

A Kernel Variogram Estimator for Clustered Data

RAQUEL MENEZES

Department of Mathematics for Science and Technology, University of Minho

PILAR GARCIA-SOIDÁN

Department of Statistics and Operations Research, University of Vigo

MANUEL FEBRERO-BANDE

Department of Statistics and Operations Research, University of Santiago de Compostela

ABSTRACT. The variogram provides an important method for measuring the dependence of attribute values between spatial locations. Suppose that the nature of the sampling process leads to the presence of clustered data; it would be advisable to use a variogram estimator that aims to adjust for clustering of samples. In this setting, the use of a non-parametric weighted estimator, obtained by considering an inverse weight to a given neighbourhood density combined with the kernel method, seems to have a satisfactory behaviour in practice. This paper pursues a theoretical study of the cluster robust estimator, by proving that it is asymptotically unbiased as well as consistent and by providing criteria for selection of the bandwidth parameter and the neighbourhood radius. Numerical studies are also included to illustrate the performance of the considered estimator and the suggested approaches.

Key words: cluster, isotropy, kernel method, variogram

1. Introduction

Among existing geostatistical methods, variogram analysis provides a useful tool for summarizing spatial data and measuring the dependence of attribute values between spatial locations. Under certain regularity conditions, the variogram is used with kriging to predict the value of the spatial variable at an unsampled point; see, for instance, Stein (1999). The approach differs from classical regression in that local features can affect the solution. The measurements in the vicinity of the investigated point are expected to be more closely related than others to the unknown true value; so the relative contribution of each individual data value depends on its relative location. Hence, the estimation of the variogram plays a decisive role in kriging theory, as the optimal values of these relative contributions require the variogram function, which is, in general, unknown.

The importance of an accurate estimation of the variogram is not only restricted to geostatistics, but has implications for a broad spectrum of areas, such as hydrology, atmospheric science, possibly epidemiology, etc; whenever one wishes to make predictions for quantities that are continuously defined over some convex domain; see, for instance, Cressie (1993) and references therein.

Typically, one assumes that the sampling points are uniformly spread over the observation region. However, the sampling strategy may originate non-uniform sampling density, leading to the presence of clustered data. Observations that are clustered in space are in most cases driven by external factors, such as the existence of specific geographic or demographic spots; or they may be needed to better characterize short-range variability, which requires a

denser sampling, but sometimes too costly to cover the whole study region. In this context, the behaviour of the traditional variogram estimators may significantly decay, as shown in the foregoing works such as Kovitz & Christakos (2004) or Menezes (2005), and as exemplified by the numerical study which we shall present.

Our paper describes a modification of the kernel estimator aiming to adjust for clustering of samples and minimize the negative impact on variogram estimation, which was first approached in Menezes & Tawn (2003) (or see Menezes, 2005). In here, we proceed with the theoretical study of the suggested variogram estimator, proving its consistency and unbiasedness. Additionally, we propose values for its unknown *smoothing parameters*, two user-adjustable quantities that affect the performance of the variogram estimator.

The rest of this paper is structured as follows. In section 2 we place the work in the context of the existing literature, pointing out the similarities and differences with the most common variogram estimators. Section 3 introduces additional notation and summarizes the main assumptions considered in our asymptotic study. In section 4 the fundamental properties of the proposed nonparametric variogram estimator are established. The results derived for bias and variance are then used for the optimal bandwidth selection in section 4.1; this optimal value requires itself the value for the neighbourhood radius, the second smoothing parameter. Comments on the neighbourhood radius selector are included in section 4.2. The numerical studies are described in section 5 and the proofs are developed in section 6.

