Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 121 (2004) 65–92

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Nonparametric kernel estimation of an isotropic variogram

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

[a] *Departamento de Estatística e Investigación Operativa, Facultade de Ciencias Sociais, Campus A Xunqueira, Universidade de Vigo, C.P. 36005, Pontevedra, Spain*

[b] *Departamento de Estatística e Investigación Operativa, Facultade de Matemáticas, Campus Sur, Universidade de Santiago, C.P. 15771, Santiago de Compostela, Spain*

Received 20 July 2001; accepted 4 November 2002

Abstract

In this paper, we propose nonparametric kernel estimators of the semivariogram, under the assumption of isotropy. At first, a symmetric kernel is considered in order to construct a consistent estimator, so that the selection of the bandwidth parameter is treated via the MSE or the MISE criteria. Next, the use of a boundary kernel will be suggested in order to obtain satisfactory estimates near the semivariogram endpoint. In all cases, an adaptation of Shapiro and Botha's fit is proposed to produce valid semivariogram estimators. Finally, we describe a numerical study carried out to illustrate the performance of the kernel estimators.

MSC: primary 62G05; secondary: 62G10

Keywords: Bandwidth parameter; Boundary; Conditionally negative definite; Intrinsic stationarity; Isotropy; Kernel; Random process; Variogram

1. Introduction

Let $\{Z(s)/s \in D \subset \mathbb{R}^d\}$ be a spatial random process, where D is a bounded region with positive d-dimensional volume. Suppose that n data, $Z(s_1), Z(s_2), \ldots, Z(s_n)$, are collected, at known spatial locations $s_1, s_2, \ldots, s_n$, respectively.

* 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.

0378-3758/$ - see front matter
doi:10.1016/S0378-3758(02)00507-4

A random process is defined as intrinsic or intrinsically stationary if the following conditions are satisfied:

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

The latter assumptions convey to the fact that the first two moments of the difference $Z(s_i) - Z(s_j)$ depend only on the relative location, $s_i - s_j$, of the two variables. The function γ is called the semivariogram (and 2γ is the variogram).

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 (1993) and references therein. In particular, the semivariogram estimation plays a crucial role for spatial prediction, since the kriging equations depend on the semivariogram function which is, in general, unknown.

The semivariogram γ must satisfy the conditionally negative definiteness property:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \gamma(s_i - s_j) \leqslant 0 \quad (m \leqslant n) \tag{1}$$

for any $\{s_i \in \mathbb{R}^d / 1 \leqslant i \leqslant m\}$ and for any $\{a_i \in \mathbb{R}/1 \leqslant i \leqslant m\}$, such that $\sum_{i=1}^{m} a_i = 0$. Otherwise, negative mean-squared prediction errors may be obtained; therefore, property (1) will be also required from the semivariogram estimator.

Condition (ii) may be replaced by the more restrictive condition:

(ii′) $\mathrm{Var}[Z(s_i) - Z(s_j)] = 2\gamma(\|s_i - s_j\|)$, for all $s_i, s_j \in D$.

Then, the intrinsic random process is said to be isotropic. In this case, the first two moments of $Z(s_i) - Z(s_j)$ will be dependent only on the distance of the spatial locations, $\|s_i - s_j\|$.

For the sake of simplicity, we have considered isotropic models in this work; however, this assumption is not so restrictive in practice. In fact, a different semivariogram may be fit in each of several directions, in case that the random process proves to be anisotropic; besides, a semivariogram estimator may be constructed as given in (10), so that not only the distances between spatial locations but also their directions have an effect on the estimation.

A natural and unbiased estimator based on the method of moments, due to Matheron (1963), is the empirical semivariogram given by

$$\hat{\gamma}(s) = \frac{1}{2|N(s)|} \sum_{(s_i, s_j) \in N(s)} (Z(s_i) - Z(s_j))^2, \quad s \geqslant 0, \tag{2}$$

where $|N(s)|$ is the number of pairs in

$$N(s) = \{(s_i, s_j)/\|s_i - s_j\| = s,\ 1 \leqslant i, j \leqslant n\}. \tag{3}$$

When data are irregularly spaced, the latter estimator is usually smoothed by considering a tolerance region $T(s)$ around s, rather than $N(s)$.

An alternative estimator has been proposed in Cressie and Hawkins (1980), by using instead the root square of the differences $|Z(s_i) - Z(s_j)|$; however, this estimator can be destroyed by a single outlier in the data. In this sense, a more robust variogram estimator is suggested in Genton (1998a), based on a highly robust estimator of scale.

The semivariogram estimators mentioned above cannot be used directly for spatial prediction, since condition (1) typically lacks. In that case, the procedure of estimation should be modified in order to obtain a semivariogram estimator with this requirement. One idea is based on first choosing a parametric family, as proposed in Cressie (1985) or in Genton (1998b), and then selecting that semivariogram in the family considered which best fit the data. However, care must be taken about judging the quality of a parametric estimator obtained from the empirical semivariogram, due in part to the fact that the latter estimator is a poor tool for distinguishing the degree of smoothness of a differentiable process; see Stein (1999).

An alternative may be that of considering a broad class of valid semivariograms to fit the data, not depending on a small number of parameters. An explicit discussion of such a method is given in Shapiro and Botha (1991) up to the three-dimensional case, although it can be easily extended to higher dimensions; see, for instance, Christakos (1992). This mechanism of estimation has been used in Gorsich and Genton (1999) as a tool to choose among valid variogram models.

A different procedure in the context of covariance estimation is used in Hall et al. (1994) or in Hall and Patil (1994), for the particular case $d=1$ or a general dimension d, respectively. In both papers, valid kernel estimators for the covariance function are obtained by inverting the Fourier transform of an appropriate correction of the original estimator.

Both methods above are based on the spectral representation of positive-definite functions derived from Bochner's theorem. In particular, Shapiro and Botha's method requires a choice of nodes which will play an important role in the estimator obtained. There are several criteria to do this; for instance, we may consider a number of equispaced nodes, although it might convey to spurious oscillations of the estimator. Another idea may be that of choosing the nodes as the roots of some Bessel functions, as suggested in Genton and Gorsich (2002); the latter procedure produces an orthogonal discretization so that a very small number of nodes are necessary to obtain a good nonparametric fit.

In this work, we analyze the Nadaraya–Watson estimators of the semivariogram proposed in Febrero-Bande et al. (1998) and applied in Diblasi and Bowman (2001). The referred estimators were constructed trying to mimic the nonparametric kernel covariance estimators proposed in Hall et al. (1994) and in Hall and Patil (1994), but adapted to the isotropic setting.

We will prove that the kernel semivariograms enjoy good properties, such as asymptotically unbiasedness or consistency. In fact, the assumption of isotropy will have an effect on the convergence rates achieved for the bias and variance of the semivariogram estimators, extending to this setting those obtained in Hall et al. (1994) for the covariance estimators in the one-dimensional case. In addition, we will treat the problem of the unsatisfactory behavior of the estimators near endpoints, so that we will apply the procedure given in Kyung-Joon and Shucany (1998) for estimation near the boundary.

The next step must be that of transforming the kernel semivariograms into valid estimators. In this respect, Shapiro and Botha's method is sometimes criticized because it does not produce smooth estimators, although we have adapted this procedure because of its simplicity. Proceeding in this way, we have extended the work in Shapiro and Botha (1991) or Christakos (1992) in the sense that the kernel semivariograms may be also considered to produce the data to be fit, besides the empirical or the robust estimators, with the advantage that the kernel estimators are properly defined at all s up to some magnitude. The convergence rates achieved by the valid estimators are also established, which is not always clear, as remarked in Hall et al. (1994).

The contents of this paper has been organized as follows. Section 2 is devoted to notation and technical aspects. The Nadaraya–Watson estimator has been introduced in Section 3, where some of its properties outside the boundary have been established; the semivariogram estimation close to the endpoint 0 is studied in Section 3.1. Section 3.2 describes an adaptation of Shapiro and Botha's method applied to obtain valid estimators which may be used for bandwidth selection, as discussed in Section 3.3. The numerical studies are detailed in Section 4 and the proofs are developed in Section 5.

2. Notation and technical details

Firstly, the main hypotheses to be required will be presented.

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

We will impose that the random process $\{Z(s)/s \in D \subset \mathbb{R}^d\}$ is intrinsic as well as isotropic, where the semivariogram γ satisfies the following property:

(S2) γ admits three continuous derivatives in a neighborhood of s, for all $s > 0$.

In addition, the observation region will be considered to be increasing, in the way proposed in Hall et al. (1994) and in Hall and Patil (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.

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

(S4) $f_0(x) > d_1$, for all $x \in D_0$ and for some positive constant d_1.

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:

(S5) $s_i = \lambda u_i$, for $1 \leqslant i \leqslant n$.

Write f_i, $1 \leqslant i \leqslant 3$, for the respective densities of $U_1 - U_2$, $(U_1 - U_2, U_1 - U_3)$ and $(U_1 - U_2, U_1 - U_3, U_1 - U_4)$. We will ask the following assumptions for these densities:

(S6) $f_1(0) > 0$ and f_1 is continuously differentiable in a neighborhood of 0.
(S7) f_2 and f_3 are continuously differentiable in a neighborhood of 0.

Finally, some hypotheses will be required as regards the fourth-order moments of the random process:

(S8) There exists a bounded and continuously differentiable function $g : \mathbb{R}^4 \to \mathbb{R}$ satisfying that:

$$\begin{aligned}&\mathrm{Cov}[(Z(s_i) - Z(s_j))^2, (Z(s_k) - Z(s_l))^2]\\ &\quad = g(\|s_i - s_k\|, \|s_j - s_l\|, \|s_i - s_l\|, \|s_j - s_k\|).\end{aligned}$$

(S9) Given any positive constant d_2, then

$$\int_{\substack{\|s_1-s_2\| \leqslant d_2\\ \|s_3-s_4\| \leqslant d_2}} |g(\|s_1-s_3\|, \|s_2-s_4\|, \|s_1-s_4\|, \|s_2-s_3\|)|\, \mathrm{d}s_1\, \mathrm{d}s_2\, \mathrm{d}s_3\, \mathrm{d}s_4 < \infty.$$

For instance, in the context of a Gaussian process, one has that

$$\begin{aligned}&\mathrm{Cov}[(Z(s_i) - Z(s_j))^2, (Z(s_k) - Z(s_l))^2]\\ &\quad = 2(\gamma(\|s_i - s_k\|) + \gamma(\|s_j - s_l\|) - \gamma(\|s_i - s_l\|) - \gamma(\|s_j - s_k\|))^2.\end{aligned}$$

Hence, take

$$g(x_1, x_2, x_3, x_4) = 2(\gamma(x_1) + \gamma(x_2) - \gamma(x_3) - \gamma(x_4))^2. \tag{4}$$

If the semivariogram γ is bounded and admits one continuous derivative, then condition S8 will be satisfied. Moreover, we may require that γ is finite ranged or that it has an asymptotic range with an exponentially decreasing rate of convergence; then, hypothesis S9 will hold for a Gaussian process.

