_i)^{1+\epsilon}$, then $\sum_{\mathbf{n}\in\mathbb{N}^N} \frac{1}{\widehat{\mathbf{n}} u_{\mathbf{n}}} < \infty$. Some consistency results are obtained under the following assumptions:

A1: *the density functions $f_{X,Y}$ and f are continuous on $\mathbb{R}^{d+1}$ and $\mathbb{R}^d$, respectively;*

A2: *the density and the regression functions satisfy the Lipschitz condition, thus*

$$|f(x) - f(y)| \le C\|x - y\| \quad \textit{and} \quad |r(x) - r(y)| \le C\|x - y\|, \qquad \forall x, y \in \mathbb{R}^d;$$

A3: *the functions $K_1(\cdot)$ and $K_2(\cdot)$ are bounded integrable kernels on $\mathbb{R}$. Moreover, the kernel $K_1(\cdot)$ satisfy some Lipschitz condition;*

A4: *there exist some constants C_{1i} and C_{2i} with $0 < C_{1i} < C_{2i} < \infty$, for $i = 1, 2$, such that*

$$C_{11}\mathbf{1}_{[0,1]}(s's) \le K_1(s) \le C_{21}\mathbf{1}_{[0,1]}(s's) \qquad \textit{for } s \in \mathbb{R}^d,$$
$$C_{12}\mathbf{1}_{[0,1]}(t) \le K_2(t) \le C_{22}\mathbf{1}_{[0,1]}(t) \qquad \textit{for } t \in \mathbb{R},$$

where s' denotes the transpose of s;

A5: ***Local dependence condition.*** *The joint probability density $f_{X_{\mathbf{i}},X_{\mathbf{j}}}$ of $(X_{\mathbf{i}}, X_{\mathbf{j}})$ exists, is bounded and $\forall u, v \in \mathbb{R}^d$, for some constant $C > 0$, verifies*

$$|f_{X_{\mathbf{i}},X_{\mathbf{j}}}(u, v) - f_{X_{\mathbf{i}}}(u) f_{X_{\mathbf{j}}}(v)| < C;$$

A6: *$\psi(n,m) \le C\min(n,m)$ and $\widehat{\mathbf{n}} b_{\mathbf{n}}^{d\theta_1} \rho_{\mathbf{n}}^{N\theta_1} \log \widehat{\mathbf{n}}^{\theta_2} u_{\mathbf{n}}^{\theta_3} \to \infty$ with the mixing coefficient $\theta > N(q+2)$, $q > 1$ and*

$$\theta_1 = \frac{qN - \theta}{N(q+2) - \theta} > 0; \quad \theta_2 = \frac{\theta - 2N}{N(q+2) - \theta} < 0; \quad \theta_3 = \frac{2N}{N(q+2) - \theta} < 0;$$

A7: *$\psi(n,m) \le C(n+m+1)^{\tilde{\beta}}$ and $\widehat{\mathbf{n}} b_{\mathbf{n}}^{d\theta'_1} \rho_{\mathbf{n}}^{N\theta'_1} \log \widehat{\mathbf{n}}^{\theta'_2} u_{\mathbf{n}}^{\theta'_3} \to \infty$ with the mixing coefficient $\theta > N(q + 2\tilde{\beta} + 1)$, $q > 1$, $\tilde{\beta} > 1$ and*

$$\theta'_1 = \frac{N(q-1) - \theta}{N(q+2\tilde{\beta}+1) - \theta} > 0; \quad \theta'_2 = \frac{\theta - N}{N(q+2\tilde{\beta}+1) - \theta} < 0; \quad \theta'_3 = \frac{2N}{N(q+2\tilde{\beta}+1) - \theta} < 0.$$

Remarks. These assumptions are classically used in spatial nonparametric modeling.

– The assumptions **A2** and **A3** allow us to control the bias of the estimator. The Lipschitz condition **A2** allows the precise rate of convergence to be found, whereas a continuity-type model would give only convergence results.
– Assumption **A4** is imposed for the sake of simplicity and brevity of the proofs. We will use Assumption **A4** both to control the bias and the distances between sites. This condition is verified, for example, if K_2 is defined by $K_2(t) = \mathbf{1}_{[0,1]}(t)$ or any function defined as $K_2(t) = (u(t))\mathbf{1}_{[0,1]}(t)$ where u is a non-increasing function such that $u(1) > 0$.
– The **local dependence condition (A5)** is a classical condition in kernel estimation based on dependent data (see, e.g., [2]). The difference between this condition and the mixing condition is: condition **A5** controls the dependency through the distance between $f_{X_{\mathbf{i}}X_{\mathbf{j}}}$ and $f_{X_{\mathbf{i}}} f_{X_{\mathbf{j}}}$ when the mixing condition controls the dependency through the distance between $P(A \cap B)$ and $P(A)P(B)$ (as previously defined). Naturally, both conditions are linked. The link between them can be found, for example, in [6]. Like the mixing condition, condition **A5** is used to control the variance term of the estimation.
– The assumptions **A6** and **A7** are classical technical assumptions that appear (in the calculations when studying the asymptotic behavior of the estimator) in the particular case where the mixing coefficient is such that $\varphi(i)$ verifies: $\varphi(i) \le Ci^{-\theta}$, for some $\theta > 0$ (see [12] and [13] for some examples). Each of these conditions is related to a specific case of mixing in the spatial context and are used respectively in [12] and [14].

