Available at www.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
ELSEVIER
Statistics & Probability Letters 64 (2003) 169–179
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Local linear regression estimation of the variogram

Pilar H. García-Soidán[a,*,1,2], Wenceslao González-Manteiga[b,1,3,4], Manuel Febrero-Bande[b,1,2,3,4]

[a] Departamento de Estatística e I.O., Facultade de Ciencias Sociais, Universidade de Vigo, Campus A Xunqueira, Pontevedra 36005, Spain

[b] Departamento de Estatística e I.O., Facultade de Matemáticas, Universidade de Santiago de Compostela, Campus Sur, Santiago de Compostela 15771, Spain

Received December 2000; received in revised form March 2003

Abstract

In this work, we introduce the local linear semivariogram. Several properties of this estimator are established and compared with those of the Nadaraya–Watson semivariogram. Finally, an adaptation of Shapiro and Botha's fit is applied to produce a valid estimator.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Intrinsic stationarity; Isotropy; Local regression; Variogram

1. Introduction

Let $\{Z(\mathrm{s})/\mathrm{s} \in D \subset \mathbb{R}^d\}$ be a spatial random process, where D is a bounded region with positive d-dimensional volume. A random process is defined as intrinsically stationary if the following conditions are satisfied:

(i) $E[Z(\mathrm{s}_i) - Z(\mathrm{s}_j)] = 0$, for all $\mathrm{s}_i, \mathrm{s}_j \in D$.
(ii) $\mathrm{Var}[Z(\mathrm{s}_i) - Z(\mathrm{s}_j)] = 2\gamma(\mathrm{s}_i - \mathrm{s}_j)$, for all $\mathrm{s}_i, \mathrm{s}_j \in D$.

The function γ is called the semivariogram (and 2γ is the variogram).

* Corresponding author.
E-mail address: pgarcia@uvigo.es (P.H. García-Soidán).
[1] Supported in part by Grant PB98-0182-C02-02 from DGESIC PGC.
[2] Supported in part by Grant ERB-FMRX-CT960095 from European Union.
[3] Supported in part by Grant PGIDT99MA 20701.
[4] Supported in part by Grant BFM2002-03213.

0167-7152/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0167-7152(03)00149-4

In order to guarantee the nonnegativity of the mean-squared prediction errors, the semivariogram is required to satisfy the conditionally negative definiteness property, namely

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \gamma(\mathrm{s}_i - \mathrm{s}_j) \leqslant 0 \quad (m \leqslant n) \tag{1}$$

for any $\{\mathrm{s}_i \in \mathbb{R}^d / 1 \leqslant i \leqslant m\}$ and for any $\{a_i \in \mathbb{R}/1 \leqslant i \leqslant m\}$, such that $\sum_{i=1}^{m} a_i = 0$. A semivariogram satisfying condition (1) is said to be valid.

On the other hand, condition (ii) may be replaced by the more restrictive condition

(ii$'$)$\mathrm{Var}[Z(\mathrm{s}_i) - Z(\mathrm{s}_j)] = 2\gamma(\|\mathrm{s}_i - \mathrm{s}_j\|)$, for all $\mathrm{s}_i, \mathrm{s}_j \in D$.

Then, the intrinsic random process is said to be isotropic.

Estimation of the semivariogram is a fundamental problem in inference for intrinsic random processes, with applications in a broad spectrum of areas such as geostatistics, hydrology, atmospheric science, etc; see, for instance, Cressie (1991) and references therein. For the sake of simplicity, we will focus our attention on the estimation of isotropic semivariograms.

Suppose that n data, $Z(\mathrm{s}_1)$, $Z(\mathrm{s}_2), \ldots, Z(\mathrm{s}_n)$, are collected, at known spatial locations $\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_n$, respectively. The idea of averaging the square differences, $(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2$, may lead to construct nonparametric estimators of the semivariogram as follows:

$$\tilde{\gamma}(s) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(s)\,(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(s)}, \quad s \geqslant 0, \tag{2}$$

where $w_{i,j}(s) \geqslant 0$, for all i, j and $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(s) > 0$.

For instance, take

$$w_{i,j}(s) = \mathrm{I}_{\{\|\mathrm{s}_i - \mathrm{s}_j\| = s\}}$$

to yield the empirical semivariogram, due to Matheron (1963). In Cressie and Hawkins (1980), the square root of the absolute differences, $|Z(\mathrm{s}_i) - Z(\mathrm{s}_j)|$, are appropriately averaged instead to provide robust variogram estimators. When data are irregularly spaced, the latter estimators are usually smoothed by considering a tolerance region around s.