2. Variogram estimators

For the sake of simplicity, we have considered isotropic processes, where the variogram becomes a function of the distance, instead of the difference, of the spatial locations. This restriction can be relaxed by assuming geometric anisotropy or by fitting a distinct variogram in each of several directions. Let $\{Z(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^d\}$ be a spatial random process, where D is a bounded region with positive d -dimensional volume. Suppose that n data, $Z(\mathbf{x}_1), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_n)$, are collected at known spatial locations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. The random process is intrinsically stationary if the following conditions are satisfied:

- (i) $E[Z(\mathbf{x}_i) - Z(\mathbf{x}_j)] = 0, \quad \forall \mathbf{x}_i, \mathbf{x}_j \in D;$
- (ii) $\text{Var}[Z(\mathbf{x}_i) - Z(\mathbf{x}_j)] = 2\gamma(\|\mathbf{x}_i - \mathbf{x}_j\|), \quad \forall \mathbf{x}_i, \mathbf{x}_j \in D.$

The function $2\gamma(\cdot)$ defines the variogram function and $\gamma(\cdot)$ is termed the semivariogram. So the variogram of an intrinsic and isotropic process reduces to

$$2\gamma(\|\mathbf{x}_i - \mathbf{x}_j\|) = E[(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2].$$

Consequently, we can estimate the variogram from sample data replacing the previous theoretical expectation by the corresponding sample average. The variogram estimator most commonly adopted was proposed in Matheron (1963) and it can be represented by the weighted average

$$2\hat{\gamma}(u) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(u)[Z(\mathbf{x}_i) - Z(\mathbf{x}_j)]^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(u)}, \tag{1}$$

where $w_{ij}(u) = I_{\{\|\mathbf{x}_i - \mathbf{x}_j\| = u\}}$. In practice, this estimator is usually smoothed by taking a tolerance region T around u ; it can then be represented by $2\hat{\gamma}(u(l))$ with $w_{ij}(u(l)) = I_{\{\|\mathbf{x}_i - \mathbf{x}_j\| \in T(u(l))\}}$ and $l = 1, \dots, K$. Journel & Huijbregts (1978) recommends that the number of distinct pairs $(\mathbf{x}_i, \mathbf{x}_j)$ in each $T(u(l))$ be at least 30; the regions $\{T(u(l)) : l = 1, \dots, K\}$ are chosen to be disjoint and to accomplish the previous recommendation. Hence, the Matheron estimator can be obtained from simple averages within distance bins. This approach is convenient, but the

estimator can continue to be highly variable. Namely, it can be affected by atypical values because of the squared term, and thus alternative estimators have been proposed aiming to be more robust to outliers; for example, by working instead with square root absolute differences as described in Cressie (1993).

Suppose now that we take the weights in expression (1) as

$$w_{ij}(u) = K \left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right),$$

where K is a symmetric, zero-mean and bounded density function, with compact support $[-C, C]$. The positive number h is usually called bandwidth. The resulting variogram estimator is commonly referred to as the kernel estimator. The Matheron estimator would correspond to the election of K as the uniform kernel, with a tolerance region being given by $(u - hC, u + hC)$. The choice of the bandwidth h would then be similar to the choice of the bin size that defines the tolerance of each size-class $T(u/l)$.

The alternative election of a unimodal kernel function K , such as the Gaussian or the Epanechnikov kernels, allows for more sophisticated smoothing methods. The binning process is not needed and an estimate of the variogram function can be obtained in any distance. In this case, the choice of bandwidth, h , becomes a crucial task. Moreover, note that the difference between choosing a uniform or a non-uniform kernel means to adopt a simple or a weighted average for each estimate, respectively. The weights are at their maximum when the distance between two points is close to u , and zero values if $|(u - \|\mathbf{x}_i - \mathbf{x}_j\|)/h| > C \iff \|\mathbf{x}_i - \mathbf{x}_j\| \notin [u - hC, u + hC]$. Consequently, it offers a smoother estimation of the variogram.

The binned variogram from Matheron together with the kernel smoothed variogram are included in most software available to practicing statisticians/geostatisticians. An example is the geoR library from R, which offers a kernel estimator for exploratory purposes with the bandwidth being chosen by the user (Ribeiro & Diggle, 2001). Garcia-Soidán *et al.* (2004) suggests a transformed version of the kernel estimator, not restricted to exploratory aims but allowed to be used in kriging. They adapt the Nadaraya–Watson (NW) regression estimation to the context of spatial data and study several properties of the resulting estimator; an asymptotically optimal bandwidth selector is then obtained. The performance of this kernel estimator and the one from Matheron is compared in Menezes *et al.* (2005), where a numerical study is carried out under different kinds of spatial dependence situations, suggesting the usual superiority of the former estimator.

