Nonparametric regression estimation and prediction for spatial processes

Sophie Dabo-Niang 1 University Charles De Gaulle, Lille 3 and Anne-Françoise Yao2 University Aix-Marseille 2

1 Laboratoire GREMARS, Université Charles De Gaulle, Lille 3, maison de la recherche, domaine du pont de bois, BP 60149, 59653 Villeneuve d’ascq cedex, France
2 Laboratoire LMGEM, Université Aix-Marseille 2, Campus de Luminy, case 901, 13288 Marseille cedex 09, France

Abstract: We investigate a kernel estimate of the spatial regression function
\[ r(x) = E(Y_u | X_u = x), \quad x \in \mathbb{R}^d \]
of a stationary multidimensional spatial process, \( \{Z_u = (X_u, Y_u), \; u \in \mathbb{R}^N\} \). The weak and the strong consistencies of the estimate are shown under sufficient conditions on mixing coefficients and on the bandwidth, when the process is observed over a rectangular domain of \( \mathbb{R}^N \). Special attention is paid to apply these results to predict unsampled locations of a random field \( (F_u)_{u \in \mathbb{R}^N} \).

Keywords: Spatial process; Kernel regression estimation; nonparametric spatial prediction.

1 Introduction

The spatial regression estimation as well as prediction is an interesting and crucial problem in statistical inference for a number of applications, where the influence of a vector of covariates on some response variable is to be studied in a context of spatial dependence. Spatial data are modeled as finite realizations of random fields and are collected from different spatial location on the earth, as in a variety of fields, including soil science, geology, oceanography, econometrics, epidemiology, environmental science, forestry and many others. The literature on spatial models is relatively abundant, see for example Guyon (1995), Anselin and Florax (1995), Cressie (1991) or Ripley (1981) for a list of references. However, the nonparametric treatment of such data is limited. Not many theoretical works have been devoted so far to nonparametric regression estimation and prediction for spatial processes. Let \( Z_i = (X_i, Y_i), \; i \in \mathbb{Z}^N, \; (N, d \geq 1) \) be a stationary \( \mathbb{R}^d \times \mathbb{R} \)-valued measurable random field defined on a probability space \( (\Omega, \mathcal{A}, P) \) and observed over a rectangular domain \( I_n = \{i = (i_1, \ldots, i_N) \in \mathbb{Z}^N, \; 1 \leq i_k \leq n_k, \; k = 1, \ldots, N\} \), \( n = (n_1, \ldots, n_N) \in \mathbb{Z}^N \). A point \( i \) will be referred to as a site.
Key references on spatial nonparametric estimation of the density of \( X_i \), are Tran (1990), Tran and Yakowitz (1993), Carbon, Hallin and Tran (1996), Carbon, Tran and Wu (1997), Hallin, Lu and Tran (2004a). Except some papers, Biau and Cadre (2004), Lu and Chen (2002, 2004), Hallin, Lu and Tran (2004b), little attention has been paid to estimate nonparametrically the regression function \( E[Y_i|X_i = x] \).

As explained in Biau (2003), there are some issues where it is not reasonable to assume that the stochastic process under study is varying in \( Z^N \), such as in oil engineering, geostatistics,... where data are thought of as realizations of random process \( (Z_u, u \in I) \), \( I \) being a subset of \( \mathbb{R}^N \). To the best of our knowledge, the only paper dealing with kernel estimation problems in continuously indexed random fields is Biau (2003).

We are mainly concerned in this work with weak and strong consistencies as well as convergence rates of the kernel regression and predictor in the case of continuous indexed random fields.

2 Presentation of estimates

Consider \( Z_u = (X_u, Y_u) \), \( u \in \mathbb{R}^N \) be a \( \mathbb{R}^d \times \mathbb{R} \)-valued measurable and stationary spatial process \( (N \geq 1, d \geq 1) \) defined on a probability space \( (\Omega, \mathcal{A}, P) \). A point \( u = (u_1, ..., u_N) \in \mathbb{R}^N \) will be referred to as a site. We assume that the \( Z_u \)'s have the same distribution as \( (X, Y) \) admitting an unknown density \( f_{X,Y} \) with respect to Lebesgue measure \( \lambda \) over \( \mathbb{R}^{d+1} \). We assume that \( Y \) is integrable and we denote by \( f \) the density of \( X \).

The regression function \( r(\cdot) \) of \( Y \) given \( X \) is defined by

\[
  r(x) = \begin{cases} 
    \frac{\varphi(x)}{f(x)} & \text{if } f(x) \neq 0; \\
    EY & \text{if } f(x) = 0 
  \end{cases}
\]

where \( \varphi(x) = \int_R y f_{X,Y}(x, y) dy \), \( x \in \mathbb{R}^d \). We define the rectangular region \( \mathcal{I}_T \) by \( \mathcal{I}_T = \{u \in \mathbb{R}^N : 0 \leq u_i \leq T_i, i = 1, ..., N\} \) where \( T = (T_1, ..., T_N) \).

We write \( T \to +\infty \) if \( \min_{i=1,...,N} T_i \to +\infty \) and we set \( \hat{T} = T_1 \times ... \times T_N \).

We suppose that a continuous spatial sample is available on \( \mathcal{I}_T \). The kernel density and regression estimators based on \( (Z_u = (X_u, Y_u), u \in \mathcal{I}_T) \) are respectively defined by

\[
  f_T(x) = \frac{1}{T h_T^d} \int_{\mathcal{I}_T} K\left(\frac{x - X_u}{h_T}\right) d\mathbf{u}
\]

and

\[
  r_T(x) = \begin{cases} 
    \frac{\varphi_T(x)/f_T(x)}{f_T(x)} & \text{if } f_T(x) \neq 0; \\
    \frac{1}{T} \int Y_u d\mathbf{u} & \text{if } f_T(x) = 0. 
  \end{cases}
\]

with

\[
  \varphi_T(x) = \frac{1}{T h_T^d} \int_{\mathcal{I}_T} Y_u K\left(\frac{x - X_u}{h_T}\right) d\mathbf{u}, \; x \in \mathbb{R}^d,
\]
where \( \lim_{T \to +\infty} h_T = 0 (+) \) and \( K : R^d \to R \) is a bounded integrable kernel such that \( \int K(x)\,dx = 1 \).

Consider a bounded open set \( S \subset R^N \) with smooth boundary \( \partial S \) and closure \( \overline{S} = S \cup \partial S \). Let \( D_S = \{ V \subset S, V \text{ is a bounded open set with } \partial V \} \).

Now, let \( (\xi_t, t \in R^N) \) be a \( R \)-valued strictly stationary random spatial process, assume to be observed in \( S \). We want to predict the square integrable value, \( \xi_{t_0} \), at a given non-observed fixed point , \( t_0 \notin \overline{S} \) admitting a bounded neighborhood \( V_{t_0} \) such that \( S \cap V_{t_0} \) is not empty and belongs to \( D_S \) (not containing \( t_0 \)).

In this work, we will be concerned with consistency of the kernel regression estimate \( r_T \) and with a predictor of \( \xi_{t_0} \) based on \( r_T \).

References


4 Nonparametric spatial regression and prediction


