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\textsuperscript{a,c}, Camille Ternynck\textsuperscript{a}, Anne-Francoise Yao\textsuperscript{b}

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

A R T I C L E   I N F O

Article history:
Accepted 30 January 2014
Accepted after revision 3 April 2015
Available online 29 April 2015
Presented by Paul Deheuvels

A B S T R A C T

In this note, we propose a nonparametric spatial estimator of the regression function $x \rightarrow r(x) := \mathbb{E}[Y_i | X_i = x]$, $x \in \mathbb{R}^d$, of a stationary $(d + 1)$-dimensional spatial process $\{(Y_i, X_i), i \in \mathbb{Z}^N\}$, at a point located at some station $j$. 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 $\alpha$-mixing sequence.

© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

Dans cette note, nous proposons un estimateur non paramétrique spatial de la fonction de régression $x \rightarrow r(x) := \mathbb{E}[Y_i | X_i = x]$, $x \in \mathbb{R}^d$, d’un champ stationnaire $\{(Y_i, X_i), i \in \mathbb{Z}^N\}$ de dimension $(d + 1)$, à un point localisé à un site donné $j$. 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 $\alpha$-mélangeants.

© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

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 \cite{biau2015} 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/© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
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 $\alpha$-mixing sequence.

2. Kernel spatial estimator of the regression function

We consider a spatial process $(Z_i = (X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}, i \in \mathbb{Z}^N)$ 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 given. We are interested in the following regression model $Y_i = r(X_i) + e_i$, where $r(\cdot) = \mathbb{E}(Y|X = x)$ is an unknown function, with real values, defined by $r(x) = \psi(x)/\rho(x)$ where $\psi(x) = \int y f_{X,Y}(x, y) \, dy, x \in \mathbb{R}^d, (e_i)_{i \in \mathbb{Z}^N}$ is a centered spatial process independent of $(X_i)_{i \in \mathbb{Z}^N}$. The process is observed over the domain $\mathcal{I}_n = \{ i = (i_1, \ldots, i_N) \mid 1 \leq i_k \leq n_k, k = 1, \ldots, N \}$. We denote $\mathbf{n} = (n_1, \ldots, n_N)$; let $\mathbf{n} := n_1 \times \ldots \times n_N$ be the sample size. From now on, we assume for simplicity that $n_1 = n_2 = \ldots = n_N = n$ (e.g., [7–9]) and write $n \to \infty$ 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 $Y_j$ under the condition that $X_j = x$ (as in [16]), which we denote in what follows $x_j$; on the matter of the concerned location $j$, see Remark 1. Considering normalized sites, the kernel estimator of $r(x_j)$, is defined as

$$r_n(x_j) = \frac{\varphi_n(x_j)}{f_n(x_j)} \text{ if } f_n(x_j) \neq 0, \quad r_n(x_j) = \overline{Y} \quad \text{(empirical mean), otherwise, where}$$

$$\varphi_n(x_j) = \frac{1}{a_n b_n^d} \sum_{i \in \mathcal{I}_n} Y_i K_1 \left( b_n^{-1} (x_j - X_i) \right) K_2 \left( \rho_n^{-1} \frac{\| j - i \|}{n} \right),$$

$$f_n(x_j) = \frac{1}{a_n b_n^d} \sum_{i \in \mathcal{I}_n} K_1 \left( b_n^{-1} (x_j - X_i) \right) K_2 \left( \rho_n^{-1} \frac{\| j - i \|}{n} \right),$$

with $a_n = \sum_{i \in \mathcal{I}_n} K_2 \left( \rho_n^{-1} \frac{\| j - i \|}{n} \right)$. In addition, $K_1$ and $K_2$ are kernels respectively defined on $\mathbb{R}^d$ and $\mathbb{R}$, $b_n$ and $\rho_n$ are bandwidths tending to zero. Note that $K_2 \left( \rho_n^{-1} \frac{\| j - i \|}{n} \right) = K_2 \left( \frac{\| j - i \|}{\rho_n n} \right)$, where $\frac{1}{n} = (\frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_N})$. For each site $j$, the estimator $r_n(x_j)$ is a function of the number $k_n = k_{n,j} = \sum_{i | j - i| \leq d_n} 1$ of neighbors sites $i$, for which the distance between $i$ and $j$ is less or equal to distance $d_n > 0$ such that $d_n \to \infty$ as $n \to \infty$. More precisely, in what follows, we assume that $k_n = C_N d_n^N (1 + o(1))$ as $d_n \to \infty$ 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, $d_n$ is chosen to be $\eta \rho_n$ involving that $d_n^N = \overline{n} \rho_n^N$ and $k_n = O(\overline{n} \rho_n^N)$. 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 $Y_j$ in some unobserved location $j$. More precisely, we suppose that the field $(X_i, Y_i)_{i \in \mathbb{Z}^N}$ is observed on the set $\mathcal{O}_n$ contained in $\mathcal{I}_n$. The main purpose is to predict the unobserved value $Y_j$ given $X_j$ for a location $j \in \mathcal{I}_n$ but $j \notin \mathcal{O}_n$.