In any case, both foregoing estimators are badly affected by clustered data, as illustrated in Fig. 1, where the notation ‘NW kernel’ identifies the Nadaraya–Watson kernel estimator. Two distinct sampling designs were looked at: in case A, the complete spatial randomness (CSR) design; and in case B, a design presenting a strong cluster. Gaussian data were generated on those locations, considering the same theoretical semivariogram model for the dependence structure. This theoretical model was plotted against the estimates obtained from Matheron and NW kernel estimators. The worst estimates of case B are, at least partially, justified by the sample locations not being sufficiently spatially representative of the overall data, as they are irregularly distributed over the observation region. The Matheron estimator seems to be more vulnerable than the NW kernel estimator to unequal samples density, although the latter is also affected. These results were confirmed in the numerical study presented in section 5.

A compensation for the unpopulated areas is then proposed, by suggesting an inverse weight to a given neighbourhood density and, simultaneously, joining the benefits of the NW kernel estimator even though not directly related to the clustering issue. In (1), it is suggested that

$$w_{ij}(u) = \frac{1}{\sqrt{n_i \times n_j}} \times K \left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right), \quad n_i = \sum_k I_{\{\|\mathbf{x}_i - \mathbf{x}_k\| \leq \delta\}}, \quad (2)$$

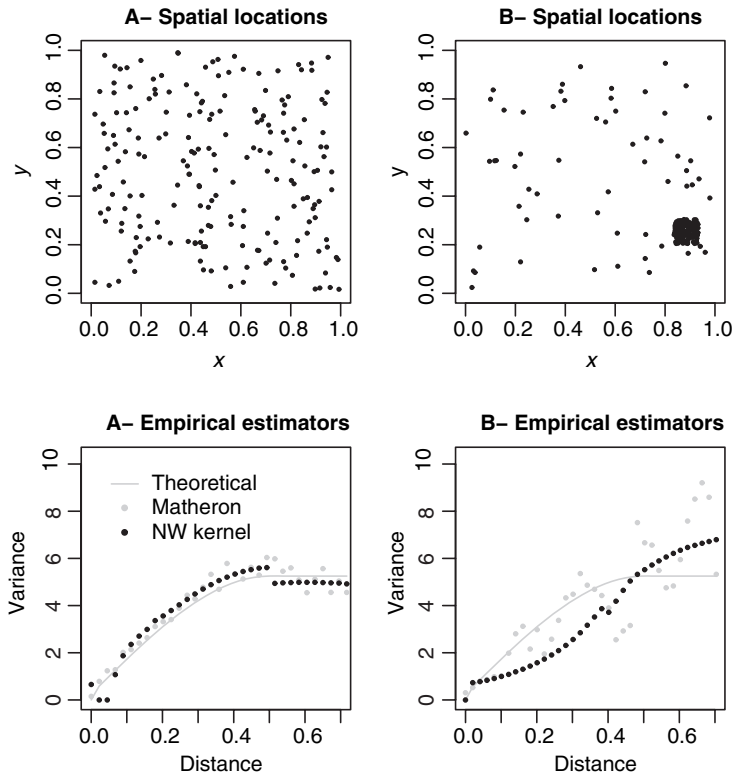


Fig. 1. Behaviour of $\hat{\gamma}$ under non-clustered (CSR) and clustered sampling.

where n_i represents the number of points that fall within the circle of radius δ and centre \mathbf{x}_i (Fig. 2). The idea is to adjust the contribution from any sample point \mathbf{x}_i according to its number of neighbours n_i ; as the variogram estimator sums contributions from pairs of data points $(\mathbf{x}_i, \mathbf{x}_j)$, the product $n_i n_j$ represents the joint contribution of the pair; the square root is adopted as a scale corrector. The resulting variogram downweights squared increments of $Z(\cdot)$ based on the geometric mean of the local densities of observations around each of the observation locations making up the increment.