An alternative estimator is suggested in García-Soidán et al. (2003), which results from adapting the Nadaraya–Watson regression estimator to the context of spatial data, trying to mimic the non-parametric kernel covariance estimators proposed in Hall et al. (1994) or in Hall and Patil (1994). The estimator obtained, which will be denoted by $\hat{\gamma}_h(s)$, may be written as given in (2) by selecting $w_{i,j}(s)$ as

$$w_{i,j}(s) = K\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - s}{h}\right), \tag{3}$$

where K denotes a symmetric density function and $h = h_n$ is the bandwidth parameter.

The aim of this paper is to construct a semivariogram estimator by using the local polynomial fitting, since it provides a kernel method with attractive properties; see Fan and Gijbels (1996) for a description of this procedure in a regression setting. For the sake of simplicity, we will apply the local linear estimation, so that we will suppose that locally the semivariogram function can be approximated by

$$\gamma(s) \approx \sum_{k=0}^{1} \gamma^{(k)}(s_0)(s - s_0)^k$$

for s in a neighborhood of s_0, by using Taylor's expansion. The latter polynomial may be fitted locally by a weighted least-squares problem, say, by obtaining β_1 and β_2 that minimizes

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{1}{2}(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2 - \sum_{k=0}^{1}\beta_k(\|\mathrm{s}_i - \mathrm{s}_j\| - s)^k\right] K\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - s}{h}\right). \tag{4}$$

Define

$$w_{0,i,j}(s) = K\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - s}{h}\right)\sum_{k=1}^{n}\sum_{l=1}^{n} K\left(\frac{\|\mathrm{s}_k - \mathrm{s}_l\| - s}{h}\right)(\|\mathrm{s}_k - \mathrm{s}_l\| - s)$$
$$(\|\mathrm{s}_k - \mathrm{s}_l\| - \|\mathrm{s}_i - \mathrm{s}_j\|),$$

$$w_{1,i,j}(s) = K\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - s}{h}\right)\sum_{k=1}^{n}\sum_{l=1}^{n} K\left(\frac{\|\mathrm{s}_k - \mathrm{s}_l\| - s}{h}\right)(\|\mathrm{s}_i - \mathrm{s}_j\| - \|\mathrm{s}_k - \mathrm{s}_l\|).$$

The minimizers of (4) will be given by

$$\hat{\beta}_k = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{k,i,j}(s)(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{0,i,j}(s)}.$$

Then, $\hat{\beta}_k$ may be considered as an estimator of $\gamma^{(k)}(s)$, for $k = 1, 2$.

The latter means that the local linear estimator will be constructed by taking $w_{i,j}(s)$ in (2) as $w_{0,i,j}(s)$. The resulting kernel estimator will be denoted by $\tilde{\gamma}_h(s)$.

From the two kernel methods mentioned above, several features make it advisable the use of the local linear estimation proposed in this work. On one hand, the bias at the boundary is of the same order as that in the interior, unlike the Nadaraya–Watson estimator $\hat{\gamma}_h(s)$; the latter is an important merit, since the boundary can be quite substantial in the number of data points involved. Moreover, boundary modifications may be a difficult task, specially for higher dimensions; for instance, the Nadaraya–Watson estimation requires the use of an specific combination of boundary kernels, given in (6), to retain rates of convergence. In addition, the performance of the local linear semivariogram $\tilde{\gamma}_h(s)$ outside the boundary may be better than that of $\hat{\gamma}_h(s)$, for an appropriate selection of the kernel function, as remarked in Section 3.

The properties of the Nadaraya–Watson estimator have not been developed in this paper, since they are detailed in García-Soidán et al. (2003). Thus, we will only introduce and compare them with those of the local linear estimator. On the other hand, arguments to prove the properties related to the local linear estimator are similar as those given in Hall et al. (1994); therefore, we will basically focus on remarking the particularities of this setting and, more specifically, on deriving the bias and the variance of $\tilde{\gamma}_h(s)$.

Finally, we will take into account in the present paper that all the semivariogram estimators mentioned above do not necessarily satisfy condition (1). The latter means that, previously to be used for spatial prediction, the estimators should be appropriately treated in order to guarantee that requirement. See, for instance, the alternatives proposed in Cressie (1985) or Shapiro and Botha (1991), based on selecting the semivariogram which best fit the data from a family of valid semivariograms considered.