3. Main results

Let $\{Z(s)/s \in D \subset \mathbb{R}^d\}$ be an intrinsic and isotropic random process. Denote by $Z(s_1), Z(s_2), \ldots, Z(s_n)$, n values of the process observed at spatial locations $s_1, s_2, \ldots, s_n$, respectively.

The Nadaraya–Watson estimator will be defined as follows:

$$\hat{\gamma}_h(s) = \frac{\sum_{i=1}^n \sum_{j=1}^n K(\frac{s-\|s_i-s_j\|}{h})(Z(s_i) - Z(s_j))^2}{2\sum_{i=1}^n \sum_{j=1}^n K(\frac{s-\|s_i-s_j\|}{h})}, \quad s \geqslant 0, \tag{5}$$

where $h = h_n$ represents the bandwidth parameter.

The following convergence rates will be assumed:

$$\{h+(nh)^{-1}+\lambda^{-1}\}\stackrel{n\to\infty}{\longrightarrow}0. \tag{6}$$

As suggested in Hall et al. (1994), we will also require that $\lambda^d = \mathrm{o}(n)$, to bring about a large number of pairs, (s_i, s_j), whose distance, $\|s_i - s_j\|$, is close to s, for any given s. Moreover, for some constant $c > 0$, we will take

$$\lambda^d = cnh + \mathrm{o}(nh). \tag{7}$$

Bear in mind that the semivariogram domain is restricted to the nonnegative values, in case of isotropy, and that the kernel function operates on the distances $\|s_i - s_j\| \in [s - Ch, s + Ch]$. Thus, some properties of estimator (5) will be derived outside the boundary, $s \geqslant Ch$; in particular, we will prove that $\hat{\gamma}_h(s)$ is asymptotically unbiased as well as consistent. The arguments used in the proofs of both properties are based on the fact that interval $[s - Ch, s + Ch]$ is wholly contained within the domain of γ, obtaining the order h^2 for its bias. As we will give account in Section 3.1, the latter bias order does not remain valid close to the endpoint, $s < Ch$; however, by using an specific combination of boundary kernels, it is possible to achieve the same rate of convergence.

Theorem 3.1. *Assume that conditions* S1–S6 *are satisfied. In addition, suppose the convergence rates stated in* (6) *and* (7). *Then, for $s \geqslant Ch$, one has*

$$E[\hat{\gamma}_h(s)] = \gamma(s) + \tfrac{1}{2} c_K \gamma''(s) h^2 + \mathrm{o}(h^2),$$

where $c_K = \int_{-C}^{C} t^2 K(t)\,\mathrm{d}t$.

Theorem 3.2. *Assume the hypotheses required in Theorem* 3.1. *In addition, suppose that conditions* S7–S9 *are satisfied. It follows for $s \geqslant Ch$*:

$$\begin{aligned}\mathrm{Var}[\hat{\gamma}_h(s)] &= \frac{d_K g(0,0,s,s)}{2 f_1(0) s^{d-1} A_d} n^{-2}\lambda^d h^{-1} + \frac{f_2(0,0) B_d(s)}{(f_1(0) A_d)^2} n^{-1} + \frac{f_3(0,0,0) C_d(s)}{4 (f_1(0) A_d)^2} \lambda^{-d} \\ &\quad + \mathrm{o}(n^{-2}\lambda^d h^{-1} + n^{-1} + \lambda^{-d} + h^4) \\ &= \frac{f_3(0,0,0) C_d(s)}{4 (f_1(0) A_d)^2} \lambda^{-d} + \mathrm{o}(\lambda^{-d} + h^4),\end{aligned}$$

where A_d, $B_d(s)$ and $C_d(s)$ are as given in (21), (30) *and* (32), *respectively, and* $d_K = \int_{-C}^{C} (K(t))^2\,\mathrm{d}t$.

The proofs of the above results are sketched in Sections 5.1 and 5.2.

Remark 3.3. Theorem 3.2 makes it clear that the order of the variance of $\hat{\gamma}_h(s)$ is $\lambda^{-d} = \mathrm{O}((nh)^{-1})$, if we assume the convergence rates stated in (6) and (7).

Now, we will establish the order of $\mathrm{Cov}[\hat{\gamma}_h(s)\hat{\gamma}_h(s')]$, which will be used in Section 3.2 to state the properties of the valid estimator.

Theorem 3.4. *Assume the hypotheses required in Theorem* 3.2. *Then, for* $s, s' \geqslant Ch$:

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

where $C_d'(s,s')$ *is given in* (34).

See Section 5.3 for a sketch of this proof.

Next, the problem of estimation of the bandwidth parameter will be treated. From the different criteria proposed in the classical theory of kernel method, we have considered those based on minimizing asymptotically the MSE or the MISE of kernel-type estimators, in order to determine the local or the global bandwidth parameter, respectively, which will be stated in terms of both the sample size n and the scale factor λ. The expressions obtained will be dependent on unknown characteristics of the random process, whose specific estimation will be studied in Section 3.3.

At first, we will deal with the selection of the local bandwidth parameter. From Theorems 3.1 and 3.2, one has for $s \geqslant Ch$:

$$\begin{aligned}\mathrm{MSE}[\hat{\gamma}_h(s)] &= (E[\hat{\gamma}_h(s)] - \gamma(s))^2 + \mathrm{Var}[\hat{\gamma}_h(s)]\\ &= \frac{c_K^2(\gamma''(s))^2}{4}\,h^4 + \frac{f_3(0,0,0)C_d(s)}{4c(f_1(0)A_d)^2}\,(nh)^{-1} + \mathrm{o}(h^4 + (nh)^{-1})\end{aligned}$$

by considering the convergence rates assumed in (6) and (7).

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

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

In particular, when the random process is Gaussian, function g is as given in (4) and may be used to compute $C_d(s)$ in h_{AMSE}.

With this selection of the bandwidth parameter, it follows that:

$$\mathrm{MSE}[\hat{\gamma}_{h_{\mathrm{AMSE}}}(s)] = \mathrm{O}(n^{-4/5}).$$

Suppose instead that our interest is to consider a global bandwidth parameter; in this case, we should minimize

$$\mathrm{MISE}[\hat{\gamma}_h] = \int_R (E[\hat{\gamma}_h(s)] - \gamma(s))^2\,\mathrm{d}s + \int_R \mathrm{Var}[\hat{\gamma}_h(s)]\,\mathrm{d}s$$

for some $R \subset [0, +\infty)$. For instance, we may take $R = [a, m]$, where $m = \sup\{\|s_i - s_j\| / s_i, s_j \in D\}$ and some constant a, $0 < a < m$.

By similar arguments as those used above, when considering the MSE criterion, we obtain

$$h_{\mathrm{AMISE}} = \left(\frac{f_3(0,0,0)\int_R C_d(s)\,\mathrm{d}s}{4c(c_K f_1(0)A_d)^2\int_R(\gamma''(s))^2\,\mathrm{d}s}\right)^{1/5} n^{-1/5}. \tag{9}$$

Remark 3.5. In case of anisotropy, an estimator may be defined, as proposed in Hall et al. (1994) for the univariate case. The point is to construct a kernel estimator where the differences between spatial locations are considered, so that not only their distances but also their respective directions have an effect on the estimation, as follows:

$$\check{\gamma}_h(s) = \frac{\sum_{i=1}^n \sum_{j=1}^n \mathscr{K}(\frac{s-(s_i-s_j)}{h})(Z(s_i)-Z(s_j))^2}{2\sum_{i=1}^n \sum_{j=1}^n \mathscr{K}(\frac{s-(s_i-s_j)}{h})}, \quad s \in \mathbb{R}^d, \tag{10}$$

where $\mathscr{K}$ denotes a d-variate kernel density.

Although we have not developed asymptotic theory for the latter estimator, the main difference with respect to estimator (5) is the distinct influence of the dimension d of the observation region in the convergence rates achieved.

3.1. Estimation near the endpoint

The problem of the unsatisfactory behavior of the estimators near endpoints has been treated in the context of curve estimation; see, for example, Müller (1993), Jones (1993) and Kyung-Joon and Shucany (1998). The procedures proposed are based on appropriate modifications of the estimator considered, in order to achieve consistency as well as to retain rates of convergence.

As pointed out in Section 3, the semivariogram domain is restricted to the nonnegative values, under the assumption of isotropy, so that the zero distance is an endpoint. Thus, we will apply the procedure given in Kyung-Joon and Shucany (1998) for estimation near the boundary, which suggests the use of a boundary kernel function instead of a symmetric one.

At first, bear in mind that the kernel density in (5) only operates on the distances $\|s_i - s_j\| \in [s - Ch, s + Ch]$ and this interval contains negative values for those points s in the boundary, $s < Ch$. Therefore, we will introduce the following index:

$$q = \min\left\{\frac{s}{h}, C\right\} \in [0, C]$$

and take $c_{i,K} = \int_{-C}^{q} z^i K(z)\,\mathrm{d}z$. We may conclude that the following order holds for $s > 0$.

Theorem 3.6. *Assume the hypotheses required in Theorem* 3.1. *Then, one has for* $s > 0$:

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

A sketch of this proof will be outlined in Section 5.4.

Theorem 3.6 gives account of the fact that the bias of $\hat{\gamma}_h(s)$ is of order h rather than h^2 for $s < Ch$, since $c_{1,K} \neq 0$. Then, proceeding as in Kyung-Joon and Shucany (1998), we will consider a new symmetric kernel function L, satisfying the following conditions:

(S10) L is a univariate, symmetric and bounded density function, $L \neq K$, with compact support $[-C, C]$, such that $L(0) > 0$ and $r = c_{1,K}c_{0,L}(c_{0,K}c_{1,L})^{-1} \neq 1$.

