Modelling spatial dependence using kernel variogram estimators

Pilar García-Soidán

1 Universidad de Vigo, Facultad de Ciencias Sociales, 36005 Pontevedra, Spain

Abstract: In this work, several measures of the Nadaraya-Watson variogram estimators are established, via the $L_2$ norm, which will allow to construct non-parametric tests for modelling the spatial dependence. In particular, we will focus on developing some approaches for assessing isotropy or testing the goodness of fit of a parametric model selected for the variogram.

Keywords: Goodness of fit; Intrinsic process; Isotropy; Variogram.

1 Introduction

Independence between observations cannot be often assumed in the analysis of spatial data, so that the specific type of dependence observed must be modelled. For intrinsic random processes, the variogram provides a measure of the spatial dependence, although several difficulties arise when proceeding to its estimation. In a first step, the nonparametric estimators may be used for this purpose, such as the empirical variogram, appearing in Matheron (1963), or the more robust estimators suggested in Cressie and Hawkins (1980) and Genton (1998). On the other hand, two kernel-type estimators are introduced in García-Soidán et al. (2003) or García-Soidán et al. (2004) and García-Soidán (2006), which result from adapting the local linear or the Nadaraya-Watson estimators to the context of spatial data, respectively.

However, the nonparametric estimators cannot be used directly for spatial prediction, since the conditionally negative definiteness property typically lacks. We can cope with this problem by choosing a valid parametric model and then selecting that variogram in the family considered which best fit the data, by using any of the different criteria proposed; see Menezes et al. (2005) for a review of several approaches, which are put into comparison in a numerical study covering different spatial dependence situations.

The parametric estimation demands an appropriate mechanism to select from different variogram models and the graphical diagnostics, typically applied in practice, are often difficult to assess. A goodness of fit test has been suggested in Maglione and Dibiasi (2004), where the random process is assumed to be gaussian and isotropic. Then, in this work, we will propose procedures to check whether or not the theoretical semivariogram of
a general intrinsic random process follows a parametric model, for both
the isotropic and the anisotropic settings, by carrying out the following
contrast:
\begin{align*}
H_0 : \gamma &\in \Gamma_\theta = \{\gamma_\theta(\cdot)/\theta \in \Theta \subset \mathbb{R}^p\} \text{ versus } \\
H_1 : \gamma &\notin \Gamma_\theta
\end{align*}

(1)

In practice, isotropy is a common assumption when analyzing the spa-
tial variation, as it simplifies interpretation and computation; however,
we should check for isotropy before working under this requirement. The
isotropic condition is typically checked through graphical methods, al-
though the latter procedures rarely provide definitive results. A formal
approach to test for isotropy has been introduced in Guan et al. (2004),
for strictly stationary random processes. Thus, our aim will be to provide
a nonparametric test for assessing the hypothesis of isotropy of an intrinsic
random process.

To develop the testing approaches mentioned above, the process generating
the design points is assumed to follow the asymptotic regime described in
Hall et al. (1994), say, that it produces an increasingly dense set of random
locations that are dispersed across an expanding observation region.

2 Main results

Let \( \{Z(s)/s \in D \subset \mathbb{R}^d\} \) be a spatial random process. The assumption that
a random process is intrinsic conveys that the first two moments of the
difference \( Z(s) - Z(t) \) depend only on the relative location, \( s - t \), of both
variables; moreover, in case of isotropy, the referred two moments are only
dependent on the distance of the spatial locations, \( \|s - t\| \).

Then, suppose that the random process \( \{Z(s)/s \in D \subset \mathbb{R}^d\} \) is intrinsic
with semivariogram \( \gamma \). Denote by \( Z(s_1), Z(s_2), \ldots, Z(s_n) \), \( n \) data collected,
at the respective spatial locations \( s_1, s_2, \ldots, s_n \). Several approaches have
been proposed for estimation of the semivariogram, although we will restrict
our attention to the Nadaraya-Watson estimators, for both the anisotropic
and the isotropic settings, which will be respectively denoted as \( \hat{\gamma}_h(s) \) and
\( \hat{\gamma}_{h,1}(s) \), with \( s \) denoting a vector in \( \mathbb{R}^d \) and \( s \) a real value.