An additional advantage of the local linear fitting is that it provides a practical rule for selection of an appropriate semivariogram model. At this respect, bear in mind that the local linear fitting allows to obtain a direct estimator of $\gamma'(s)$, given by $\hat{\beta}_1$. This estimated derivative may be used as a tool to choose among several families of valid semivariograms, similarly as proposed in Gorsich and Genton (1999), where an adaptation of Shapiro and Botha's method is applied on the empirical estimator for this purpose.

The structure of this paper is as follows. Section 2 is devoted to notation and technical aspects. Asymptotic properties of the local linear estimator are discussed in Section 3 and compared with those of the Nadaraya–Watson estimator; in addition, an adaptation of Shapiro and Botha's is applied to transform the local linear estimator into a valid semivariogram. In the appendix, we will sketch the proofs to derive the bias and the variance of the local linear estimator.

2. Notation and technical details

Firstly, the main hypotheses will be presented.

(S1) K is a univariate, symmetric and bounded density function, with compact support $[-C, C]$ and such that $K(0) > 0$.

The random process $\{Z(\mathrm{s})/\mathrm{s} \in D \subset \mathbb{R}^d\}$ will be required to satisfy the conditions below:

(S2) The random process is intrinsic as well as isotropic, where the semivariogram γ admits three continuous derivatives in a neighborhood of s, for all $s > 0$.

The observation region will be considered to be increasing, in the way proposed in Hall et al. (1994), so that it will allow to achieve consistent estimation.

(S3) $D = D_n = \lambda D_0$, for some $\lambda = \lambda_n$ diverging to $+\infty$ and for some fixed and bounded region $D_0 \subset \mathbb{R}^d$ containing a sphere with positive d-dimensional volume.

As regards the fourth-order moments of the random process, we will require

(S4) There exists a bounded and continuously differentiable function g such that
$\mathrm{Cov}[(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2, (Z(\mathrm{s}_k) - Z(\mathrm{s}_l))^2] = g(\|\mathrm{s}_i - \mathrm{s}_k\|, \|\mathrm{s}_j - \mathrm{s}_l\|, \|\mathrm{s}_i - \mathrm{s}_l\|, \|\mathrm{s}_j - \mathrm{s}_k\|)$.

(S5) Given any positive constant d_1, then
$\int_{\substack{\|\mathrm{s}_1-\mathrm{s}_2\| \leqslant d_1 \\ \|\mathrm{s}_3-\mathrm{s}_4\| \leqslant d_1}} |g(\|\mathrm{s}_1 - \mathrm{s}_3\|, \|\mathrm{s}_2 - \mathrm{s}_4\|, \|\mathrm{s}_1 - \mathrm{s}_4\|, \|\mathrm{s}_2 - \mathrm{s}_3\|)| \, \mathrm{ds}_1 \, \mathrm{ds}_2 \, \mathrm{ds}_3 \, \mathrm{ds}_4 < \infty$.

Hypotheses (S4) and (S5) are not too restrictive in practice. For instance, an isotropic Gaussian process satisfies that $g(x_1, x_2, x_3, x_4) = 2(\gamma(x_1) + \gamma(x_2) - \gamma(x_3) - \gamma(x_4))^2$. Then, both conditions hold for a bounded semivariogram which admits one continuous derivative and is finite-ranged or, even, if it has an asymptotic range with an exponentially decreasing rate of convergence.

The following convergence rates will be assumed:

(S6) $\{h + (nh)^{-1} + \lambda^{-1} + n^{-1}\lambda^d\} \stackrel{n\to\infty}{\to} 0$.

Moreover, for some constant $c > 0$, we will take

$$\lambda^d = cnh + \mathrm{o}(nh). \tag{5}$$

A random design will be assumed for the spatial locations, as suggested in Hall et al. (1994). Then, let f_0 represent a density function defined on D_0, satisfying

(S7) $f_0(\mathrm{x}) > d_2$, for all $\mathrm{x} \in D_0$ and for some positive constant d_2.

Denote by $U_1, U_2, \ldots, U_n$ a random sample of size n from f_0 and by $u_1, u_2, \ldots, u_n$ a realization of it. To model this situation, we will take

(S8) $\mathrm{s}_i = \lambda\, u_i$, for $1 \leqslant i \leqslant n$.

Write f_i, $1 \leqslant i \leqslant 3$, for the density of $(U_1 - U_2, \ldots, U_1 - U_{i+1})$. We will assume

(S9) $f_1(0) > 0$ and f_1 is continuously differentiable in a neighborhood of 0.

(S10) f_2 and f_3 are continuously differentiable in a neighborhood of 0.

The uniform or the normal densities are examples of functions f_0 satisfying conditions (S7), (S9) and (S10).