To achieve the forecasting at the site $j$, we propose to use the regression function estimator $r_n(x_j)$. Then, the prediction of the value of the field $(Y_i)_{i \in \mathbb{Z}^N}$ at the location $j \notin \mathcal{O}_n$ is written

$$\hat{Y}_j = r_n(X_j) = \frac{\sum_{i \in \mathcal{O}_n} Y_i K_1 \left( \frac{X_j - X_i}{\rho_n} \right) K_2 \left( \rho_n^{-1} \frac{\| j - i \|}{n} \right)}{\sum_{i \in \mathcal{O}_n} K_1 \left( \frac{X_j - X_i}{\rho_n} \right) K_2 \left( \rho_n^{-1} \frac{\| j - i \|}{n} \right)}.$$  \hspace{1cm} (1)
One can derive an asymptotic result such as almost complete convergence and consistency in $L^q$ norm ($q \in \mathbb{N}^*$) for $\hat{Y}_j$ from the kernel regression estimate, given below.

- More generally, one can extend $\hat{Y}_j$ by considering $\hat{Y}_j = \frac{\sum_{k \in \mathcal{C}_n} v_k K_1 \left( \frac{j+k}{\sqrt{n}} \right) K_2 \left( \frac{j+k}{\sqrt{n}} \right) - t}{\sum_{k \in \mathcal{C}_n} K_1 \left( \frac{j+k}{\sqrt{n}} \right) K_2 \left( \frac{j+k}{\sqrt{n}} \right)}$, where sites $i$ and $j$ 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 $(Z_i)$ satisfies the following $\alpha$-mixing condition: there exists a function $\psi(x) \downarrow 0$ as $x \to \infty$, such that $\alpha(\sigma(S), \sigma(S')) \leq \psi(\text{Card}(S), \text{Card}(S')) \psi(\text{dist}(S, S'))$, where $S$ and $S'$ are two finite sets of sites, $\text{Card}(S)$ denotes the cardinality of $S$, $\sigma(S) = [Z_i, i \in S]$ denotes a $\sigma$-field generated by $Z_i$, dist$(\cdot, \cdot)$ is the Euclidean distance, $\psi(\cdot)$ is a positive symmetric function nondecreasing in each variable. We will assume that $\psi(i)$ tends to zero at a polynomial rate, i.e., $\psi(i) \leq Ci^{-\theta}$. Let $u_n = \prod_{i=1}^{N} (\log n_i)(\log \log n_i)^{1+\epsilon}$, then $\sum_{n \in \mathbb{N}^*} \frac{1}{n u_n} < \infty$. 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

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

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 $C_{11}$ and $C_2$ with $0 < C_{11} < C_2 < \infty$, for $i = 1, 2$, such that

$$C_{11} \mathbf{1}_{[0,1]}(s') \leq K_1(s) \leq C_{21} \mathbf{1}_{[0,1]}(s') \quad \text{for } s \in \mathbb{R}^d,$$

$$C_{12} \mathbf{1}_{[0,1]}(t) \leq K_2(t) \leq C_{22} \mathbf{1}_{[0,1]}(t) \quad \text{for } t \in \mathbb{R},$$

where $s'$ denotes the transpose of $s$;

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

$$|f_{X_i,X_j}(u, v) - f_{X_i}(u) f_{X_j}(v)| < C;$$

A6: $\psi(n, m) \leq C \min(n, m)$ and $\hat{\theta}_{n}^{\rho_n} \log \hat{\theta}_{n}^{\rho_n} \to \infty$ with the mixing coefficient $\theta > N(q+2), q > 1$ and

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

A7: $\psi(n, m) \leq C \min(n, m)$ and $\hat{\theta}_{n}^{\rho_n} \log \hat{\theta}_{n}^{\rho_n} \to \infty$ with the mixing coefficient $\theta > N(q+2+\beta), q > 1, \beta > 1$ and

$$\theta_1' = \frac{N(q-1) - \theta}{N(q+2b+1) - \theta} > 0; \quad \theta_2' = \frac{-\theta - N}{N(q+2b+1) - \theta} < 0; \quad \theta_3' = \frac{2N}{N(q+2b+1) - \theta} < 0.$$