Now, we will construct an specific linear combination of both kernels

$$H_q(z) = \frac{c_{0,K}^{-1}K(z) - rc_{0,L}^{-1}L(z)}{1-r} \quad \text{if } z \in [-C, q].$$

Write $\hat{\gamma}_{q,h}(s)$ for the estimator obtained when using H_q instead of K in (5)

$$\hat{\gamma}_{q,h}(s) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} H_q(\frac{s-\|s_i-s_j\|}{h})(Z(s_i)-Z(s_j))^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} H_q(\frac{s-\|s_i-s_j\|}{h})}, \quad s \geqslant 0. \tag{11}$$

This particular selection of the boundary kernel H_q produces a semivariogram estimator that makes it negligible the term of order h in the bias, which appears for a point s in the boundary when using instead a symmetric kernel. Therefore, $\hat{\gamma}_{q,h}(s)$ preserves the same convergence orders for all $s > 0$, as stated below.

Theorem 3.7. *Assume the hypotheses required in Theorem* 3.2. *In addition, suppose that condition* S10 *is satisfied. It follows for* $s, s' > 0$:

$$E[\hat{\gamma}_{q,h}(s)] = \gamma(s) + \frac{c'_{H_q}\gamma''(s)}{2}h^2 + o(h^2),$$

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

$$\text{Cov}[\hat{\gamma}_{q,h}(s), \hat{\gamma}_{q,h}(s')] = \frac{f_3(0,0,0)C'_d(s,s')}{4(f_1(0)A_d)^2}\lambda^{-d} + o(\lambda^{-d} + h^4),$$

where $c'_{H_q} = (c_{2,K}c_{1,L} - c_{1,K}c_{2,L})(c_{0,K}c_{1,L} - c_{1,K}c_{0,L})^{-1}$.

See Section 5.5 for a proof of this theorem.

Remark 3.8. An immediate consequence of Theorem 3.7 is that, for all $s > 0$, the bandwidth parameters that asymptotically minimize MSE and MISE of $\hat{\gamma}_{q,h}(s)$ are as given in (8) and (9), respectively, just replacing constant c_K by its corresponding counterpart c'_{H_q}.

Remark 3.9. In addition, for appropriate compactly supported and symmetric kernel functions K and L, the semivariogram estimator constructed from the linear combination H_q may equal that obtained from K outside the boundary, $s \geqslant Ch$; in fact, a sufficient condition for the latter property to be held would be that $r = 0$ when $q = C$. For instance, take K and L to be the Epanechnikov kernel and the uniform kernel of order 2, respectively, with compact support $[-1, 1]$; therefore, $r = 3(1-q)(2q-4)^{-1}$, which equals 0 if $q = 1$, so that $H_q = K$ outside the boundary.

3.2. Valid kernel estimator

Recall that a semivariogram estimator will be required to satisfy condition (1). In this sense, we have adapted Shapiro and Botha's method in order to obtain a valid kernel estimator satisfying the latter property.

Since γ satisfies the conditionally negative definiteness property, condition (1), it follows:

$$\gamma(s) = \int_0^{+\infty} (1 - g_d(st))\,\mathrm{d}F(t), \qquad s \geqslant 0$$

for some bounded function F, which is monotonically increasing on $[0, +\infty]$, where

$$g_d(s) = \left(\frac{2}{s}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) \mathscr{J}_{(d-2)/2}(s) \tag{12}$$

and $\mathscr{J}_v$ represents the Bessel function of the first kind of order v.

Therefore, given $C_1(n) > 0$ it follows that:

$$\left|\gamma(s) - \int_0^{C_1(n)} (1 - g_d(st))\,\mathrm{d}F(t)\right| \leqslant \int_{C_1(n)}^{+\infty} |1 - g_d(st)|\,\mathrm{d}F(t) \leqslant C_2 \int_{C_1(n)}^{+\infty} \mathrm{d}F(t)$$

for some positive constant C_2 independent of n.

On the other hand, take $m_1 = m_1(n) = C_1(n)\delta_n^{-1}$ and $t_j = j\delta_n$, for some $\delta_n > 0$ and $1 \leqslant j \leqslant m_1(n)$. Then, one has

$$\left|\int_0^{C_1(n)} (1 - g_d(st))\,\mathrm{d}F(t) - \sum_{j=1}^{m_1} (1 - g_d(st_j))z_j\right| \leqslant C_3\delta_n(F(C_1(n)) - F(0))$$

for some positive constant C_3 independent of n, where $z_j = \int_{(j-1)\delta_n}^{j\delta_n} \mathrm{d}F(t)$.

In consequence, one has

$$\gamma(s) = \sum_{j=1}^{m_1} (1 - g_d(st_j))z_j + \varepsilon_n(s) = x(s)z + \varepsilon_n(s) = x(s)z + \mathrm{O}(a_n) \tag{13}$$

uniformly in s, for some $a_n > 0$, just considering appropriate $C_1(n)$ and δ_n, where $z = (z_1, \ldots, z_{m_1})^{\mathrm{T}}$ and $x(s)$ is given as follows:

$$x(s) = (1 - g_d(st_1), \ldots, 1 - g_d(st_{m_1})). \tag{14}$$

Bearing the above in mind, a projection of γ will be obtained. Thus, select a finite set of distances r_i and weights w_i, $1 \leqslant i \leqslant m_2$, for some positive integer $m_2 = m_2(n)$. Then, we should find those values y_j, $1 \leqslant j \leqslant m_1$, such that they minimize

$$\sum_{i=1}^{m_2} w_i \left(\gamma(r_i) - \sum_{j=1}^{m_1} (1 - g_d(r_i t_j))y_j\right)^2 = (\vec{\gamma} - Xy)^{\mathrm{T}} W (\vec{\gamma} - Xy),$$

where g_d is defined in (12), $\vec{\gamma} = (\gamma(r_1), \ldots, \gamma(r_{m_2}))^{\mathrm{T}}$, $y = (y_1, \ldots, y_{m_1})^{\mathrm{T}}$, X is the $m_2 \times m_1$ matrix whose ith row is $x(r_i)$, as given in (14), and W is the $m_2 \times m_2$ diagonal matrix of weights w_i.

Then, the projection of γ is given by

$$\gamma^*(s) = \sum_{j=1}^{m_1} (1 - g_d(st_j))y_j^* = x(s)y^*, \quad s \geqslant 0, \tag{15}$$

where $y^* = (y_1^*, \ldots, y_{m_1}^*)^{\mathrm{T}} = B\vec{\gamma}$ and $B = (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W$.

It is simply to check that γ^* satisfies the conditionally negative definiteness property, condition (1). Moreover, the procedure followed to construct the projection of γ is the basis of Shapiro and Botha's method used to yield a valid semivariogram estimator, which we will apply next.

Along this section, consider that $\tilde{\gamma}_h$ may represent $\hat{\gamma}_h$ or $\hat{\gamma}_{q,h}$, as given in (5) and (11), respectively. Proceeding in a similar way as above, the problem of constructing a semivariogram estimator from $\tilde{\gamma}_h$, which satisfies property (1), may be reduced to find y_j, $1 \leqslant j \leqslant m_1$, such that they minimize

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

which provides an estimator of γ^* rather than of γ.

An optimal solution is given by $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_{m_1})^{\mathrm{T}} = B\vec{\tilde{\gamma}}$, where $\vec{\tilde{\gamma}} = (\tilde{\gamma}_h(r_1), \ldots, \tilde{\gamma}_h(r_{m_2}))^{\mathrm{T}}$. Then, the resulting valid semivariogram estimator has the following explicit representation:

$$\bar{\gamma}_h(s) = \sum_{j=1}^{m_1} (1 - g_d(st_j))\hat{y}_j = x(s)B\vec{\tilde{\gamma}}, \quad s \geqslant 0. \tag{16}$$

Several possibilities are suggested for the selection of weights. For instance, all the weights may equal 1 or each weight could be taken as an estimate of a quantity proportional to the inverse of the variance of the estimator considered, which would be equivalent to using the ordinary or the weighted least-squares criteria, respectively; see Cressie (1993). In this sense, a second alternative would be that of selecting w_i as an estimator of $(C_d(r_i))^{-1}$, defined in (32); the latter estimator may be obtained by assuming a Gaussian process so that g would be given as in (4) and $\gamma(s)$ might be estimated by $\sum_{j=1}^{m_1} (1 - g_d(st_j))y_j$.

Moreover, this procedure may be iterated by updating the weights in each stage. In the first iteration, all the weights may be taken to equal 1, in order to obtain an estimator $\bar{\gamma}_h^{(1)}$; then, the weights can be recalculated as described in the second alternative above and used in a second iteration, and so forth.

Next, the convergence rates for $\bar{\gamma}_h$ will be stated under the following assumption, which is a typical requirement in a regression setting:

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

Theorem 3.10. *Assume the hypotheses required in Theorem* 3.7. *In addition, suppose that condition* S11 *is satisfied. Then, it follows for* $s, s' > 0$:

$$E[\bar{\gamma}_h(s)] - \gamma(s) = \mathrm{O}(h^2 + m_2 a_n),$$

$$\mathrm{Var}[\bar{\gamma}_h(s)] = \mathrm{O}(\lambda^{-d}),$$

$$\mathrm{Cov}[\bar{\gamma}_h(s), \bar{\gamma}_h(s')] = \mathrm{O}(\lambda^{-d}).$$

The proof of this theorem will be outlined in Section 5.6.

Remark 3.11. According to Theorem 3.10, the order of the variance of $\tilde{\gamma}_h(s)$ holds for $\bar{\gamma}(s)$; whereas, a term of order $m_2 a_n$ must be added for $\bar{\gamma}(s)$ as regards the bias. However, the latter term involves parameters $m_2(n)$, $C_1(n)$ as well as δ_n and, consequently, $m_1(n)$, which may be appropriately selected to produce a negligible order, $\mathrm{o}(h^2)$.

3.3. Bandwidth selectors

Let h_{AMSE} and h_{AMISE} be the bandwidth parameters that asymptotically minimize MSE and MISE criteria, respectively, as given in (8) and (9). Both expressions involve unknown characteristics of the random process, dependent on functions γ'' as well as g, that must be estimated; thus, in this section, several bandwidth selectors will be presented. We have considered plug-in methods that basically consist of replacing γ'' and g by estimators $\breve{\gamma}''$ and $\breve{g}$, respectively, in (8) and (9), in order to obtain the bandwidth selectors $\hat{h}_{\mathrm{AMSE}}$ and $\hat{h}_{\mathrm{AMISE}}$.

