Journal of Multivariate Analysis 156 (2017) 103–115

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On local linear regression for strongly mixing random fields

Mohamed El Machkouri [a,*], Khalifa Es-Sebaiy [b], Idir Ouassou [b]

[a] *Université de Rouen Normandie, LMRS UMR CNRS 6085, avenue de l'Université, 76801 Saint-Étienne-du-Rouvray, France*

[b] *National School of Applied Sciences-Marrakesh, Cadi Ayyad University, Av. Abdelkrim Khattabi, 40000 Guéliz-Marrakech, Morocco*

ARTICLE INFO

Article history:
Received 22 February 2016
Available online 20 February 2017

AMS 2000 subject classifications:
62G05
60J25
62G07

Keywords:
Local linear regression estimation
Strong mixing
Random fields
Asymptotic normality

ABSTRACT

We investigate the local linear kernel estimator of the regression function g of a stationary and strongly mixing real random field observed over a general subset of the lattice $\mathbb{Z}^d$. Assuming that g is differentiable with derivative g', we provide a new criterion on the mixing coefficients for the consistency and the asymptotic normality of the estimators of g and g' under mild conditions on the bandwidth parameter. Our results improve the work of Hallin et al. (2004) in several directions.

1. Introduction and main results

In a variety of fields such as soil science, geology, oceanography, econometrics, epidemiology, image processing and many others, the aim of practioners is to handle phenomena observed on spatial sets. In particular, one of the fundamental questions is how to understand the phenomenon from a set of (dependent) observations based on regression models. In this work, we investigate the problem in the context of strongly mixing spatial processes (or random fields) and we focus on local linear regression estimation.

Let d be a positive integer and let $\{(Y_i, X_i) : i \in \mathbb{Z}^d\}$ be a strictly stationary $\mathbb{R}^2$-valued random field defined on a probability space $(\Omega, \mathcal{F}, \Pr)$. The estimation of its regression function g defined by $g(x) = \mathrm{E}(Y_0|X_0 = x)$ for almost all real x is a natural question and a very important task in statistics. The non-spatial case $d = 1$, i.e., dependent time series, has been extensively studied. One can refer, e.g., to Lu and Cheng [17], Masry and Fan [18], Robinson [21], Roussas [23] and many references therein. For $d \geqslant 2$, contributions to the case of strongly mixing random fields were made by Biau and Cadre [1], Carbon et al. [2], Dabo-Niang and Rachdi [4], Dabo-Niang and Yao [5], El Machkouri [7], El Machkouri and Stoica [10], Hallin et al. [12] and Lu and Chen [15,16].

To our knowledge, there are no results on regression estimation for dependent random fields which are not strongly mixing. However, the density estimation problem has been investigated, e.g., by El Machkouri [9] for a class of stationary ergodic random fields which contains non mixing linear random fields studied by Cheng et al. [3], Hallin et al. [11], and Wang and Woodroofe [25].

Given two σ-algebras $\mathcal{U}$ and $\mathcal{V}$, the α-mixing coefficient introduced by Rosenblatt [22] is defined by

$$\alpha(\mathcal{U}, \mathcal{V}) = \sup\{|\Pr(A \cap B) - \Pr(A)\Pr(B)|,\ A \in \mathcal{U},\ B \in \mathcal{V}\}.$$

* Corresponding author.
E-mail addresses: mohamed.elmachkouri@univ-rouen.fr (M. El Machkouri), k.essebaiy@uca.ma (K. Es-Sebaiy), i.ouassou@uca.ma (I. Ouassou).

http://dx.doi.org/10.1016/j.jmva.2017.02.002

Let p be fixed in $[1,\infty]$. The strong mixing coefficients $(\alpha_{1,p}(n))_{n\geqslant 0}$ associated to $\{(Y_i,X_i)\,:\,i\in\mathbb{Z}^d\}$ are defined by

$$\alpha_{1,p}(n)=\sup\{\alpha(\sigma(Y_k,X_k),\mathcal{F}_\Gamma),\ k\in\mathbb{Z}^d,\ \Gamma\subset\mathbb{Z}^d,\ |\Gamma|\leqslant p,\ \rho(\Gamma,\{k\})\geqslant n\},$$

where $\mathcal{F}_\Gamma=\sigma(Y_i,X_i;\ i\in\Gamma)$, $|\Gamma|$ is the number of element in Γ and the distance ρ is defined for any subsets Γ_1 and Γ_2 of $\mathbb{Z}^d$ by $\rho(\Gamma_1,\Gamma_2)=\min\{|i-j|,\ i\in\Gamma_1,\ j\in\Gamma_2\}$ with $|i-j|=\max_{1\leqslant s\leqslant d}|i_s-j_s|$ for any $i=(i_1,\ldots,i_d)$ and $j=(j_1,\ldots,j_d)$ in $\mathbb{Z}^d$. We say that the random field $(Y_i,X_i)_{i\in\mathbb{Z}^d}$ is strongly mixing if $\lim_{n\to\infty}\alpha_{1,p}(n)=0$.

