Journal of Statistical Planning and Inference 193 (2018) 15–41

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Orthogonal series estimates on strong spatial mixing data

Johannes T.N. Krebs

Department of Mathematics, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67653, Kaiserslautern, Germany

ARTICLE INFO

Article history:
Received 2 September 2016
Received in revised form 19 May 2017
Accepted 28 July 2017
Available online 9 August 2017

MSC 2010:
62G08
62H11
60G60

Keywords:
Empirical orthonormalization
L^2-consistency
Nonparametric nonlinear regression
Rate of convergence
Strong spatial mixing
Spatial processes

ABSTRACT

We study a nonparametric regression model for sample data which is defined on an N-dimensional lattice structure and which is assumed to be strong spatial mixing: we use design adapted multidimensional Haar wavelets which form an orthonormal system w.r.t. the empirical measure of the sample data. For such orthonormal systems, we consider a nonparametric hard thresholding estimator. We give sufficient criteria for the consistency of this estimator and we derive rates of convergence. The theorems reveal that our estimator is able to adapt to the local smoothness of the underlying regression function and the design distribution. We illustrate our results with simulated examples.

1. Introduction

In this article we study penalized nonparametric sieve estimators for spatial sample data which features a certain dependence structure: the data is given by the random field (X, Y) which is indexed by a set S of spatial coordinates and which is strong spatial mixing. Here, we take $S = \mathbb{Z}^N$ for some lattice dimension $N \in \mathbb{N}_+$ but our discussion is not limited to that regular case; we could also allow that the random field is only partially observed at some $S \subseteq \mathbb{Z}^N$.

The random variables $X(s)$ are $\mathbb{R}^d$-valued and have equal marginal distributions denoted by the probability measure μ on the Borel-σ-algebra of $\mathbb{R}^d$, $\mathcal{B}(\mathbb{R}^d)$. The $Y(s)$ are $\mathbb{R}$-valued, square integrable and satisfy the equation

$$Y(s) = m(X(s)) + \varsigma(X(s))\,\varepsilon(s), \quad \text{for each } s \in S \tag{1.1}$$

where $m, \varsigma : \mathbb{R}^d \to \mathbb{R}$ are functions in $L^2(\mu)$. The error terms $\varepsilon(s)$ are distributed with mean zero and variance one, i.e., $\varepsilon(s) \sim (0, 1)$. Furthermore, they are independent of X and have identical marginal distributions but may be dependent among each other such that the strong spatial mixing property remains valid. We emphasize that there is no requirement on the distribution of the error terms, e.g., a Gaussian distribution is not necessary. The same is true for the distribution of the regressors $X(s)$, it is not required that these admit a density with respect to the Lebesgue measure.

Thus, we apply the classical heteroscedastic regression model to spatial data under minimal assumptions on the random field (X, Y). An introduction to spatial statistics is given by Cressie (1993). In particular, Markov random fields are studied in the monograph of Kindermann and Snell (1980).

E-mail address: krebs@mathematik.uni-kl.de.

http://dx.doi.org/10.1016/j.jspi.2017.07.005
0378-3758/

Nonparametric regression on spatial data has gained importance, in particular, the case where the data is given on a regular lattice structure: Hallin et al. (2004) study a local linear kernel estimator under a strong spatial mixing condition. Li (2016) considers a nonparametric regression estimator for such lattice data which is constructed with wavelets.

In this article we consider a nonparametric estimator for lattice data, too, however, we do this with a penalized orthogonal series estimator. Baraud et al. (2001) consider penalized estimators for β-mixing time series $\{(X_t, Y_t) : t \in \mathbb{N}\}$ where the regressors X_t are multidimensional. Orthogonal series estimators have been studied for various data situations: for a real-valued one-dimensional regressor X a popular choice is piecewise polynomials. Comte and Rozenholc (2004) study an algorithm for the construction in the case of fixed design regression. Kohler (2003) gives a generalization to random design regression under the assumption that the error terms are bounded. Akakpo and Lacour (2011) use piecewise polynomials for conditional density estimation of a β-mixing time series $\{(X_t, Y_t) : t \in \mathbb{N}\}$. In another article (Kohler, 2008) considers Haar wavelets to construct an orthogonal series estimator in the case of a multivariate regressor X under the assumption of sub Gaussian error terms and a bounded design distribution of X. The ideas and results obtained in the latter can be considered as the starting point for our analysis.