It is important to note that the estimation of function g turns out to be rather complicated, unless we proceed as if the random process were Gaussian. In this case, function g is as given in (4) so that its estimation would be reduced to the quite simpler problem of estimation of γ. Following this procedure, next we will propose different procedures for constructing estimators of $\gamma(s)$ whose second derivatives will provide estimators of $\gamma''(s)$.

Denote by $\{r_k / 1 \leqslant k \leqslant m\}$ a subset of the collection of all the distances $\{\|s_i - s_j\| / 1 \leqslant i, j \leqslant n\}$. The first plug-in method suggested is based on considering a parametric family of valid isotropic semivariograms $\mathscr{F} = \{\gamma(\cdot) / \gamma(\cdot) = \gamma(\cdot, \theta),\ \theta \in \Theta\}$. Then, the parameter θ may be chosen by ordinary least squares, so as to minimize

$$\mathscr{Q}(\theta) = \sum_{k=1}^{m} (\hat{\gamma}(r_k) - \gamma(r_k, \theta))^2,$$

where $\hat{\gamma}$ represents the empirical semivariogram, as given in (2).

A different alternative may be that of applying the generalized least-squares criterium, so that we may take into account the correlations between the different values of the empirical semivariogram and select θ to minimize

$$\mathscr{Q}(\theta) = (\vec{\hat{\gamma}} - \vec{\gamma}_\theta)^{\mathrm{T}} V (\vec{\hat{\gamma}} - \vec{\gamma}_\theta),$$

where $\vec{\hat{\gamma}} = (\hat{\gamma}(r_1), \ldots, \hat{\gamma}(r_m))^{\mathrm{T}}$, $\vec{\gamma}_\theta = (\gamma(r_1,\theta), \ldots, (\gamma(r_m,\theta))^{\mathrm{T}}$ and V is the variance–covariance matrix of the $\hat{\gamma}(r_k)$. Since V is in general unknown, we may use instead an estimator of it.

A compromise between efficiency and simplicity is weighted least squares, namely minimization of

$$\mathscr{Q}(\theta) = \sum_{k=1}^{m} w_k(\hat{\gamma}(r_k) - \gamma(r_k,\theta))^2,$$

where the weights could be taken as $w_k = |N(r_k)|(\gamma(r_k,\theta))^{-2}$ and $N(r_k)$ is defined as in (3); see Cressie (1993).

Proceeding as described above, we may obtain an estimator $\hat{\theta}$ of θ. In consequence, $\gamma(s,\hat{\theta})$ and $\gamma''(s,\hat{\theta})$ could be used in (8) and (9) to obtain the bandwidth $\hat{h}_{\mathrm{AMSE}}$ and $\hat{h}_{\mathrm{AMISE}}$, which will be referred to as the parametric selectors of h in the first stage.

A second plug-in method could be that of considering kernel-type estimators of $\gamma(s)$, since they are properly defined on $[0,+\infty)$. There are several alternatives based either on using the original kernel estimator $\hat{\gamma}_{h_1}(s)$ or on correcting the latter estimator near the endpoint by using $\hat{\gamma}_{q,h_1}$, as defined in (5) or (11).

The estimators proposed before demand the specification of a pilot bandwidth h_1. An idea could be that of estimating h_1 as suggested for h in the first part of this section, by selecting a parametric family, bearing in mind that now the dominant terms of the variance–covariance matrix of the kernel estimators are proportional to $C'_d(r_i,r_j)$. Once h_1 estimated, we may use it to approximate h, obtaining a parametric selector in the second stage.

On the other hand, a kernel-type estimator may be applied for approximation of h_1, dependent on a new bandwidth h_2. In that case, we may consider again a parametric estimator to approximate h_2 and use it to estimate both h_1 and h; the latter bandwidth will be called the parametric selector in the third stage. Another option is that of applying the kernel method again, which will require the specification of a bandwidth h_3 and, depending on when the parametric estimator is used, the estimation of the bandwidth parameter h will be completed in the fourth or the following stages.

This procedure of estimation of the original bandwidth may be iterated, in the sense that it is possible to delay the stage where the parametric semivariogram is applied. Proceeding in this way, it seems reasonable that the latter conveys to a parallel reduction on the effect of a bad specification of the parametric family, which is used to approximate the theoretical semivariogram, although we have not investigated this point.

4. Numerical study

We now describe the numerical study conducted to illustrate the performance of the semivariogram estimators proposed. Four semivariogram models have been considered:

- Spherical semivariogram (Sph):

$$\gamma(s,\theta_0,\theta_1,\theta_2)=\begin{cases}0 & \text{if } s=0,\\ \theta_0+\theta_1(\frac{3}{2}\frac{s}{\theta_2}-\frac{1}{2}(\frac{s}{\theta_2})^3) & \text{if } 0\leqslant s\leqslant\theta_2,\\ \theta_0+\theta_1 & \text{if } s>\theta_2.\end{cases}$$

- Exponential semivariogram (Exp):

$$\gamma(s,\theta_0,\theta_1,\theta_2)=\begin{cases}0 & \text{if } s=0,\\ \theta_0+\theta_1(1-\exp(-\frac{s}{\theta_2})) & \text{if } s\geqslant 0.\end{cases}$$

- Rational quadratic semivariogram (Rat):

$$\gamma(s,\theta_0,\theta_1,\theta_2)=\begin{cases}0 & \text{if } s=0,\\ \theta_0+\theta_1\frac{s^2}{1+s^2\theta_2} & \text{if } s\geqslant 0.\end{cases}$$

- Wave semivariogram (Wav):

$$\gamma(s,\theta_0,\theta_1,\theta_2)=\begin{cases}0 & \text{if } s=0,\\ \theta_0+\theta_1(1-\frac{\theta_2}{s}\sin(\frac{s}{\theta_2})) & \text{if } s\geqslant 0.\end{cases}$$

In particular, the theoretical semivariograms used in the simulation study have a nugget of 0.25, a sill of 5.5 and a range (referred to the minimum value for which the semivariogram reaches either the sill or 95% of the sill, in case that the range is not finite) of 0.5.

In all cases, a uniform distribution on $[0,1]\times[0,1]$ was assumed for the spatial locations s_i, $1\leqslant i\leqslant n$, to obtain samples of size $n=50,250$. Then, the data $Z(s_i)$, $1\leqslant i\leqslant n$, were generated from a stationary Gaussian process with zero mean and an isotropic semivariogram, as specified above, and this procedure was repeated to obtain 100 independent samples from the Gaussian process.

For each data set, we considered several methods for estimation of the semivariogram. At first, the empirical estimator was computed; then, a parametric estimator was obtained by selecting a parametric semivariogram, among the four models proposed, and applying the weighted least-squares fit to the empirical estimator (2), which provided estimates of parameters θ_0, θ_1 and θ_2. On the other hand, the kernel estimator (11) was computed, by constructing a boundary kernel H_q from the Epanechnikov kernel and the quartic kernel. With regard to the bandwidth parameter, we considered the local estimator (8) and used a parametric selector obtained by a weighted least-squares fit, as described in Section 3.3. In addition, we used Shapiro and Botha's method to obtain a valid semivariogram estimator from the kernel semivariogram.

For each sample, the integrated quadratic error (or $\mathscr{L}_2$-distance) of the parametric estimator as well as that of the valid kernel semivariogram, denoted respectively by ISE_1 or ISE_2, have been approximated numerically. Table 1 shows the mean values obtained for the latter integrated quadratic errors.

Table 1

Theor. model	Value	Param. model ($n = 50$)				Param. model ($n = 250$)			
		Sph.	Exp.	Rat.	Wav.	Sph.	Exp.	Rat.	Wav.
Sph.	Mean(ISE$_1$)	2.34	1.90	1.94	1.57	1.74	1.83	1.31	2.03
Sph.	Mean(ISE$_2$)	1.90	1.47	1.56	1.40	1.60	1.35	1.13	1.90
Sph.	$\frac{\text{Mean(ISE}_1)}{\text{Mean(ISE}_2)}$	1.23	1.29	1.25	1.13	1.09	1.35	1.15	1.07
Exp.	Mean(ISE$_1$)	1.24	1.37	1.66	1.12	0.82	0.90	0.78	1.10
Exp.	Mean(ISE$_2$)	1.25	1.08	1.23	1.01	0.77	0.74	0.66	0.94
Exp.	$\frac{\text{Mean(ISE}_1)}{\text{Mean(ISE}_2)}$	0.99	1.27	1.34	1.11	1.07	1.22	1.19	1.17
Rat.	Mean(ISE$_1$)	1.19	0.93	1.00	1.78	2.14	1.29	1.13	1.66
Rat.	Mean(ISE$_2$)	1.22	1.03	0.88	1.51	1.78	1.23	1.00	1.59
Rat.	$\frac{\text{Mean(ISE}_1)}{\text{Mean(ISE}_2)}$	0.98	0.91	1.34	1.18	1.20	1.05	1.13	1.05
Wav.	Mean(ISE$_1$)	3.37	4.03	3.17	5.07	3.95	4.13	4.47	2.86
Wav.	Mean(ISE$_2$)	3.14	4.23	2.77	4.28	3.88	4.23	3.78	2.86
Wav.	$\frac{\text{Mean(ISE}_1)}{\text{Mean(ISE}_2)}$	1.07	0.95	1.14	1.18	1.02	0.98	1.19	1.00

In general, we may appreciate the better behavior of the valid kernel estimator than that of the parametric semivariogram, as it provides less or equal mean of the integrated quadratic error in 75% of the cases considered for samples of size $n=50$ and in 93.75% for $n = 250$. For the spherical semivariogram, the valid kernel estimator has always the best performance, regardless of the parametric model that is considered. The latter also happens to the exponential and the rational quadratic semivariograms, when using large samples ($n = 250$). The advantage to the kernel estimation is also clear for the wave semivariogram, although the difference with the parametric estimation tends to decrease as the number of data increases.

Moreover, in view of Table 1, the values achieved for the parametric estimator exceed in at least a 5% those obtained for the valid kernel semivariogram in 12 and 13 of the 16 studies carried out for 50 and 250 data, respectively.