3. Main results

In what follows, some properties of the local linear regression estimator $\tilde{\gamma}_h(s)$ will be derived; in particular, that it is asymptotically unbiased as well as consistent for $s > 0$.

Theorem 3.1. *Suppose that conditions* (S1)–(S10) *are satisfied. Then, for* $s > 0$

$$E[\tilde{\gamma}_h(s)] = \gamma(s) + \frac{c_{2,K}^2 - c_{1,K}c_{3,K}}{2(c_{0,K}c_{2,K} - c_{1,K}^2)}\,\gamma''(s)\,h^2 + \mathrm{o}(h^2),$$

$$\mathrm{Var}[\tilde{\gamma}_h(s)] = \frac{f_3(0,0,0)B_d(s,s)}{4(f_1(0)A_d)^2}\,\lambda^{-d} + \mathrm{o}(\lambda^{-d} + h^4),$$

where $c_{i,K} = \int_{-C}^{q} z^i K(z)\,\mathrm{d}z$, $q = \min\{sh^{-1}, C\} \in [0, C]$ *and* A_d *and* $B_d(s,s)$ *are defined in* (A.3).

The proof of the theorem above will be sketched in Section A.1.

Remark 3.2. From Theorem 3.1, the bias of $\tilde{\gamma}_h(s)$ is of the order $\mathrm{O}(h^2)$, for all $s > 0$; however, we may only guarantee that the same rate remains valid for the Nadaraya–Watson estimator $\hat{\gamma}_h(s)$ outside the boundary; see García-Soidán et al. (2003). In fact

$$E[\hat{\gamma}_h(s)] = \gamma(s) - \frac{c_{1,K}\gamma'(s)}{c_{0,K}}\,h + \frac{c_{2,K}\gamma''(s)}{2c_{0,K}}h^2 + \mathrm{o}(h^2).$$

According to Remark 3.2, there is a need for an appropriate modification of estimator $\hat{\gamma}_h(s)$ close to the endpoint 0, $s < Ch$; otherwise, bias order would be h rather than h^2. Then, consider two symmetric kernel functions K and L, $K \neq L$ but with the same support $[-C, C]$, and construct an specific linear combination of their corresponding boundary kernels

$$H_q(z) = \frac{c_{0,K}^{-1}K(z) - rc_{0,L}^{-1}L(z)}{1 - r}, \quad \text{where } r = c_{1,K}c_{0,L}(c_{0,K}c_{1,L})^{-1} \neq 1, \tag{6}$$

if $z \in [-C, q]$. Now, write $\hat{\gamma}_{q,h}(s)$ for semivariogram estimator obtained when using

$$w_{i,j}(s) = H_q\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - s}{h}\right)$$

instead of (3) in (2). As proved in García-Soidán et al. (2003), it follows that

$$E[\hat{\gamma}_{q,h}(s)] = \gamma(s) + \frac{c_{2,K}c_{1,L} - c_{1,K}c_{2,L}}{2(c_{0,K}c_{1,L} - c_{1,K}c_{0,L})}\gamma''(s)h^2 + \mathrm{o}(h^2),$$

$$\mathrm{Var}[\hat{\gamma}_{q,h}(s)] = \frac{f_3(0,0,0)B_d(s,s)}{4(f_1(0)A_d)^2}\lambda^{-d} + \mathrm{o}(\lambda^{-d} + h^4).$$

Remark 3.3. Proceeding as above, the same bias order is preserved for all $s > 0$, when using $\hat{\gamma}_{q,h}(s)$. However, the latter gives account of one of the main advantages of estimator $\tilde{\gamma}_h(s)$, say, the absence of boundary effects.

Remark 3.4. The dominant term of the variance of $\tilde{\gamma}_h(s)$ equals that obtained for the Nadaraya–Watson estimator $\hat{\gamma}_{q,h}(s)$. It is also remarkable the fact that the latter variances can be of the order $\lambda^{-d} = \mathrm{O}((nh)^{-1})$, under the convergence rate stated in (5).

Remark 3.5. The relations obtained allow to compare both kernel semivariograms. For this purpose, we may consider the same kernel function K for both estimators; then, $h^{-4}(\gamma''(s))^{-2}[\mathrm{MSE}(\tilde{\gamma}_h(s)) - \mathrm{MSE}(\hat{\gamma}_{q,h}(s))]$ is asymptotically proportional to

$$H_{K,L}(s) = \left[\left(\frac{c_{2,K}^2 - c_{1,K}c_{3,K}}{2(c_{0,K}c_{2,K} - c_{1,K}^2)}\right)^2 - \left(\frac{c_{2,K}c_{1,K}^{-1}c_{1,L} - c_{2,L}}{2(c_{0,K}c_{1,K}^{-1}c_{1,L} - c_{0,L})}\right)^2\right].$$

We may require that there exists $\lim_{q\to C} c_{1,K}^{-1}c_{1,L} = b \in [0, +\infty]$, where $b \neq 1$ according to (6). In case that $b = \infty$, both semivariograms will have a similar behavior outside the boundary, since $H_{K,L}(s) = 0$.