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 spatial-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 $f_{X_i} X_j$ and $f_x f_x X_j$ 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 $\psi(i)$ verifies: $\psi(i) \leq Ci^{-\theta}$, 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, \( r_n(x) \) converges almost completely (a.c.) to \( r(x) \) and

\[
|r_n(x) - r(x)| = 0 \left( b_n + \sqrt{\frac{\log \hat{n}}{n^2 n}} \right) \text{ a.c.}
\]

Pattern of the proof. We write

\[
|r_n(x) - r(x)| \leq \left( \frac{1}{f_n(x)} |\psi_n(x) - \phi(x)| + \frac{\phi(x)}{|f_n(x)| f(x)} |f_n(x) + f(x)| \right) 1_{[\Sigma_{i=1}^n W_{ni} \neq 0]} + \hat{Y} 1_{[\Sigma_{i=1}^n W_{ni} = 0]},
\]

where \( \hat{Y} \) is the empirical mean of the \( Y_i \).

We study the term \( |f_n(x) - f(x)| \) since it is a particular case of \( |\psi_n(x) - \phi(x)| \) when \( Y_i \) is equal to 1. The result is obtained studying separately the bias and the variance terms. It is easy to show that the bias \( |E(f_n(x)) - f(x)| = O(b_n). \) For the variance, an adjustment of Lemma 3.2 in [6] is used to obtain that

\[
P = P(|f_n(x)| - E(f_n(x))| > \epsilon) \\
\leq C N \hat{n}^{-a + 2N+2} \frac{C}{a_n j \bar{\psi}^{2/3} n} \psi((\bar{i} - 1)p^N, p^N) \varphi(p) \tilde{n}^{1/2}
\]

with \( a = \frac{6}{2N+4} + 2N+2 CN^2 \) and \( \epsilon = \delta \left( \frac{\log \hat{n}}{n^2 n} \right)^{1/2} \), \( \delta > 0 \), and \( p = \left( \frac{\hat{n}^2 n^2}{\log n} \right)^{1/2} \). In both assumptions on \( |\psi(n, m)\) (A6 and A7) and by appropriate choice of \( \delta > 2N+1 C N \), the bound of \( \sum_{n=(n_1, \ldots, n_k) \in \mathbb{N}^N} P \) is the general term of a convergent series.

We have

\[
P \left( \sum_{i \in I} W_{ni} = 0 \right) \leq P \left[ |f_n(x) - E[f_n(x)]| > \epsilon \right] \text{ for n large enough.}
\]

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

Theorem 3.2. Under Assumptions A1–A5 and A6 or A7, \( r_n(x) \) converges in mean of order \( q \) to \( r(x) \) and

\[
\|r_n(x) - r(x)\|_q = 0 \left( b_n + \sqrt{\frac{1}{\hat{n} \bar{d}_n^2 N}} \right), \quad q > 1,
\]

Pattern of the proof. Let \( W_{ni} = \frac{K_1(b_2 n^{-2/3} Y_i - X_i) \bar{r}_n^{-1} \left( \frac{1}{n} \right)}{\sum_{i \in I} K_1(b_2 n^{-2/3} Y_i - X_i) \bar{r}_n^{-1} \left( \frac{1}{n} \right)} \) and by adopting the convention \( 0/0 = 0 \), we have \( \sum_{i \in I} W_{ni} = 0 \) or 1. Consequently, we can deal with the following decomposition:

\[
\|r_n(x) - r(x)\|_q \leq E^{1/q} \left[ \sum_{i \in I} W_{ni} \left[ E(Y_i|X_i) - r(x) \right] 1_{[\Sigma_{i=1}^n W_{ni} = 1]} \right]^{q} + E^{1/q} \left[ \sum_{i \in I} W_{ni} \left[ Y_i - E(Y_i|X_i) \right] 1_{[\Sigma_{i=1}^n W_{ni} = 1]} \right]^{q} + E^{1/q} \left[ \left( \frac{1}{n} \right) \sum_{i \in I} Y_i - r(x) \right] 1_{[\Sigma_{i=1}^n W_{ni} = 0]} \right]^{q}.
\]

The study of the three terms of the right-hand-side gives the following result \( E^{1/q} [r_n(x) - r(x)]^q = O(b_n^q) + O \left( \left( \hat{n} \bar{d}_n^2 N \right)^{-1/2} \right) + O \left( \left( \hat{n} \bar{d}_n^2 N \right)^{-1/2} \right) \) 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