It is also significant that the advantage given to the parametric semivariogram, when following the theoretical model, does not produce a better approximation as expected. In fact, observe that the kernel estimation produces the lowest values for the mean of the integrated quadratic error, with the exception of the equality achieved for the wave semivariogram and $n = 250$.

The same conclusions as regards the better performance of the valid kernel estimators remain valid when the parametric family is wrongly selected, as we can appreciate in a third of the studies for 50 data, which amounts to just one case for a larger sample size. In this respect, recall that the bad specification of the parametric family has a second-order effect on the kernel estimator, since it only affects estimation of values associated to the bandwidth parameter. A similar phenomenon is observed for plug-in rules in density estimation, where the unknown constants in the bandwidth parameter are computed under the assumption of normality, although the underlying distribution is not normal or even unimodal.

A final remark about the numerical study will be referred to the computational cost of the estimators used in it. Once the data have been collected, the time needed to compute the parametric semivariogram for one sample is of 0.02 and 0.06 s for 50 and 250 data, respectively. In case that the kernel estimation is considered, the simulation time amounts to 3.06 and 5.65 s, for the respective sizes $n = 50$ and 250, since the valid kernel semivariogram involves estimation of the bandwidth parameter. The kernel method might have worked even better if we had used instead a parametric selector of h in the second or the following stages, with the corresponding increasing of the computational time.

5. Proofs

5.1. Proof of Theorem 3.1

We will proceed as in Hall et al. (1994) to derive this proof, together with the following lemma.

Lemma 5.1. *Let $\{X_n\}$ be a sequence of uniformly bounded random variables such that $X_n = o(1)$ a.s. Then, $EX_n^r = o(1)$, for all r.*

Denote by

$$a(s) = \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{s - \|s_i - s_j\|}{h}\right),$$

$$b(s) = \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{s - \|s_i - s_j\|}{h}\right) \gamma(\|s_i - s_j\|).$$

Recall that the spatial locations has been taken as $s_i = \lambda u_i$, $1 \leqslant i \leqslant n$; see conditions S3 and S5. Then, $E[\hat{\gamma}_h(s)/U_1, \ldots, U_n] = b(s)(a(s))^{-1}$ and, consequently

$$E[\hat{\gamma}_h(s)/U_1, \ldots, U_n] - \gamma(s) = \frac{b(s) - a(s)\gamma(s)}{a(s)}. \tag{17}$$

We will check that the following orders hold for $s \geqslant Ch$:

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

$$b(s) - a(s)\gamma(s) = \frac{1}{2} c_K f_1(0) s^{d-1} A_d \gamma''(s) n^2 \lambda^{-d} h^3 + o(n^2 \lambda^{-d} h^3) \quad \text{a.s.}$$

A sketch of these proofs will be outlined below.

Then, by considering the latter relations and applying Lemma 5.1 to (17), one obtains

$$E[\hat{\gamma}_h(s)] - \gamma(s) = E\left[\frac{b(s) - a(s)\gamma(s)}{a(s)}\right] = \frac{1}{2}\, c_K \gamma''(s) h^2 + \mathrm{o}(h^2),$$

which would complete the Proof of Theorem 3.1.

5.1.1. Order of $a(s)$

It is easy to see that for $s \geqslant Ch$, we have that

$$a(s) = \sum_{i \neq j} K\left(\frac{s - \|s_i - s_j\|}{h}\right) + nK\left(\frac{s}{h}\right) = \sum_{i \neq j} K\left(\frac{s - \|s_i - s_j\|}{h}\right)$$

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

Write

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

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

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

It follows that

$$\alpha = E\left[K\left(\frac{s - \lambda\|U_1 - U_2\|}{h}\right)\right] = \int K\left(\frac{s - \lambda\|y\|}{h}\right) f_1(y)\,\mathrm{d}y. \tag{18}$$

Now, convert $y = (y^{(1)}, \ldots, y^{(d)})$ to spherical polar coordinates with the transformation

$$y^{(i)} = r\cos\theta_i \prod_{j=0}^{i-1} \sin\theta_j, \tag{19}$$

where $\sin\theta_0 = \cos\theta_d = 1$, $0 \leqslant \theta_{d-1} < 2\pi$ and $0 \leqslant \theta_i < \pi$, for $i = 1, \ldots, d-2$.

The Jacobian of the transformation is given by $r^{d-1} J_d(\theta_1, \ldots, \theta_{d-1})$, where

$$J_d(\theta_1, \ldots, \theta_{d-1}) = (\sin\theta_1)^{d-2} (\sin\theta_2)^{d-3} \ldots \sin\theta_{d-2} \tag{20}$$

as used in Schulman (2000) and proved in Fikhtengolts (1965). Then

$$\alpha = \int_0^{m_0} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} r^{d-1} J_d(\theta_1, \ldots, \theta_{d-1}) K\left(\frac{s - \lambda r}{h}\right)$$

$$\times f_1\left(r\cos\theta_1, \ldots, r\prod_{j=1}^{d-1}\sin\theta_j\right) \mathrm{d}r\, \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_{d-2}\, \mathrm{d}\theta_{d-1}$$

$$= \lambda^{-1} h \int_{(s-\lambda m_0)/h}^{s/h} \int_0^{\pi} \ldots \int_0^{\pi} \int_0^{2\pi} (\lambda^{-1}(s - ht))^{d-1} J_d(\theta_1, \ldots, \theta_{d-1}) K(t)$$

$$\times f_1\left(\lambda^{-1}(s-ht)\cos\theta_1,\ldots,\lambda^{-1}(s-ht)\prod_{j=1}^{d-1}\sin\theta_j\right)\,\mathrm{d}t\,\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{d-2}\,\mathrm{d}\theta_{d-1}$$

$$= f_1(0)s^{d-1}A_d\lambda^{-d}h + \mathrm{o}(\lambda^{-d}h)$$

on account of condition S6 and the fact that K is compactly supported, where $m_0 = \sup\{\|x_1 - x_2\|/x_1, x_2 \in D_0\}$ and

$$A_d = \int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi} J_d(\theta_1,\ldots,\theta_{d-1})\,\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{d-2}\,\mathrm{d}\theta_{d-1}. \tag{21}$$

With this notation, $a(s) - n(n-1)\alpha$ 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

$$E[Z_j/U_1,\ldots,U_{j-1}] = 0 \tag{22}$$

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

$$S_2 = Z_2, \quad S_3 = Z_2 + Z_3,\ldots, \quad S_n = Z_2 + \cdots + Z_n. \tag{23}$$

The random variables U_1 and U_2 are continuous and bounded, with common density f_0 satisfying S4; therefore, the density of $U_1 - U_2$ conditioned on $U_2 = y$, $f_{1,y}$, will be uniformly bounded in y. In consequence, there exist positive constants C_1, C_2, such that

$$\sup_y E[W_{i,j}/U_j = y] = \sup_y \int K\left(\frac{s-\lambda\|\mathrm{v}\|}{h}\right) f_{1,y}(v)\,\mathrm{d}v \leqslant C_1 s^{d-1}\lambda^{-d}h,$$

$$\sup_y E[W_{i,j}^2/U_j = y] = \sup_y \int \left(K\left(\frac{s-\lambda\|v\|}{h}\right)\right)^2 f_{1,y}(v)\,\mathrm{d}v \leqslant C_2 s^{d-1}\lambda^{-d}h.$$

On account of the latter relations, we may obtain that

$$\sup_y E[D_{i,j}^2/U_j = y] \leqslant C_3\lambda^{-d}h\max\{1, s^{2(d-1)}\} \tag{24}$$

for some constant $C_3 > 0$.

Now, by using (22), (23) and (24), we may proceed in a similar way as in proof of Theorem 3.1 of Hall et al. (1994) to conclude that

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

Hence, with probability 1:

$$a(s) = \sum_{i\neq j}(W_{i,j}-\alpha) + n(n-1)\alpha = f_1(0)s^{d-1}A_d n^2\lambda^{-d}h + \mathrm{o}(n^2\lambda^{-d}h) \quad \text{a.s.} \tag{25}$$

5.1.2. Order of $b(s)-a(s)\gamma(s)$

Bear in mind that for $s \geqslant Ch$ and large n:

$$b(s)-a(s)\gamma(s)=\sum_{i\neq j} K\left(\frac{s-\|s_i-s_j\|}{h}\right)(\gamma(\|s_i-s_j\|)-\gamma(s)).$$

Take again

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

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

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

Then, $b(s)-a(s)\gamma(s)-n(n-1)\alpha$ represents an observed value of

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

Observe that

$$\alpha=E\left[K\left(\frac{s-\lambda\|U_1-U_2\|}{h}\right)(\gamma(\lambda\|U_1-U_2\|)-\gamma(s))\right]. \tag{26}$$

Now, we may again convert $y=(y^{(1)},\dots,y^{(d)})$ to spherical polar coordinates, as given in (19), and do the change of variable $t=(s-\lambda r)/h$, to obtain that

$$\begin{aligned}\alpha&=\lambda^{-1}h\int_{(s-\lambda m_0)/h}^{s/h}\int_0^{\pi}\dots\int_0^{\pi}\int_0^{2\pi}(\lambda^{-1}(s-ht))^{d-1}J_d(\theta_1,\dots,\theta_{d-1})K(t)\\&\qquad\times(\gamma(s-ht)-\gamma(s))f_1\left(\lambda^{-1}(s-ht)\cos\theta_1,\dots,\lambda^{-1}(s-ht)\prod_{j=1}^{d-1}\sin\theta_j\right)\\&\qquad\times \mathrm{d}t\,\mathrm{d}\theta_1\dots\mathrm{d}\theta_{d-2}\,\mathrm{d}\theta_{d-1}\\&=\frac{1}{2}c_K f_1(0)s^{d-1}A_d\gamma''(s)\lambda^{-d}h^3+\mathrm{o}(\lambda^{-d}h^3)\end{aligned}$$

by using condition S2, where $m_0=\sup\{\|x_1-x_2\|/x_1,x_2\in D_0\}$.