On the other hand, if $b < \infty$, one has for $s \geqslant Ch$,

$$H_{K,L}(s) = \left[\frac{(\int_{-C}^{C} z^2K(z)\,\mathrm{d}z)^2}{4} - \left(\frac{b\int_{-C}^{C} z^2K(z)\,\mathrm{d}z - \int_{-C}^{C} z^2L(z)\,\mathrm{d}z}{2(b-1)}\right)^2\right]. \tag{7}$$

The latter means that the performance of the local linear semivariogram outside the boundary, in comparison with that of the Nadaraya–Watson estimator, will be dependent on the sign of $H_{K,L}(s)$; in particular, an improvement of $\tilde{\gamma}_h(s)$ is expected when (7) is negative.

Next, we will deal with the selection of the bandwidth parameter. We will restrict to the points s outside the boundary; otherwise, the asymptotically optimal bandwidth parameter obtained would be dependent on h itself.

From Theorem 3.1 and condition (5), one has for $s \geqslant Ch$,

$$\mathrm{MSE}(\tilde{\gamma}_h(s)) = \frac{1}{4}\left(\gamma''(s)\int_{-C}^{C} z^2 K(z)\,\mathrm{d}z\right)^2 h^4 + \frac{f_3(0,0,0)B_d(s,s)}{4c(f_1(0)A_d)^2}(nh)^{-1} + \mathrm{o}((nh)^{-1} + h^4).$$

Then, the bandwidth parameter that asymptotically minimizes $\mathrm{MSE}(\tilde{\gamma}_h(s))$ is

$$h_{\mathrm{AMSE}} = \left(\frac{f_3(0,0,0)B_d(s,s)}{4c(f_1(0)A_d\gamma''(s)\int_{-C}^{C} z^2K(z)\,\mathrm{d}z)^2}\right)^{1/5} n^{-1/5}. \tag{8}$$

Remark 3.6. Relation (8) involves unknown characteristics of the random process, so that the bandwidth parameter will be in practice dependent on data; see Remark 3.10.

As pointed out in Section 1, a semivariogram estimator must satisfy the conditionally negative definiteness property. For that reason, an adaptation of Shapiro and Botha's method will be applied next to $\tilde{\gamma}_h(s)$, with the aim of obtaining a valid semivariogram estimator.

Then, give constants t_j $(1 \leqslant j \leqslant m_1)$ as well as distances r_i and weights w_i $(1 \leqslant i \leqslant m_2)$, for some m_1 and m_2; the problem will be reduced to find y_j such that they minimize

$$\sum_{i=1}^{m_2} w_i \left(\tilde{\gamma}_h(r_i) - \sum_{j=1}^{m_1}(1 - g_d(r_i t_j))y_j\right)^2,$$

where

$$g_d(s) = \left(\frac{2}{s}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) J_{(d-2)/2}(s)$$

and J_v represents the Bessel function of the first kind of order v.

Thus, the valid semivariogram estimator has the following explicit representation:

$$\bar{\gamma}_h(s) = \mathrm{x}(s)\mathrm{B}\vec{\tilde{\gamma}}_h, \quad s \geqslant 0$$

where $\vec{\tilde{\gamma}}_h = (\tilde{\gamma}_h(r_1), \ldots, \tilde{\gamma}_h(r_{m_2}))^{\mathrm{T}}$, $\mathrm{B} = (X^{\mathrm{T}}\mathrm{W}X)^{-1}X^{\mathrm{T}}\mathrm{W}$, $\mathrm{W} = \mathrm{diag}(w_i)$ and X is the $m_2 \times m_1$ matrix whose ith row is $\mathrm{x}(r_i) = (1 - g_d(r_i t_1), \ldots, 1 - g_d(r_i t_{m_1}))$.