The two following theorems give some results about the consistency of the estimator proposed for the regression function.

Theorem 3.1. *Under Assumptions **A1–A5** and **A6** or **A7**, $r_{\mathbf{n}}(x_{\mathbf{j}})$ converges almost completely (a.c.) to $r(x_{\mathbf{j}})$ and*

$$|r_{\mathbf{n}}(x_{\mathbf{j}})-r(x_{\mathbf{j}})|=O\left(b_{\mathbf{n}}+\sqrt{\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}\,b_{\mathbf{n}}^{d}\,\rho_{\mathbf{n}}^{N}}}\right)\ a.c.$$

Pattern of the proof. We write

$$|r_{\mathbf{n}}(x_{\mathbf{j}})-r(x_{\mathbf{j}})|\leq\left(\frac{1}{f_{\mathbf{n}}(x_{\mathbf{j}})}|\varphi_{\mathbf{n}}(x_{\mathbf{j}})-\varphi(x_{\mathbf{j}})|+\frac{\varphi(x_{\mathbf{j}})}{f_{\mathbf{n}}(x_{\mathbf{j}})f(x_{\mathbf{j}})}|f_{\mathbf{n}}(x_{\mathbf{j}})+f(x_{\mathbf{j}})|\right)\mathbf{1}_{[\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}W_{\mathbf{ni}}\neq 0]}+\overline{Y}\mathbf{1}_{[\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}W_{\mathbf{ni}}=0]},\qquad(2)$$

where $\overline{Y}$ is the empirical mean of the $Y_{\mathbf{i}}$.

We study the term $|f_{\mathbf{n}}(x_{\mathbf{j}})-f(x_{\mathbf{j}})|$ since it is a particular case of $|\varphi_{\mathbf{n}}(x_{\mathbf{j}})-\varphi(x_{\mathbf{j}})|$ when $Y_{\mathbf{i}}$ is equal to 1. The result is obtained studying separately the bias and the variance terms. It is easy to show that the bias $|\mathbb{E}(f_{\mathbf{n}}(x_{\mathbf{j}}))-f(x_{\mathbf{j}})|=O(b_{\mathbf{n}})$. For the variance, an adjustment of Lemma 3.2 in [6] is used to obtain that

$$\begin{aligned}P&=\mathbb{P}(|f_{\mathbf{n}}(x_{\mathbf{j}})-\mathbb{E}(f_{\mathbf{n}}(x_{\mathbf{j}}))|>\epsilon)\\&\leq C_N\widehat{\mathbf{n}}^{-a}+2^{N+2}\frac{C}{a_{\mathbf{n},\mathbf{j}}b_{\mathbf{n}}^{d}}\psi([\widehat{\mathbf{t}}-1]p^N,p^N)\varphi(p)\widehat{\mathbf{n}}\epsilon^{-1}\end{aligned}$$

with $a=\dfrac{\delta^2}{2^{2N+4}C+2^{N+2}C_N\delta}$ and $\epsilon=\delta\left(\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}b_{\mathbf{n}}^{d}\rho_{\mathbf{n}}^{N}}\right)^{1/2}$, $\delta>0$, and $p=\left(\frac{\widehat{\mathbf{n}}b_{\mathbf{n}}^{d}\rho_{\mathbf{n}}^{N}}{\log\widehat{\mathbf{n}}}\right)^{\frac{1}{2N}}$. In both assumptions on $\psi(n,m)$ (**A6** and **A7**) and by appropriate choice of $\delta>2^{N+1}C_N$, the bound of $\sum_{\mathbf{n}=(n_1,\dots,n_N)\in\mathbb{N}^N}P$ is the general term of a convergent series.

We have

$$P\left(\left[\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}W_{\mathbf{ni}}=0\right]\right)\leq\mathbb{P}\left[|f_{\mathbf{n}}(x_{\mathbf{j}})-\mathbb{E}[f_{\mathbf{n}}(x_{\mathbf{j}})]|>\epsilon\right]\qquad\text{for }\mathbf{n}\text{ large enough.}$$

So the last term of (2) is a.c. zero for large $\mathbf{n}$.