From this point on, we would follow as in the previous section, just by considering that now $\sup_y E[D_{i,j}^2/U_j=y]\leqslant C_4\lambda^{-d}h^3\max\{1,s^{2(d-1)}\}$, for some positive constant C_4, which allows to conclude that

$$(n^2\lambda^{-d}h^3)^{-1}\sum_{i\neq j}(W_{i,j}-\alpha)\to 0\quad \text{a.s.}$$

Therefore, we obtain that

$$b(s) - a(s)\gamma(s) = \sum_{i \neq j} (W_{i,j} - \alpha) + n(n-1)\alpha$$

$$= \frac{1}{2} c_K f_1(0) s^{d-1} A_d \gamma''(s) n^2 \lambda^{-d} h^3 + o(n^2 \lambda^{-d} h^3) \quad \text{a.s.}$$

5.2. Proof of Theorem 3.2

Consider that

$$\mathrm{Var}[\hat{\gamma}_h(s)] = \mathrm{Var}[E[\hat{\gamma}_h(s)/U_1, \ldots, U_n]] + E[\mathrm{Var}[\hat{\gamma}_h(s)/U_1, \ldots, U_n]]. \tag{27}$$

By using (17) and Lemma 5.1, it is straightforward to see that for $s \geqslant Ch$:

$$\mathrm{Var}[E[\hat{\gamma}_h(s)/U_1, \ldots, U_n]] = o(h^4).$$

On the other hand, we will check that for $s \geqslant Ch$ one has

$$\mathrm{Var}[\hat{\gamma}_h(s)/U_1, \ldots, U_n] = \frac{d_K g(0,0,s,s)}{2 f_1(0) s^{d-1} A_d} n^{-2} \lambda^d h^{-1} + \frac{f_2(0,0) B_d(s)}{(f_1(0) A_d)^2} n^{-1}$$
$$+ \frac{f_3(0,0,0) C_d(s)}{4 (f_1(0) A_d)^2} \lambda^{-d} + o(n^{-2} \lambda^d h^{-1} + n^{-1} + \lambda^{-d}) \quad \text{a.s.}$$

Consequently, it would be enough to apply Lemma 5.1 to (27), which would allow to conclude Theorem 3.2 is valid.

In order to achieve the above, we will bear in mind that

$$\mathrm{Var}[\hat{\gamma}_h(s)/U_1, \ldots, U_n]$$
$$= (2a(s))^{-2} \sum_{i \neq j, k \neq l} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s - \|s_k - s_l\|}{h}\right)$$
$$\times \mathrm{Cov}[(Z(s_i) - Z(s_j))^2, (Z(s_k) - Z(s_l))^2]$$
$$= (2a(s))^{-2} \sum_{i \neq j, k \neq l} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s - \|s_k - s_l\|}{h}\right)$$
$$\times g(\|s_i - s_k\|, \|s_j - s_l\|, \|s_i - s_l\|, \|s_j - s_k\|) = \frac{2e_1(s) + 4e_2(s) + e_3(s)}{4(a(s))^2}$$

for some function g as given in conditions S8 and S9, where

$$e_1(s) = \sum_{i \neq j} \left(K\left(\frac{s - \|s_i - s_j\|}{h}\right)\right)^2 g(0, 0, \|s_i - s_j\|, \|s_i - s_j\|),$$

$$e_2(s) = \sum_{(i,j,l)\in E_2} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s - \|s_i - s_l\|}{h}\right)$$
$$\times g(0, \|s_j - s_l\|, \|s_i - s_l\|, \|s_i - s_j\|)$$
$$e_3(s) = \sum_{(i,j,k,l)\in E_3} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s - \|s_k - s_l\|}{h}\right)$$
$$\times g(\|s_i - s_k\|, \|s_j - s_l\|, \|s_i - s_l\|, \|s_j - s_k\|)$$

and the sets E_2 and E_3 are given as follows:

$$E_2 = \{(i,j,l)/i \neq j, l \text{ and } j \neq l\},$$
$$E_3 = \{(i,j,k,l)/i \neq k, j, l,\ \ j \neq k, l \text{ and } k \neq l\}. \tag{28}$$

Following (25), it is immediate to see that for $s \geqslant Ch$:

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

Then, in the following sections, we will prove that

$$e_1(s) = d_K f_1(0)s^{d-1}A_d g(0,0,s,s)n^2\lambda^{-d}h + \mathrm{o}(n^2\lambda^{-d}h) \quad \text{a.s.,}$$
$$e_2(s) = f_2(0,0)s^{2(d-1)}B_d(s)n^3\lambda^{-2d}h^2 + \mathrm{o}(n^3\lambda^{-2d}h^2) \quad \text{a.s.,}$$
$$e_3(s) = f_3(0,0,0)s^{2(d-1)}C_d(s)n^4\lambda^{-3d}h^2 + \mathrm{o}(n^4\lambda^{-3d}h^2) \quad \text{a.s.,}$$

which will allow to state the validity of Theorem 3.2.

5.2.1. *Order of $e_1(s)$*

Observe that for $s \geqslant Ch$:

$$E\left[\left(K\left(\frac{s - \lambda\|U_1 - U_2\|}{h}\right)\right)^2 g(0,0,\lambda\|U_1 - U_2\|, \lambda\|U_1 - U_2\|)\right]$$
$$= \int \left(K\left(\frac{s - \lambda\|y\|}{h}\right)\right)^2 g(0,0,\lambda\|y\|,\lambda\|y\|) f_1(y)\,\mathrm{d}y$$
$$= d_K f_1(0)s^{d-1}A_d g(0,0,s,s)\lambda^{-d}h + \mathrm{o}(\lambda^{-d}h)$$

by proceeding in a similar way as in Section 5.1.1, since conditions S6 and S8 are satisfied, where $d_K = \int_{-C}^{C}(K(t))^2\,\mathrm{d}t$.

Hence, $e_1(s) = d_K f_1(0)s^{d-1}A_d g(0,0,s,s)n^2\lambda^{-d}h + \mathrm{o}(n^2\lambda^{-d}h)$ a.s.

5.2.2. *Order of $e_2(s)$*

Note that for $s \geqslant Ch$ one has

$$\alpha = E\left[K\left(\frac{s - \lambda\|U_1 - U_2\|}{h}\right) K\left(\frac{s - \lambda\|U_1 - U_3\|}{h}\right)\right.$$
$$\left.\times g(0, \lambda\|U_2 - U_3\|, \lambda\|U_1 - U_3\|, \lambda\|U_1 - U_2\|)\right]$$

$$= \int \int K\left(\frac{s-\lambda\|y_1\|}{h}\right) K\left(\frac{s-\lambda\|y_2\|}{h}\right) g(0, \lambda\|y_1 - y_2\|, \lambda\|y_2\|, \lambda\|y_1\|)$$
$$\times f_2(y_1, y_2)\, \mathrm{d}y_1\, \mathrm{d}y_2.$$

Then, we will convert $y_k = (y^{(1,k)}, \ldots, y^{(d,k)})$ to spherical polar coordinates, for $d \geqslant 2$ and $k = 1, 2$, with the transformation

$$y^{(i,k)} = r_k \cos\theta_{i,k} \prod_{j=0}^{i-1} \sin\theta_{j,k}, \tag{29}$$

where $\sin\theta_{0,k} = \cos\theta_{d,k} = 1$, $0 \leqslant \theta_{d-1,k} < 2\pi$ and $0 \leqslant \theta_{i,k} < \pi$, for $i = 1, \ldots, d-2$.

The Jacobian of the transformation is given by $r_1^{d-1} r_2^{d-1} J_d(\theta_{1,1}, \ldots, \theta_{d-1,1}) \times J_d(\theta_{1,2}, \ldots, \theta_{d-1,2})$, where J_d is defined as in (20). Then, for $d \geqslant 2$:

$$\alpha = \int_0^{m_0} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} \int_0^{m_0} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} r_1^{d-1} r_2^{d-1}$$
$$\times J_d(\theta_{1,1}, \ldots, \theta_{d-1,1}) J_d(\theta_{1,2}, \ldots, \theta_{d-1,2})$$
$$\times g\left(0, \lambda\left\| r_2\cos\theta_{1,2} - r_1\cos\theta_{1,1}, \ldots, r_2 \prod_{j=0}^{d-1}\sin\theta_{j,2} - r_1\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|, \lambda r_2, \lambda r_1\right)$$
$$\times K\left(\frac{s-\lambda r_1}{h}\right) K\left(\frac{s-\lambda r_2}{h}\right)$$
$$\times f_2\left(r_1\cos\theta_{1,1}, \ldots, r_1\prod_{j=1}^{d-1}\sin\theta_{j,1}, r_2\cos\theta_{1,2}, \ldots, r_2\prod_{j=1}^{d-1}\sin\theta_{j,2}\right)$$
$$\times \mathrm{d}r_1\, \mathrm{d}\theta_{1,1} \ldots \mathrm{d}\theta_{d-2,1}\, \mathrm{d}\theta_{d-1,1}\, \mathrm{d}r_2\, \mathrm{d}\theta_{1,2} \ldots \mathrm{d}\theta_{d-2,2} \mathrm{d}\theta_{d-1,2}$$
$$= \lambda^{-2} h^2 (\lambda^{-1}s)^{2(d-1)}$$
$$\times \int_{(s-\lambda m_0)/h}^{s/h} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} \int_{(s-\lambda m_0)/h}^{s/h} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} J_d(\theta_{1,1}, \ldots, \theta_{d-1,1})$$
$$\times J_d(\theta_{1,2}, \ldots, \theta_{d-1,2})$$
$$\times g\left(0, s\left\|\cos\theta_{1,2} - \cos\theta_{1,1}, \ldots, \prod_{j=0}^{d-1}\sin\theta_{j,2} - \prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|, s, s\right)$$
$$\times K(t_1) K(t_2)$$
$$\times f_2\left(\lambda^{-1}s\cos\theta_{1,1}, \ldots, \lambda^{-1}s\prod_{j=1}^{d-1}\sin\theta_{j,1}, \lambda^{-1}s\cos\theta_{1,2}, \ldots, \lambda^{-1}s\prod_{j=1}^{d-1}\sin\theta_{j,2}\right)$$

$$\times \mathrm{d}t_1 \mathrm{d}\theta_{1,1}\ldots \mathrm{d}\theta_{d-2,1}\mathrm{d}\theta_{d-1,1}\,\mathrm{d}t_2\,\mathrm{d}\theta_{1,2}\ldots \mathrm{d}\theta_{d-2,2}\,\mathrm{d}\theta_{d-1,2} + \mathrm{o}(\lambda^{-2d}h^2)$$
$$= f_2(0,0)s^{2(d-1)}B_d(s)\lambda^{-2d}h^2 + \mathrm{o}(\lambda^{-2d}h^2)$$

because of conditions S7 and S9 and the fact that K is compactly supported, where $m_0 = \sup\{\|x_1 - x_2\|/x_1,\ x_2 \in D_0\}$ and

$$B_d(s) = \int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi} J_d(\theta_{1,1},\ldots,\theta_{d-1,1})J_d(\theta_{1,2},\ldots,\theta_{d-1,2})$$
$$\times g\left(0, s\left\|\cos\theta_{1,2} - \cos\theta_{1,1},\ldots,\prod_{j=0}^{d-1}\sin\theta_{j,2} - \prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|, s, s\right)$$
$$\times \mathrm{d}\theta_{1,1}\ldots \mathrm{d}\theta_{d-2,1}\,\mathrm{d}\theta_{d-1,1}\,\mathrm{d}\theta_{1,2}\ldots \mathrm{d}\theta_{d-2,2}\,\mathrm{d}\theta_{d-1,2}. \tag{30}$$

In consequence, $e_2(s) = f_2(0,0)s^{2(d-1)}B_d(s)n^3\lambda^{-2d}h^2 + \mathrm{o}(n^3\lambda^{-2d}h^2)$ a.s.