Before we give a more thorough introduction to the results of this article, we mention that there exist alternative approaches to construct orthogonal series estimators for a random (univariate) regressor X. Kerkyacharian and Picard (2004) consider warped wavelets in the case where the regressor X admits a density on a compact real interval. Kulik and Raimondo (2009) use this concept to study time series with long range dependence errors. Delouille et al. (2001) construct a soft thresholding regression estimator for univariate i.i.d. sample data. They derive rates of convergence for Hölder continuous regression functions in a model where the design variables X are supposed to admit a density which has bounded support. Girardi and Sweldens (1997) show that design adapted Haar wavelets can generate even a multiresolution analysis in the one-dimensional case. Masry (2000) studies α-mixing stationary processes and derives rates of convergence for regression functions which belong to a multidimensional Besov space.

In this article, we transfer the ideas of Kohler (2008) to the spatial setting where the sample data is no longer independently distributed but where the dependence vanishes with an increasing lattice distance between the random variables. We relax most restrictions which are usually made in the context of nonparametric regression on dependent data. Most notably, the design distribution (which is the distribution of the $X(s)$) does not need to be known and is not restricted to a bounded domain. Furthermore, the distribution does not need to admit a density w.r.t. the Lebesgue measure as it is for instance assumed in Hallin et al. (2004). Li (2016) assumes in the spatial wavelet regression model that the $X(s)$ admit a density which is known. We do not do this here. Additionally, we do not require the error terms in the regression model to be bounded or sub Gaussian; we develop our results here for a general class of error terms which satisfies a certain condition on the tail distribution. In addition in order to show that the estimator is consistent in the L^2-sense, we do not need a bounded regression function.

In this paper we establish general consistency results for our nonparametric regression estimators and we derive rates of convergence. Since our assumptions on the distribution of the regressor X and on the error terms ε are less restrictive than usual, we obtain, however, a sub-optimal rate of convergence, when compared to the results of Stone (1982). We shall discuss this further in the corresponding parts of the article.

The remainder of the paper is organized as follows: we give the notation and definitions which we use throughout the article in Section 2. In Section 3 we present the main results: we give a general consistency theorem for our nonparametric estimator and derive a rate of convergence theorem. In Section 4, we give numerical applications and make the comparison with i.i.d. data. The proofs of our theorems are presented in Section 5. Appendices A and B contain certain deferred proofs and further background material which proves to be useful in the broader context of random fields. Furthermore, we provide in a supplemental (Krebs, 2016b) some technical results concerning our simulation procedure.

2. Notation and definitions

We work on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ which is equipped with a generic random field Z. Z is $\mathbb{R}^d$-valued and is indexed by $\mathbb{Z}^N$, for both $N, d \in \mathbb{N}_+$. This means $Z = \{Z(s) : s \in \mathbb{Z}^N\}$ and $Z(s) : \Omega \to \mathbb{R}^d$ is Borel-measurable for each $s \in \mathbb{Z}^N$. The random field Z is stationary (or homogeneous) if for each translation $t \in \mathbb{Z}^N$ and for each collection of finite points $s_1, \ldots, s_n$ the joint distribution of $\{Z(s_1 + t), \ldots, Z(s_n + t)\}$ coincides with the joint distribution of $\{Z(s_1), \ldots, Z(s_n)\}$, i.e.,

$$\mathcal{L}(Z(s_1 + t), \ldots, Z(s_n + t)) = \mathcal{L}(Z(s_1), \ldots, Z(s_n)).$$

We denote by $\|\cdot\|_p$ the Euclidean p-norm on $\mathbb{R}^N$ and by d_p the corresponding metric for $p \in [1, \infty]$ with the extension $d_p(I, J) := \inf\{d_p(s, t), s \in I, t \in J\}$ for subsets I, J of $\mathbb{R}^N$. Furthermore, write $s \le t$ for $s, t \in \mathbb{R}^N$ if and only if for each $1 \le k \le N$ the single coordinates satisfy $s_k \le t_k$. We denote the indicator function of a set A by $\mathbb{1}\{A\}$ and abbreviate for a subset I of $\mathbb{Z}^N$ by $\mathcal{F}(I) = \sigma\{Z(s) : s \in I\}$ the σ-algebra generated by the $Z(s), s \in I$.

As a measure of spatial dependence we use the α-mixing coefficient. This coefficient is introduced by Rosenblatt (1956); in the spatial context, it is given for $k \in \mathbb{N}$ as

$$\alpha(k) := \sup_{\substack{I,J \subseteq \mathbb{Z}^N, \\ d_\infty(I,J) \ge k}} \sup_{\substack{A \in \mathcal{F}(I), \\ B \in \mathcal{F}(J)}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|. \tag{2.1}$$