The following assumption will allow to state the rates of convergence of $\bar{\gamma}_h(s)$:

(S11) $|\mathrm{x}(s)\mathrm{B}\vec{\gamma}|$, $|\mathrm{x}(s)\mathrm{B}\vec{\gamma''}|$ and $\|\mathrm{x}(s)\mathrm{BC}\|_s$ are bounded sequences, for all $s > 0$, where $\vec{\gamma} = (\gamma(r_1), \ldots, \gamma(r_{m_2}))^{\mathrm{T}}$, $\vec{\gamma''} = (\gamma''(r_1), \ldots, \gamma''(r_{m_2}))^{\mathrm{T}}$, CC^{T} is the symmetric matrix of terms $B_d(r_i, r_j)$, as given in (A.3), and $\|\cdot\|_s$ is the supremum norm.

Theorem 3.7. *Suppose that conditions* (S1)–(S11) *are satisfied. Then, for* $s > 0$:

$$E[\bar{\gamma}_h(s)] = \gamma(s) + \mathrm{O}(h^2), \quad \mathrm{Var}[\bar{\gamma}_h(s)] = \mathrm{O}(\lambda^{-d}).$$

The proof of Theorem 3.7 is the same as that of Theorem 3.10 in García-Soidán et al. (2003).

Remark 3.8. From Theorem 3.7, we may conclude that the bias and variance of $\bar{\gamma}_h(s)$ have similar orders as those of $\tilde{\gamma}_h(s)$, for all $s > 0$.

Remark 3.9. Several possibilities are suggested for the selection of weights. For instance, all weights may equal 1 or each weight could be approximately proportional to the inverse of the variance of the estimator, which would correspond to the ordinary or the weighted least-squares criteria, respectively. Moreover, in García-Soidán et al. (2003), it is suggested to iterate the procedure of selection, by updating the weights in each stage.

Remark 3.10. Let h_{AMSE} be the bandwidth parameter obtained in (8). The latter relation is dependent on unknown functions γ'' as well as g, which must be estimated; however, estimation of g turns out to be rather complicated, unless we proceed as if the random process were Gaussian, so that function g is given by $g(x_1,x_2,x_3,x_4)=2(\gamma(x_1)+\gamma(x_2)-\gamma(x_3)-\gamma(x_4))^2$. In that case, we may consider either a parametric model or a valid kernel semivariogram to construct the required estimators (plug-in method).

Appendix A.

To derive the proof of Theorem 3.1, we will use the following lemma.

Lemma 3.11. *Let $\{X_n\}$ be a sequence of uniformly bounded random variables such that $X_n = \mathrm{o}(1)$ a.s. Then, $E[X_n^r] = \mathrm{o}(1)$, for all r.*

A.1. Proof of Theorem 3.1

Write $\mathrm{s}_{i,j} = \|\mathrm{s}_i - \mathrm{s}_j\|$ and define

$$a(s) = \sum_{i,j,k,l} K\left(\frac{\mathrm{s}_{i,j}-s}{h}\right) K\left(\frac{\mathrm{s}_{k,l}-s}{h}\right)(\mathrm{s}_{k,l}-s)(\mathrm{s}_{k,l}-\mathrm{s}_{i,j}),$$

$$b(s) = \sum_{i,j,k,l} K\left(\frac{\mathrm{s}_{i,j}-s}{h}\right) K\left(\frac{\mathrm{s}_{k,l}-s}{h}\right)(\mathrm{s}_{k,l}-s)(\mathrm{s}_{k,l}-\mathrm{s}_{i,j})\gamma(\mathrm{s}_{i,j}),$$

$$c(s) = \sum_{i_1,j_1,k_1,l_1,i_2,j_2,k_2,l_2} K\left(\frac{\mathrm{s}_{i_1,j_1}-s}{h}\right) K\left(\frac{\mathrm{s}_{i_2,j_2}-s}{h}\right) K\left(\frac{\mathrm{s}_{k_1,l_1}-s}{h}\right) K\left(\frac{\mathrm{s}_{k_2,l_2}-s}{h}\right)$$
$$(\mathrm{s}_{k_1,l_1}-s)(\mathrm{s}_{k_1,l_1}-\mathrm{s}_{i_1,j_1})(\mathrm{s}_{k_2,l_2}-s)(\mathrm{s}_{k_2,l_2}-\mathrm{s}_{i_2,j_2})g(\mathrm{s}_{i_1,i_2},\mathrm{s}_{j_1,j_2},\mathrm{s}_{i_1,j_2},\mathrm{s}_{j_1,i_2}).$$

According to condition (S8), we take $\mathrm{s}_i = \lambda u_i$; then

$$E[\tilde{\gamma}_h(s)/U_1,\ldots,U_n] - \gamma(s) = \frac{b(s)-a(s)\gamma(s)}{a(s)},$$

$$\mathrm{Var}[\tilde{\gamma}_h(s)/U_1,\ldots,U_n] = \frac{c(s)}{(2a(s))^2}. \tag{A.1}$$