5.2.3. Order of $e_3(s)$

We will take into account that for $s \geqslant Ch$:

$$\alpha = E\left[K\left(\frac{s-\lambda\|U_1-U_2\|}{h}\right)K\left(\frac{s-\lambda\|U_3-U_4\|}{h}\right)\right.$$
$$\left.\times g(\lambda\|U_1-U_3\|,\lambda\|U_2-U_4\|,\lambda\|U_1-U_4\|,\lambda\|U_2-U_3\|)\right]$$
$$= \int\int\int K\left(\frac{s-\lambda\|y_1\|}{h}\right)K\left(\frac{s-\lambda\|y_2\|}{h}\right)$$
$$\times g(\lambda\|y_3\|,\lambda\|y_3+y_2-y_1\|,\lambda\|y_3+y_2\|,\lambda\|y_3-y_1\|)$$
$$\times f_3(y_1,y_3,y_3+y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}y_3. \tag{31}$$

We will again convert y_k, $k=1,2,3$, with the transformation given in (29). Then, by using the changes of variable $t_1=(s-\lambda r_1)/h$, $t_2=(s-\lambda r_2)/h$ and $t=\lambda r_3$, we may obtain that

$$\alpha = \lambda^{-3}h^2(\lambda^{-1}s)^{2(d-1)}\lambda^{-(d-1)}$$
$$\times\int_{(s-\lambda m_0)/h}^{s/h}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi}\int_{(s-\lambda m_0)/h}^{s/h}\int_0^{\pi}\cdots$$
$$\int_0^{\pi}\int_0^{2\pi}\int_0^{\lambda m_0}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi} K(t_1)K(t_2)$$
$$\times J_d(\theta_{1,1},\ldots,\theta_{d-1,1})J_d(\theta_{1,2},\ldots,\theta_{d-1,2})J_d(\theta_{1,3},\ldots,\theta_{d-1,3})$$

$$\times g\left(t, \left\| t\cos\theta_{1,3} + s\cos\theta_{1,2} - s\cos\theta_{1,1}, \ldots,\right.\right.$$

$$\left.t\prod_{j=0}^{d-1}\sin\theta_{j,3} + s\prod_{j=0}^{d-1}\sin\theta_{j,2} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|,$$

$$\left\| t\cos\theta_{1,3} + s\cos\theta_{1,2}, \ldots, t\prod_{j=0}^{d-1}\sin\theta_{j,3} + s\prod_{j=0}^{d-1}\sin\theta_{j,2}\right\|,$$

$$\left.\left\| t\cos\theta_{1,3} - s\cos\theta_{1,1}, \ldots, t\prod_{j=0}^{d-1}\sin\theta_{j,3} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|\right)$$

$$\times f_3\left(\lambda^{-1}s\cos\theta_{1,1}, \ldots, \lambda^{-1}s\prod_{j=0}^{d-1}\sin\theta_{j,1}, 0, \ldots, 0,\right.$$

$$\left.\lambda^{-1}s\cos\theta_{1,2}, \ldots, \lambda^{-1}s\prod_{j=0}^{d-1}\sin\theta_{j,2}\right)$$

$$\times \mathrm{d}t_1\,\mathrm{d}\theta_{1,1}\ldots\mathrm{d}\theta_{d-2,1}\,\mathrm{d}\theta_{d-1,1}\,\mathrm{d}t_2\,\mathrm{d}\theta_{1,2}\ldots\mathrm{d}\theta_{d-2,2}\,\mathrm{d}\theta_{d-1,2}\,\mathrm{d}t\,\mathrm{d}\theta_{1,3}\ldots$$

$$\times \mathrm{d}\theta_{d-2,3}\,\mathrm{d}\theta_{d-1,3} + \mathrm{o}(\lambda^{-3d}h^2)$$

$$= f_3(0,0,0)s^{2(d-1)}C_d(s)\lambda^{-3d}h^2 + \mathrm{o}(\lambda^{-3d}h^2),$$

where $m_0 = \sup\{\|x_1 - x_2\|/x_1, x_2 \in D_0\}$, on account of conditions S7 and S9, together with the fact that K is compactly supported, where J_d is defined as in (20) and

$$C_d(s) = \int_0^{+\infty}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi} J_d(\theta_{1,1},\ldots,\theta_{d-1,1})$$

$$\times J_d(\theta_{1,2},\ldots,\theta_{d-1,2})J_d(\theta_{1,3},\ldots,\theta_{d-1,3})$$

$$\times g\left(t, \left\| t\cos\theta_{1,3} + s\cos\theta_{1,2} - s\cos\theta_{1,1}, \ldots,\right.\right.$$

$$\left.t\prod_{j=0}^{d-1}\sin\theta_{j,3} + s\prod_{j=0}^{d-1}\sin\theta_{j,2} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|,$$

$$\left\| t\cos\theta_{1,3} + s\cos\theta_{1,2}, \ldots, t\prod_{j=0}^{d-1}\sin\theta_{j,3} + s\prod_{j=0}^{d-1}\sin\theta_{j,2}\right\|,$$

$$\left.\left\| t\cos\theta_{1,3} - s\cos\theta_{1,1}, \ldots, t\prod_{j=0}^{d-1}\sin\theta_{j,3} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|\right)$$

$$\times \mathrm{d}t\, \mathrm{d}\theta_{1,1} \ldots \mathrm{d}\theta_{d-2,1}\, \mathrm{d}\theta_{d-1,1}\, \mathrm{d}\theta_{1,2} \ldots \mathrm{d}\theta_{d-2,2}\, \mathrm{d}\theta_{d-1,2}$$

$$\times \mathrm{d}\theta_{1,3} \ldots \mathrm{d}\theta_{d-2,3}\, \mathrm{d}\theta_{d-1,3}. \tag{32}$$

Consequently, $e_3(s) = f_3(0,0,0)s^{2(d-1)}C_d(s)n^4\lambda^{-3d}h^2 + \mathrm{o}(n^4\lambda^{-3d}h^2)$ a.s.

5.3. Proof of Theorem 3.4

Take into account that

$$\mathrm{Cov}[\hat{\gamma}_h(s), \hat{\gamma}_h(s')] = \mathrm{Cov}[E[\hat{\gamma}_h(s)/U_1,\ldots,U_n], E[\hat{\gamma}_h(s')/U_1,\ldots,U_n]]$$
$$+ E[\mathrm{Cov}[(\hat{\gamma}_h(s), \hat{\gamma}_h(s'))/U_1,\ldots,U_n]].$$

Again, use relation (17) and Lemma 5.1 to conclude that, for $s, s' \geqslant Ch$:

$$\mathrm{Cov}[E[\hat{\gamma}_h(s)/U_1,\ldots,U_n], E[\hat{\gamma}_h(s')/U_1,\ldots,U_n]] = \mathrm{o}(h^4).$$

We will check that for $s, s' \geqslant Ch$, $s \neq s'$:

$$\mathrm{Cov}[(\hat{\gamma}_h(s), \hat{\gamma}_h(s'))/U_1,\ldots,U_n] = \frac{f_3(0,0,0)C'_d(s,s')}{4(f_1(0)A_d)^2}\lambda^{-d} + \mathrm{o}(\lambda^{-d}) \quad \text{a.s.} \tag{33}$$

so that the proof of Theorem 3.4 would be concluded just by applying Lemma 5.1 to (33).

To see the above, consider that

$$\mathrm{Cov}[(\hat{\gamma}_h(s), \hat{\gamma}_h(s'))/U_1,\ldots,U_n]$$
$$= \frac{1}{4a(s)a(s')}\sum_{i\neq j, k\neq l} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s' - \|s_k - s_l\|}{h}\right)$$
$$\times \mathrm{Cov}[(Z(s_i) - Z(s_j))^2, (Z(s_k) - Z(s_l))^2]$$
$$= \frac{1}{4a(s)a(s')}\sum_{i\neq j, k\neq l} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s' - \|s_k - s_l\|}{h}\right)$$
$$\times g(\|s_i - s_k\|, \|s_j - s_l\|, \|s_i - s_l\|, \|s_j - s_k\|)$$
$$= \frac{2p_1(s,s') + 4p_2(s,s') + p_3(s,s')}{4a(s)a(s')}$$

for some function g as given in conditions S8 and S9, where

$$p_1(s,s') = \sum_{i\neq j} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s' - \|s_i - s_j\|}{h}\right)$$
$$\times g(0, 0, \|s_i - s_j\|, \|s_i - s_j\|),$$

$$p_2(s,s') = \sum_{(i,j,l)\in E_2} K\left(\frac{s - \|s_i - s_j\|}{h}\right) K\left(\frac{s' - \|s_i - s_l\|}{h}\right)$$
$$\times g(0, \|s_j - s_l\|, \|s_i - s_l\|, \|s_i - s_j\|)$$

$$p_3(s,s') = \sum_{(i,j,k,l)\in E_3} K\left(\frac{s-\|s_i-s_j\|}{h}\right) K\left(\frac{s'-\|s_k-s_l\|}{h}\right)$$
$$\times g(\|s_i-s_k\|,\|s_j-s_l\|,\|s_i-s_l\|,\|s_j-s_k\|)$$

and the sets E_2 and E_3 are as given in (28).