Theorem 3.2. *Under Assumptions **A1–A5** and **A6** or **A7**, $r_{\mathbf{n}}(x_{\mathbf{j}})$ converges in mean of order q to $r(x_{\mathbf{j}})$ and*

$$\|r_{\mathbf{n}}(x_{\mathbf{j}})-r(x_{\mathbf{j}})\|_q\;=\;O\left(b_{\mathbf{n}}+\sqrt{\frac{1}{\widehat{\mathbf{n}}b_{\mathbf{n}}^{d}\rho_{\mathbf{n}}^{N}}}\right),\ q>1.$$

Pattern of the proof. Let $W_{\mathbf{ni}}=\dfrac{K_1\left(b_{\mathbf{n}}^{-1}(x_{\mathbf{j}}-X_{\mathbf{i}})\right)K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}K_1\left(b_{\mathbf{n}}^{-1}(x_{\mathbf{j}}-X_{\mathbf{i}})\right)K_2\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{j}-\mathbf{i}}{\mathbf{n}}\right\|\right)}$ and by adopting the convention $0/0=0$, we have $\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}W_{\mathbf{ni}}=0$ or 1. Consequently, we can deal with the following decomposition:

$$\begin{aligned}\|r_{\mathbf{n}}(x_{\mathbf{j}})-r(x_{\mathbf{j}})\|_q\leq\mathbb{E}^{1/q}&\left[\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}W_{\mathbf{ni}}\left[\mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})-r(x_{\mathbf{j}})\right]\right)\mathbf{1}_{[\sum_{\mathbf{i}}W_{\mathbf{ni}}=1]}\right]^q\\&+\mathbb{E}^{1/q}\left[\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}W_{\mathbf{ni}}\left[Y_{\mathbf{i}}-\mathbb{E}(Y_{\mathbf{i}}|X_{\mathbf{i}})\right]\right)\mathbf{1}_{[\sum_{\mathbf{i}}W_{\mathbf{ni}}=1]}\right]^q\\&+\mathbb{E}^{1/q}\left[\left(\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}Y_{\mathbf{i}}-r(x_{\mathbf{j}})\right)\mathbf{1}_{[\sum_{\mathbf{i}}W_{\mathbf{ni}}=0]}\right]^q.\end{aligned}$$

The study of the three terms of the right-hand-side gives the following result $\mathbb{E}^{1/q}\left[r_{\mathbf{n}}(x_{\mathbf{j}})-r(x_{\mathbf{j}})\right]^q=O(b_{\mathbf{n}}^{d})+O\left(\left(\widehat{\mathbf{n}}b_{\mathbf{n}}^{d}\rho_{\mathbf{n}}^{N}\right)^{-1/2}\right)+O\left(\left(\widehat{\mathbf{n}}b_{\mathbf{n}}^{d}\rho_{\mathbf{n}}^{N}\right)^{-1/2}\right)$ obtained by applying Lemma 2.2 in [10].

Details of the proofs are provided in [5].

Acknowledgements

The authors wish to thank the anonymous reviewer for his/her useful comments.

References

[1] G. Biau, B. Cadre, Nonparametric spatial prediction, Stat. Inference Stoch. Process. 7 (3) (2004) 327–349.
[2] M. Carbon, L.T. Tran, B. Wu, Kernel density estimation for random fields (density estimation for random fields), Stat. Probab. Lett. 36 (2) (1997) 115–125.
[3] L.F. Chamizo, H. Iwaniec, On the sphere problem, Rev. Mat. Iberoam. 11 (2) (1995) 417–430.
[4] S. Dabo-Niang, L. Hamdad, C. Ternynck, A.-F. Yao, A kernel spatial density estimation allowing for the analysis of spatial clustering: application to Monsoon Asia Drought Atlas data, Stoch. Environ. Res. Risk Assess. 28 (2014) 2075–2099.
[5] S. Dabo-Niang, C. Ternynck, A.-F. Yao, Nonparametric prediction of spatial multivariate data, 2015, preprint.
[6] S. Dabo-Niang, A.-F. Yao, Kernel regression estimation for continuous spatial processes, Math. Methods Stat. 16 (4) (2007) 298–317.
[7] M. El Machkouri, Nonparametric regression estimation for random fields in a fixed-design, Stat. Inference Stoch. Process. 10 (1) (2007) 29–47.
[8] M. El Machkouri, Asymptotic normality of the Parzen–Rosenblatt density estimator for strongly mixing random fields, Stat. Inference Stoch. Process. 14 (1) (2011) 73–84.
[9] M. El Machkouri, R. Stoica, Asymptotic normality of kernel estimates in a regression model for random fields, J. Nonparametr. Stat. 22 (8) (2010) 955–971.
[10] J. Gao, Z. Lu, D. Tjøstheim, Moment inequalities for spatial processes, Stat. Probab. Lett. 78 (6) (2008) 687–697.
[11] R. Menezes, P. García-Soidán, C. Ferreira, Nonparametric spatial prediction under stochastic sampling design, J. Nonparametr. Stat. 22 (3) (2010) 363–377.
[12] C.C. Neaderhouser, Convergence of block spins defined by a random field, J. Stat. Phys. 22 (6) (1980) 673–684.
[13] M. Rosenblatt, Stationary Sequences and Random Fields, Birkhäuser, Boston, 1985.
[14] H. Takahata, On the rates in the central limit theorem for weakly dependent random fields, Z. Wahrscheinlichkeitstheor. Verw. Geb. 64 (4) (1983) 445–456.
[15] C. Ternynck, Spatial regression estimation for functional data with spatial dependency, J. Soc. Fr. Stat. 155 (2) (2014) 138–160.
[16] H. Wang, J. Wang, B. Huang, Prediction for spatio-temporal models with autoregression in errors, J. Nonparametr. Stat. 24 (1) (2012) 217–244.