Declustering methods are quite intuitive, and their need is well recognized in the spatial statistics literature to estimate spatially representative mean trends for clustered data (see e.g. Goovaerts, 1997; Isaaks & Srivastava, 1989). In contrast, the corresponding need for the reliable estimation of the second-order spatial structures is not normally considered. The presence of clustered sample data is, however, not negligible at all as illustrated in Fig. 1 and analysed by Kovitz & Christakos (2004). These authors suggest a modified form of Matheron estimator that also incorporates some declustering weights, but based on zones of proximity. Each zone of any data point is defined by the area of the Voronoi polygon that contains all points closer to that interior data point than to any other data point. The performance of this modified estimator of the variogram is analysed in terms of a numerical application.

In our case, we shall prove that the variogram estimator proposed for clustered data enjoys good properties, such as asymptotic unbiasedness and consistency. This empirical variogram requires the selection of two user-adjustable values: the kernel bandwidth h and the neighbourhood radius δ . The first will be treated via the MSE, i.e. the mean square error. The latter will result from the analysis of the density estimation of the distance between points

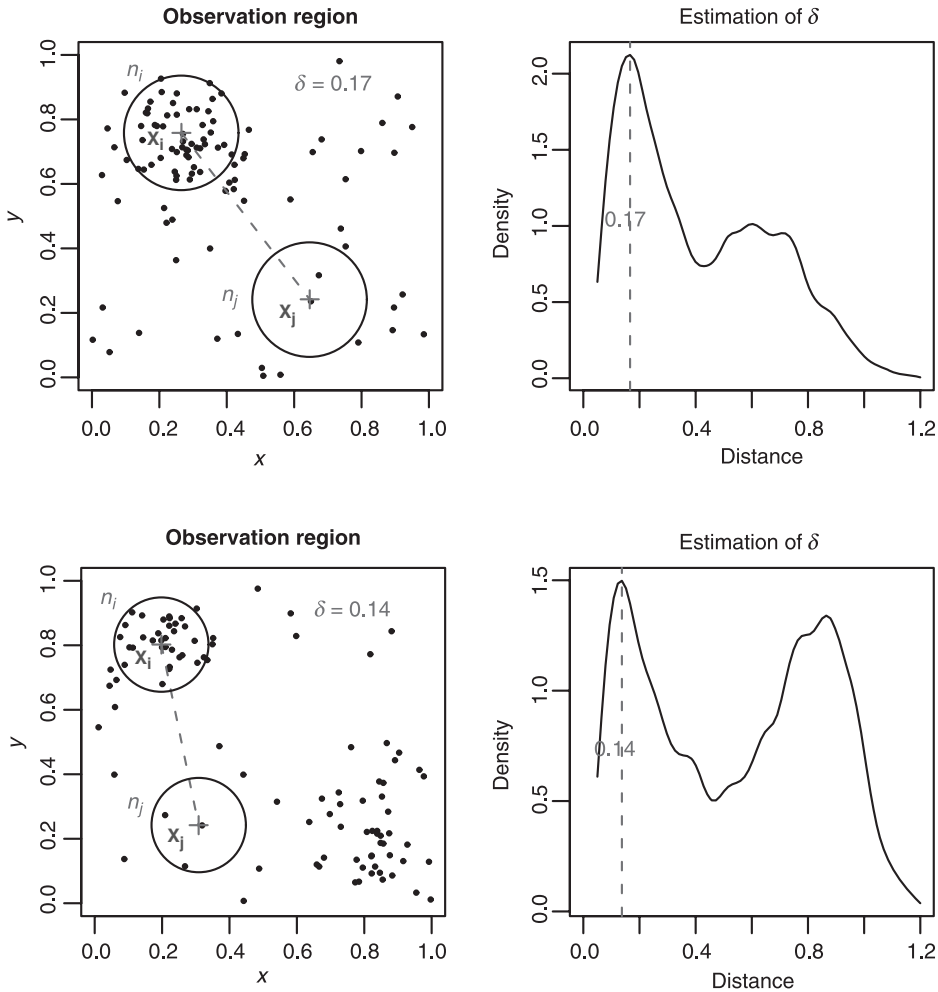


Fig. 2. Two examples of sampling designs are represented in the left panels. The corresponding density estimations of the distances are represented in the right panels. The n_i gives the number of points that fall within the circle of radius δ and centre x_i .