Fix $x\in\mathbb{R}$. Assuming that g is differentiable at x with derivate $g'(x)$, the main idea in local linear regression is to approximate $g(z)$ by $g(x)+g'(x)(z-x)$ for any z in the neighborhood of x and to estimate $(g(x),g'(x))$ instead of using a classical nonparametric estimation method (e.g., Nadaraya–Watson kernel estimator) for g itself. Following [12], we define the local linear kernel regression estimator of $\left(g(x),g'(x)\right)^\top$ by

$$\begin{pmatrix} g_n(x)\\ g_n'(x)\end{pmatrix}=\operatorname*{argmin}_{(s,t)\in\mathbb{R}^2}\sum_{i\in\Lambda_n}\{Y_i-s-t(X_i-x)\}^2K\left(\frac{X_i-x}{b_n}\right),\tag{1}$$

where b_n is the bandwidth parameter going to zero as $n\to\infty$, Λ_n is a finite subset of $\mathbb{Z}^d$ whose cardinality $|\Lambda_n|$ goes to infinity as $n\to\infty$, and K is a probability kernel, i.e., a function $K:\mathbb{R}\to\mathbb{R}$ such that $\int_{\mathbb{R}}K(s)ds=1$. We introduce the following notations:

$$u_{00}(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}K\left(\frac{X_i-x}{b_n}\right),\qquad u_{11}(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}\left(\frac{X_i-x}{b_n}\right)^2K\left(\frac{X_i-x}{b_n}\right),$$

$$u_{01}(n)=u_{10}(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}\left(\frac{X_i-x}{b_n}\right)K\left(\frac{X_i-x}{b_n}\right),$$

$$v_0(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}Y_iK\left(\frac{X_i-x}{b_n}\right),\qquad v_1(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}Y_i\left(\frac{X_i-x}{b_n}\right)K\left(\frac{X_i-x}{b_n}\right),$$

$$w_0(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}Z_iK\left(\frac{X_i-x}{b_n}\right)\quad\text{and}\quad w_1(n)=\frac{1}{|\Lambda_n|b_n}\sum_{i\in\Lambda_n}Z_i\left(\frac{X_i-x}{b_n}\right)K\left(\frac{X_i-x}{b_n}\right)$$

with $Z_i=Y_i-g(x)-g'(x)(X_i-x)$. A straightforward calculation gives

$$\begin{pmatrix} g_n(x)\\ g_n'(x)b_n\end{pmatrix}=U_n^{-1}V_n,\quad\text{where } U_n=\begin{pmatrix}u_{00}(n) & u_{10}(n)\\ u_{01}(n) & u_{11}(n)\end{pmatrix}\text{ and } V_n=\begin{pmatrix}v_0(n)\\ v_1(n)\end{pmatrix}.$$

Denoting $W_n=V_n-U_n\left(g(x),g'(x)b_n\right)^\top=(w_0(n),w_1(n))^\top$, we obtain

$$G(n,x)\equiv\begin{pmatrix} g_n(x)-g(x)\\ \left\{g_n'(x)-g'(x)\right\}b_n\end{pmatrix}=U_n^{-1}W_n.\tag{2}$$

This paper's main contribution is to provide sufficient conditions ensuring the consistency (Theorem 1) and the asymptotic normality (Theorem 2) of the estimator defined by (1) under mild conditions on the bandwidth parameter; see Assumption (A6). Our approach is based on the so-called Lindeberg method [8–10,14] instead of Bernstein's blocking method used in several previous works for proving limit theorems in the random field setting [2,12,24]. Finally, we emphasize that our work can be extended without any effort to any $(\mathbb{R}^N\times\mathbb{R})$-valued random field $\{(Y_i,X_i)\,:\,i\in\mathbb{Z}^d\}$ with arbitrary positive integer N. The present discussion is limited to the case $N=1$ only to simplify notation.

Let $K:\mathbb{R}\to\mathbb{R}$ be a probability kernel. For any $c=(c_0,c_1)\in\mathbb{R}^2$ and any s in $\mathbb{R}$, we define $K_c(s)=(c_0+c_1s)K(s)$. In the sequel, we consider the following assumptions:

(A1) For any c in $\mathbb{R}^2$, we have $\sup_{t\in\mathbb{R}}|K_c(t)|<\infty$, $\int_{\mathbb{R}}|K_c(t)|dt<\infty$ and K_c has an integrable second-order radial upper bound, i.e., the function ψ defined for any real x by $r(x)=\sup_{|t|\geqslant|x|}t^2K_c(t)$ is integrable.

(A2) g is twice differentiable and g'' is continuous.

(A3) There exists a positive constant κ such that $\sup_{k\neq 0}|f_{0,k}(x,y)-f(x)f(y)|\leq\kappa$ for any (x,y) in $\mathbb{R}^2$, where $f_{0,k}$ is the continuous joint density of (X_0,X_k) and f is the continuous marginal density of X_0.

(A4) $\mathrm{E}\left(|Y_0|^{2+\delta}\right)<\infty$ for some $\delta>0$.

(A5) $b_n\to 0$ in such a way that $|\Lambda_n|b_n^3\to\infty$.

(A6) $b_n\to 0$ in such a way that $|\Lambda_n|b_n\to\infty$ and $|\Lambda_n|b_n^5\to 0$.

Our first main result ensures the consistency of the estimator.