At the end of this section, we will check that the following order holds for $s > 0$:

$$a(s) = (f_1(0))^2 s^{2(d-1)} A_d^2 (c_{0,K} c_{2,K} - c_{1,K}^2) n^4 \lambda^{-2d} h^4 + \mathrm{o}(n^4 \lambda^{-2d} h^4) \quad \text{a.s.} \tag{A.2}$$

and, similarly, we might obtain that

$$b(s) - a(s)\gamma(s) = \tfrac{1}{2} (f_1(0))^2 s^{2(d-1)} A_d^2 (c_{2,K}^2 - c_{1,K} c_{3,K}) \gamma''(s) n^4 \lambda^{-2d} h^6 + \mathrm{o}(n^4 \lambda^{-2d} h^6) \quad \text{a.s.},$$

$$c(s) = (f_1(0))^2 f_3(0,0,0) s^{4(d-1)} B_d(s,s) A_d^2 (c_{0,K} c_{2,K} - c_{1,K}^2)^2 n^8 \lambda^{-5d} h^8 + \mathrm{o}(n^8 \lambda^{-5d} h^8) \quad \text{a.s.},$$

where

$$A_d = \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} J_d(\theta_1, \ldots, \theta_{d-1})\, \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_{d-2}\, \mathrm{d}\theta_{d-1}, \tag{A.3}$$

$$\begin{aligned} B_d(s,s') = & \int_0^{+\infty} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} J_d(\theta_{1,1}, \ldots, \theta_{d-1,1}) \\ & J_d(\theta_{1,2}, \ldots, \theta_{d-1,2}) J_d(\theta_{1,3}, \ldots, \theta_{d-1,3}) \\ & g\left(t, \left\| t\cos\theta_{1,3} + s'\cos\theta_{1,2} - s\cos\theta_{1,1}, \ldots, t\prod_{j=0}^{d-1} \sin\theta_{j,3} \right.\right. \\ & \left. + s'\prod_{j=0}^{d-1} \sin\theta_{j,2} - s\prod_{j=0}^{d-1} \sin\theta_{j,1} \right\|, \\ & \left\| t\cos\theta_{1,3} + s'\cos\theta_{1,2}, \ldots, t\prod_{j=0}^{d-1} \sin\theta_{j,3} + s'\prod_{j=0}^{d-1} \sin\theta_{j,2} \right\|, \\ & \left. \left\| t\cos\theta_{1,3} - s\cos\theta_{1,1}, \ldots, t\prod_{j=0}^{d-1} \sin\theta_{j,3} - s\prod_{j=0}^{d-1} \sin\theta_{j,1} \right\| \right) \\ & \mathrm{d}t\, \mathrm{d}\theta_{1,1} \ldots \mathrm{d}\theta_{d-2,1}\, \mathrm{d}\theta_{d-1,1}\, \mathrm{d}\theta_{1,2} \ldots \mathrm{d}\theta_{d-2,2}\, \mathrm{d}\theta_{d-1,2}\, \mathrm{d}\theta_{1,3} \ldots \mathrm{d}\theta_{d-2,3}\, \mathrm{d}\theta_{d-1,3} \end{aligned}$$

and $J_d(\theta_1, \ldots, \theta_{d-1}) = (\sin\theta_1)^{d-2} (\sin\theta_2)^{d-3} \ldots \sin\theta_{d-2}$.

Then, by considering the latter relations and applying Lemma 3.11 to (A.1), one has that

$$E[\tilde{\gamma}_h(s)] - \gamma(s) = E[E[\tilde{\gamma}_h(s)/U_1, \ldots, U_n]] - \gamma(s) = \frac{c_{2,K}^2 - c_{1,K} c_{3,K}}{2(c_{0,K} c_{2,K} - c_{1,K}^2)} \gamma''(s)\, h^2 + \mathrm{o}(h^2)$$

together with

$$\begin{aligned}\mathrm{Var}[\tilde{\gamma}_h(s)] &= \mathrm{Var}[E[\tilde{\gamma}_h(s)/U_1,\ldots,U_n]] + E[\mathrm{Var}[\tilde{\gamma}_h(s)/U_1,\ldots,U_n]]\\ &= \mathrm{o}(h^4) + \frac{f_3(0,0,0)B_d(s,s)}{4(f_1(0)A_d)^2}\lambda^{-d} + \mathrm{o}(\lambda^{-d}),\end{aligned}$$

which would complete the proof of Theorem 3.1.