We might check that

$$p_2(s,s') = f_2(0,0)(ss')^{d-1}B'_d(s,s')n^3\lambda^{-2d}h^2 + \mathrm{o}(n^3\lambda^{-2d}h^2),$$
$$p_3(s,s') = f_3(0,0,0)(ss')^{d-1}C'_d(s,s')n^4\lambda^{-3d}h^2 + \mathrm{o}(n^4\lambda^{-3d}h^2),$$

where

$$B'_d(s,s') = \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} J_d(\theta_{1,1},\ldots,\theta_{d-1,1})J_d(\theta_{1,2},\ldots,\theta_{d-1,2})$$
$$\times g\left(0, \left\| s'\cos\theta_{1,2} - s\cos\theta_{1,1},\ldots, s'\prod_{j=0}^{d-1}\sin\theta_{j,2} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|, s', s\right)$$
$$\times \mathrm{d}\theta_{1,1}\ldots\mathrm{d}\theta_{d-2,1}\,\mathrm{d}\theta_{d-1,1}\,\mathrm{d}\theta_{1,2}\ldots\mathrm{d}\theta_{d-2,2}\,\mathrm{d}\theta_{d-1,2},$$

$$C'_d(s,s') = \int_0^{+\infty} \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} J_d(\theta_{1,1},\ldots,\theta_{d-1,1})$$
$$\times J_d(\theta_{1,2},\ldots,\theta_{d-1,2})J_d(\theta_{1,3},\ldots,\theta_{d-1,3})$$
$$\times g\left(t, \left\| t\cos\theta_{1,3} + s'\cos\theta_{1,2} - s\cos\theta_{1,1},\ldots,\right.\right.$$
$$\left. t\prod_{j=0}^{d-1}\sin\theta_{j,3} + s'\prod_{j=0}^{d-1}\sin\theta_{j,2} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|,$$
$$\left\| t\cos\theta_{1,3} + s'\cos\theta_{1,2},\ldots, t\prod_{j=0}^{d-1}\sin\theta_{j,3} + s'\prod_{j=0}^{d-1}\sin\theta_{j,2}\right\|,$$
$$\left.\left\| t\cos\theta_{1,3} - s\cos\theta_{1,1},\ldots, t\prod_{j=0}^{d-1}\sin\theta_{j,3} - s\prod_{j=0}^{d-1}\sin\theta_{j,1}\right\|\right)$$
$$\times \mathrm{d}t\,\mathrm{d}\theta_{1,1}\ldots\mathrm{d}\theta_{d-2,1}\,\mathrm{d}\theta_{d-1,1}\,\mathrm{d}\theta_{1,2}\ldots\mathrm{d}\theta_{d-2,2}\,\mathrm{d}\theta_{d-1,2}\,\mathrm{d}\theta_{1,3}\ldots$$
$$\times \mathrm{d}\theta_{d-2,3}\mathrm{d}\theta_{d-1,3}. \tag{34}$$

In fact, the proofs that the latter orders hold for $p_2(s,s')$ and $p_3(s,s')$ would be similar to those of $e_2(s)$ and $e_3(s)$, respectively; see Sections 5.2.2 and 5.2.3. As regards

$p_1(s,s')$, consider that

$$E\left[K\left(\frac{s-\lambda\|U_1-U_2\|}{h}\right)K\left(\frac{s'-\lambda\|U_1-U_2\|}{h}\right)\right.$$
$$\left.\times g(0,0,\lambda\|U_1-U_2\|,\lambda\|U_1-U_2\|)\right]$$
$$=\int K\left(\frac{s-\lambda\|y\|}{h}\right)K\left(\frac{s'-\lambda\|y\|}{h}\right)g(0,0,\lambda\|y\|,\lambda\|y\|)f_1(y)\,\mathrm{d}y=0$$

for all large n, since $s\neq s'$ and the kernel function is compactly supported.

Thus, we can conclude the proof just bearing in mind that for $s,s'\geqslant Ch$:

$$(a(s)a(s'))^{-1}=(f_1(0)A_d)^{-2}(ss')^{-(d-1)}n^{-4}\lambda^{2d}h^{-2}+\mathrm{o}(n^{-4}\lambda^{2d}h^{-2})\quad \text{a.s.}$$

on account of relation (25).

5.4. *Proof of Theorem 3.6*

We would use the same arguments as in the proof of Theorem 3.1, just by taking into account that relations (18) and (26) are now given by

$$E\left[K\left(\frac{s-\lambda\|U_1-U_2\|}{h}\right)\right]=c_{0,K}f_1(0)s^{d-1}A_d\lambda^{-d}h+\mathrm{o}(\lambda^{-d}h),$$

$$E\left[K\left(\frac{s-\lambda\|U_1-U_2\|}{h}\right)(\gamma(\lambda\|U_1-U_2\|))-\gamma(s)\right]$$
$$=-c_{1,K}f_1(0)s^{d-1}A_d\gamma'(s)\lambda^{-d}h^2$$
$$+\frac{1}{2}c_{2,K}f_1(0)s^{d-1}A_d\gamma''(s)\lambda^{-d}h^3+\mathrm{o}(\lambda^{-d}h^3).$$

5.5. *Proof of Theorem 3.7*

The first part of this proof would be similar to that of Theorem 3.6, bearing in mind that now $c_{0,H_q}=1$, $c_{1,H_q}=0$ and that

$$E\left[H_q\left(\frac{s-\lambda\|U_1-U_2\|}{h}\right)(\gamma(\lambda\|U_1-U_2\|)-\gamma(s))\right]$$
$$=\frac{c_{2,K}c_{1,L}-c_{1,K}c_{2,L}}{2(c_{0,K}c_{1,L}-c_{1,K}c_{0,L})}f_1(0)s^{d-1}A_d\gamma''(s)\lambda^{-d}h+\mathrm{o}(\lambda^{-d}h)$$

since $c_{0,K}c_{1,L}\neq c_{1,K}c_{0,L}$, as stated in condition S10.

As regards the variance and the covariance, we would proceed as in the proofs of Theorems 3.2 and 3.4.

5.6. Proof of Theorem 3.10

It follows from condition S11 as well as relations (13) and (15) that

$$\gamma^*(s) = x(s)B\vec{\gamma} = x(s)B(Xz + \vec{\varepsilon_n}) = x(s)z + \mathrm{O}(m_2 a_n) = \gamma(s) + \mathrm{O}(m_2 a_n) \tag{35}$$

uniformly in s, where $\vec{\varepsilon_n} = (\varepsilon_n(r_1), \ldots, \varepsilon_n(r_{m_2}))^{\mathrm{T}}$.

From relation (35) and condition S11, together with Theorems 3.1 and 3.7, one has

$$E[\bar{\gamma}_h(s)] = x(s)BE[\vec{\bar{\gamma}}] = x(s)B\vec{\gamma} + \mathrm{O}(h^2) = \gamma(s) + \mathrm{O}(h^2 + m_2 a_n).$$

On the other hand, by S11 and Theorems 3.2, 3.4 and 3.7, we obtain that

$$\mathrm{Var}[\bar{\gamma}_h(s)] = x(s)B\,\mathrm{Var}[\vec{\bar{\gamma}}\,\vec{\bar{\gamma}}^{\mathrm{T}}]B^{\mathrm{T}}x(s)^{\mathrm{T}} = \mathrm{O}(n^{-2}\lambda^d h^{-1} + n^{-1} + \lambda^{-d}) + \mathrm{o}(h^4),$$

$$\mathrm{Cov}[\bar{\gamma}_h(s), \bar{\gamma}_h(s')] = x(s)B\,\mathrm{Var}[\vec{\bar{\gamma}}\,\vec{\bar{\gamma}}^{\mathrm{T}}]B^{\mathrm{T}}x(s')^{\mathrm{T}} = \mathrm{O}(n^{-2}\lambda^d h^{-1} + n^{-1} + \lambda^{-d}) + \mathrm{o}(h^4).$$

References

Christakos, G., 1992. Random Field Models in Earth Sciences. Academic Press, San Diego.
Cressie, N., 1985. Fitting variogram models by weighted least squares. Math. Geol. 17, 563–586.
Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.
Cressie, N., Hawkins, D., 1980. Robust estimation of the variogram. Math. Geol. 12, 115–125.
Diblasi, A., Bowman, A.W., 2001. On the use of the variogram in checking for independence in spatial data. Biometrics 57, 211–218.
Febrero-Bande, M., García-Soidán, P.H., González-Manteiga, W., 1998. El método núcleo en la estimación del variograma. Proceedings of the XXIV Congreso Nacional de Estadística e Investigación Operativa. Departamento de Estadística y Matemática Aplicada, Univ. de Almería, Spain.
Fikhtengolts, G.M., 1965. Kurs Differentsialnogo i Integralnogo Ischisleniya, Vol. 3. Nauka, Moscow, pp. 401–402.
Genton, M., 1998a. Highly robust variogram estimation. Math. Geol. 30, 213–221.
Genton, M., 1998b. Variogram fitting by generalized least squares using an explicit formula for the covariance structure. Math. Geol. 30, 323–345.
Genton, M., Gorsich, D.J., 2002. Nonparametric variogram and covariogram estimation with Fourier–Bessel matrices. Comput. Statist. Data Anal. 41, 47–57.
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.
Jones, M.C., 1993. Simple boundary correction for kernel density estimation. Statist. Comput. 3, 135–146.
Kyung-Joon, C., Shucany, W., 1998. Nonparametric kernel regression estimation near endpoints. J. Statist. Plann. Inference 66, 289–304.
Matheron, G., 1963. Principles of geostatistics. Econom. Geol. 58, 1246–1266.
Müller, H., 1993. On the boundary kernel method for non-parametric curve estimation near endpoints. Scand. J. Statist. 20, 313–328.
Schulman, A.E., 2000. Moments of spherically symmetric kernels for multivariate local regression and density estimation. Nonparametric Statist. 12, 373–376.
Shapiro, A., Botha, J., 1991. Variogram fitting with a general class of conditionally nonnegative definite functions. Comput. Statist. Data Anal. 11, 87–96.
Stein, M.L., 1999. Interpolation of Spatial Data, Some Theory for Kriging. Springer, New York.