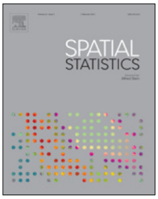
ARTICLE IN PRESS

Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

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



Convergence in distribution of the L_2-deviations of the kernel-type variogram estimators with applications

Pilar García-Soidán [*], Tomás R. Cotos-Yáñez

Department of Statistics and Operations Research, University of Vigo, Spain

A R T I C L E   I N F O

Article history:
Received 29 October 2016
Accepted 3 February 2017
Available online xxxx

Keywords:
Global measure
Goodness of fit test
Intrinsic stationarity
Variogram
Kernel estimation

A B S T R A C T

In this work, some properties of the L_2-deviations of the Nadaraya–Watson variogram estimators are analyzed, for both the anisotropic and the isotropic settings. Their convergence in distribution is established, which provides the basis for addressing practical problems, such as the construction of goodness of fit tests for the variogram and, therefore, for modeling the spatial dependence. However, the development of the latter application requires solving different issues, such as the approximation of the model parameters and the critical points. For estimation of the former ones, we propose proceeding through the least squares criteria, whose consistency will be proved, together with a reformulation of the global measures for the kernel-type estimators. Then, the resulting critical points can be approximated by appealing to the bootstrap approaches. Numerical studies with simulated and real data have been developed to illustrate the potentiality of our results, in order to check the appropriateness of a variogram model selected for the variogram.

1. Introduction

The kriging techniques allow the researchers to reconstruct a phenomenon over the whole observation region, from a finite set of data, with applications in a large spectrum of areas, such as

---

[*] Correspondence to: Faculty of Social Sciences and Communication, CP 36005 Pontevedra, Spain.
E-mail addresses: pgarcia@uvigo.es (P. García-Soidán), cotos@uvigo.es (T.R. Cotos-Yáñez).

http://dx.doi.org/10.1016/j.spasta.2017.02.002
2211-6753/

Please cite this article in press as: García-Soidán, P., Cotos-Yáñez, T.R., Convergence in distribution of the L_2-deviations of the kernel-type variogram estimators with applications. Spatial Statistics (2017), http://dx.doi.org/10.1016/j.spasta.2017.02.002

hydrology, atmospheric science, geology, etc. However, the aforementioned procedures demand an appropriate characterization of the second-order structure of the underlying random process, which can be addressed through the variogram or the covariance functions.

Let $\{Z(\mathrm{s}) \in \mathbb{R} : \mathrm{s} \in D \subset \mathbb{R}^d\}$ be a spatial random process, where D denotes the observation region. We will focus our attention on the variogram of an intrinsically stationary random process, so the following conditions are assumed:

(I1) $\mathrm{E}[Z(\mathrm{s})] = \mu$, for all $\mathrm{s} \in D$ and some constant μ.
(I2) $\mathrm{Var}[Z(\mathrm{s}) - Z(\mathrm{s}')] = 2\gamma(\mathrm{s} - \mathrm{s}')$, for all $\mathrm{s}, \mathrm{s}' \in D$ and some function γ, which is referred to as the semivariogram (2γ is the variogram).

Furthermore, an intrinsic random process is known as isotropic, when condition (I2) is replaced by:

(I2′) $\mathrm{Var}[Z(\mathrm{s}) - Z(\mathrm{s}')] = 2\gamma(\|\mathrm{s} - \mathrm{s}'\|)$, for all $\mathrm{s}, \mathrm{s}' \in D$ and some function γ, where $\|\cdot\|$ denotes the Euclidean norm.

Hypotheses I1 and I2 impose that the first two moments of $Z(\mathrm{s}) - Z(\mathrm{s}')$ depend only on $\mathrm{s} - \mathrm{s}'$ and, therefore, on the distance and the relative orientation of the lag between the locations involved. The more restrictive condition I2′ avoids the influence of the lag-orientation on both moments.

For approximation of the variogram, nonparametric procedures may be used in a first step, providing us with the empirical estimator and more robust alternatives, studied in Matheron (1963), Cressie (1993) or Genton (1998). On the other hand, kernel-type approaches can be derived by adapting the Nadaraya–Watson estimation or the local linear method to the spatial setting, as suggested in Hall and Patil (1994) or in García-Soidán et al. (2003), respectively. Alternative mechanisms for approximating the semivariogram include the constant and the variable nearest-neighbor estimators, given in Yu and Mateu (2002), which yield generalizations of the Matheron and the Nadaraya–Watson semivariograms, respectively.

The performance of different nonparametric semivariogram estimators is analyzed in Menezes et al. (2006), which have been put into comparison in a numerical study covering a range of dependence situations. Nevertheless, the above-mentioned approaches are not necessarily valid for their direct application to spatial prediction. In fact, they typically fail to fulfill the conditionally negative definiteness property, which is satisfied by the theoretical semivariogram and must be required from their estimators, in order to guarantee a solution for the kriging equations. We can cope with this problem by first choosing a valid parametric family and then selecting within it the variogram which best fits the data, as described in Cressie (1993) or extended in Shapiro and Botha (1991) to a broad class of valid variograms, not depending on a small number of parameters.

The usage of a parametric estimator may be attractive at first because of its simplicity and validity, although one of its main drawbacks is the procedure followed for the choice of a parametric model, typically addressed through graphical diagnostics, since the shape of some variogram models is similar. An alternative is developed in Gorsich and Genton (2000), based on the fact that, unlike the variogram models, their derivatives are often quite different; hence, an estimation of the first variogram derivative may help to select from different models.

The current study is aimed to obtain global measures for the Nadaraya–Watson variogram deviations, involving the L_2-norm. These results will provide the basis for developing applications, such as that of testing the goodness of fit of a variogram model, which requires solving the general contrast:

$$\begin{aligned} &H_0 : \gamma \in \Gamma_\theta = \{\gamma_\theta(\cdot) : \theta \in \Theta \subset \mathbb{R}^p\} \quad \text{versus} \\ &H_1 : \gamma \notin \Gamma_\theta. \end{aligned} \tag{1}$$

In the statistics literature, an approach for the aforementioned goal was proposed in Maglione and Diblasi (2004), specifically designed for gaussian random processes. An extensive option is analyzed in Crujeiras et al. (2010), based on constructing tests for the spectral density of spatial processes observed on a regular grid. However, an alternative methodology could be applied, which tried to mimic the goodness of fit tests suggested for curve estimation with independent data, such as those given in Fan (1994) or Härdle and Mammen (1993), for the density or the regression settings. With this aim, the convergence in distribution of the L_2-deviations of the Nadaraya–Watson estimators of the variogram must be established. In addition, the approximation of the model parameters and the critical points

Please cite this article in press as: García-Soidán, P., Cotos-Yáñez, T.R., Convergence in distribution of the L_2-deviations of the kernel-type variogram estimators with applications. Spatial Statistics (2017), http://dx.doi.org/10.1016/j.spasta.2017.02.002