Finally, we will give an sketch of the proof of relation (A.2). To do the latter, define

$$a_m(s) = \sum_{i,j} K\left(\frac{\mathrm{s}_{i,j}-s}{h}\right)(\mathrm{s}_{i,j}-s)^m = \sum_{i\neq j} K\left(\frac{\mathrm{s}_{i,j}-s}{h}\right)(\mathrm{s}_{i,j}-s)^m \quad (m=0,1,2)$$

for all large n, since the kernel function K is compactly supported.

For each m, write

$$W_{i,j} = K\left(\frac{\lambda\|U_i-U_j\|-s}{h}\right)(\lambda\|U_i-U_j\|-s)^m,$$

$$\alpha_1(y) = E[W_{i,j}/U_i = y],\ \alpha_2(y) = E[W_{i,j}/U_j = y],\ \alpha = E[W_{i,j}],$$

$$D_{i,j} = W_{i,j} - \alpha_1(U_i) - \alpha_2(U_j) + \alpha, \quad Z_j = \sum_{i=1}^{j-1} D_{i,j}.$$

Observe that $\alpha = f_1(0)s^{d-1}A_d c_{m,K}\lambda^{-d}h^{m+1} + \mathrm{o}(\lambda^{-d}h^{m+1})$.

With this notation, $a_m(s) - n(n-1)\alpha$ represents an observed value of

$$\sum_{i\neq j}(W_{i,j}-\alpha) = \sum_{j=2}^{n} Z_j + (n-1)\sum_{k=1}^{2}\sum_{i=1}^{n}(\alpha_k(U_i)-\alpha).$$

Next, take into account that

$$E[Z_j/U_1,\ldots,U_{j-1}] = 0 \tag{A.4}$$

and that the random variables $Z_2,\ldots,Z_n$ may be considered as differences of the martingales $S_2,\ldots,S_n$, where

$$S_2 = Z_2,\ S_3 = Z_2 + Z_3,\ldots,S_n = Z_2 + \cdots + Z_n. \tag{A.5}$$

The random variables U_1 and U_2 are continuous and bounded, with common density f_0 satisfying (S7) and (S9). In consequence, there exists a positive constant C_m, such that

$$\sup_y E[D_{i,j}^2/U_j = y] \leqslant C_m\lambda^{-d}h^{m+1}\max\{1, s^{2(d-1)}\}. \tag{A.6}$$

Now, by using (A.4)–(A.6), we may proceed as in the proof of Theorem 3.1 of Hall et al. (1994) to conclude that

$$(n^2\lambda^{-d}h^{m+1})^{-1}\sum_{i\neq j}(W_{i,j}-\alpha)\to 0 \quad \text{a.s.}$$

Hence

$$\begin{aligned} a_m(s) &= \sum_{i\neq j}(W_{i,j}-\alpha)+n(n-1)\alpha \\ &= f_1(0)s^{d-1}A_d c_{m,K} n^2\lambda^{-d}h^{m+1} \\ &\quad + \mathrm{o}(n^2\lambda^{-d}h^{m+1}) \quad \text{a.s.} \end{aligned}$$

Since $a(s)=a_0(s)\ a_2(s)-(a_1(s))^2$, it follows relation (A.2).

References

Cressie, N., 1985. Fitting variogram models by weighted least squares. Math. Geol. 17, 563–586.

Cressie, N., 1991. Statistics for Spatial Data. Wiley, New York.

Cressie, N., Hawkins, D., 1980. Robust estimation of the variogram I. Math. Geol. 12, 115–125.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

García-Soidán, P.H., Febrero-Bande, M., González Manteiga, W., 2003. Nonparametric kernel estimation of an isotropic semivariogram. J. Statist. Plann. Inference, to be published.

Gorsich, D.J., Genton, M., 1999. Variogram model selection via nonparametric derivative estimation. Math. Geol. 32, 249–270.

Hall, P., Patil, P., 1994. Properties of nonparametric estimators of autocovariance for stationary random fields. Probab. Theory Related Fields 99, 399–424.

Hall, P., Fisher, I., Hoffmann, B., 1994. On the nonparametric estimation of covariance functions. Ann. Statist. 22 (4), 2115–2134.

Matheron, G., 1963. Principles of geostatistics. Econom. Geol. 58, 1246–1266.

Shapiro, A., Botha, J., 1991. Variogram fitting with a general class of conditionally nonnegative definite functions. Comput. Statist. Data Anal. 11, 87–96.