Some properties of the latter kernel estimators have been developed if seve-
ral hypotheses are satisfied. In particular, an increasing observation region
\( D \) has been considered, \( D = D_n = \lambda D_0 \) for some \( \lambda \to +\infty \) and some
bounded region \( D_0 \subset \mathbb{R}^d \). Additionally, a random design has been assumed
for the spatial locations, \( s_i = \lambda u_i \), for \( 1 \leq i \leq n \), where \( u_1, \ldots, u_n \) represents
a realization of a random sample of size \( n \), \( U_1, \ldots, U_n \), drawn from a den-
sity function \( f_0 \) considered on \( D_0 \). Then, consistency and the asymptotic
normality of the Nadaraya-Watson semivariograms have been established
In this work, we will prove that, under several conditions:

\[
(n^{-2}\lambda_d h^{-d})^{-1} \int_{\|s\| \leq y} (\hat{\gamma}_h(s) - \mathcal{K}_{d,h}\gamma(s))^2 \, ds \xrightarrow{d} \int_{\|s\| \leq y} X_d(s)^2 \, ds \quad (2)
\]

with

\[
\mathcal{K}_{d,h}\gamma(s) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K_d \left( \frac{s - (s_i - s_j)}{h} \right) \gamma(s_i - s_j)}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K_d \left( \frac{s - (s_i - s_j)}{h} \right)}
\]

where \(K_d\) represents a \(d\)-variate kernel function, \(h\) is the bandwidth parameter and \(\{X_d(s)/\|s\| < y\}\) is a gaussian process with zero mean and covariance function dependent on the forth-order moments of the random process.

As a consequence of (1), we may construct a test to check the validity of a parametric model selected for the semivariogram. The latter requires an adequate estimator of the true parameter \(\theta\), under the null hypothesis \(H_0\) in (1). From the different alternatives proposed in the literature, we will consider the least squares criteria, since they are known to require the fewest distributional assumptions about the random process; in addition consistency of the resulting test will follow from the results achieved in Lahiri et al. (2002).

With regard to the unknown terms dependent on the spatial process, we will cope with the estimation of the forth-order moments via the kernel method, which will allow to retain rates of convergence.

Another application of the limit distribution above may be that of constructing a nonparametric approach to test for isotropy. For the latter purpose, we may replace \(\gamma(s)\) for \(\hat{\gamma}_{h_1}(\|s\|)\) in (2), with \(h_1\) a bandwidth parameter, and check that under isotropy the same asymptotic distribution holds.

If the random process proves to be isotropic, a more specific goodness of fit test may be used, by considering a quadratic functional as given in (2) involving \(\hat{\gamma}_{h_1}\) and:

\[
\mathcal{K}_{1,h_1}\gamma(s) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K_1 \left( \frac{s - \|s_i - s_j\|}{h_1} \right) \gamma(\|s_i - s_j\|)}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K_1 \left( \frac{s - \|s_i - s_j\|}{h_1} \right)}
\]

instead of \(\hat{\gamma}_h\) and \(\mathcal{K}_{d,h}\), respectively.

The point in this case would be to derive the limit distribution of the resulting functional to yield:

\[
(n^{-2}\lambda_d h_1^{-d})^{-1} \int_0^y (\hat{\gamma}_{h_1}(s) - \mathcal{K}_{1,h_1}\gamma(s))^2 \, ds \xrightarrow{d} \int_0^y X(s)^2 \, ds
\]

where \(\{X(s)/s \in (0,y)\}\) is a gaussian process with zero mean and a covariance function which turns out to be dependent again on the forth-order moments of the spatial process.
References