derived on the observation region. The declustering factor can be applied to any degree to which a data set may be clustered, inclusively to a complete spatial randomness design; the drawback would be the unnecessary additional computational effort, even though this is not expected to be much significant. One possibility here, is to adopt some metrics to assess the degree of clusteredness such as those described in Kovitz & Christakos (2004), which help to decide if common estimators are not recommended. An alternative simple metric can be based on the observation of the standard deviation of $\sqrt{n_i n_j}$ for all existent pairs (x_i, x_j) , as a low value is expected in the case of CSR design.

3. Assumptions

Let $\{Z(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^d\}$ be an intrinsic and isotropic random process. The observation region D will be considered to be increasing and a random design will be assumed for the spatial locations, as described next. This strategy, proposed in Hall *et al.* (1994) and adopted in

Garcia-Soidán *et al.* (2004) is a compromise between the infill asymptotics and the increasing domain asymptotics regimes, which allows to achieve consistent estimation of the variogram function.

- (A1) $D = D_n = \lambda D_0$ where $\lambda = \lambda_n$ diverges to $+\infty$ and $D_0 \subset \mathbb{R}^d$ is a bounded and fixed region.
- (A2) $\mathbf{x}_i = \lambda \mathbf{v}_i$, $i = 1 \dots n$, where \mathbf{v}_i is a realization of a random sample \mathbf{V}_i from f_0 , the density function defined on D_0 .
- (A3) For all $\mathbf{v} \in D_0$ and for some positive constants d_1 and d_2 , one has $d_1 \leq f_0(\mathbf{v}) \leq d_2$.

In (A3), we assume that the density f_0 does not exhibit holes of odds probabilities (very low or high) in D_0 , so that f_0 is regular enough. Typically, we expect the uniformity of f_0 in D_0 , in a sense of all points in D_0 have the same probability of being sampled. The clustering of data is determined by external factors to the spatial variable of interest Z . In some cases, for example of the study of air pollution, clustering can also be justified by a prior scientific knowledge of the variable of interest, such as the expected local level of air contamination. This situation results into the dependence between the sample points \mathbf{v} and the random field Z , claiming f_0 to be a function of Z . The estimation of f_0 under this dependency is beyond the scope of this work.

Several hypotheses are now required as regards the second- and fourth-order moments of the random process.

- (A4) γ admits three continuous derivatives in a neighbourhood of u , for all $u > 0$.
- (A5) There is a bounded and continuously differentiable function $g: \mathbb{R}^{3d} \rightarrow \mathbb{R}$ satisfying that $\text{Cov} [(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2, (Z(\mathbf{x}_k) - Z(\mathbf{x}_l))^2] = g(\mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_k, \mathbf{x}_i - \mathbf{x}_l) = g(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$. We also assume that for any $\delta > 0$, one has:

$$\sup_{\|\mathbf{y}_2\| \geq r \wedge \|\mathbf{y}_3\| \geq r} \{ |g(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)| : \|\mathbf{y}_1\| \leq \delta, \|\mathbf{y}_2 - \mathbf{y}_3\| \leq \delta \} \xrightarrow{r \rightarrow \infty} 0.$$

For instance, in the context of a Gaussian process, it follows that:

$$\begin{aligned} & \text{Cov} [(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2, (Z(\mathbf{x}_k) - Z(\mathbf{x}_l))^2] \\ &= 2 [\gamma(\|\mathbf{x}_i - \mathbf{x}_k\|) + \gamma(\|\mathbf{x}_j - \mathbf{x}_l\|) - \gamma(\|\mathbf{x}_i - \mathbf{x}_l\|) - \gamma(\|\mathbf{x}_j - \mathbf{x}_k\|)]^2 \end{aligned}$$

and, afterwards, one may take

$$g(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) = 2 [\gamma(\|\mathbf{y}_2\|) + \gamma(\|\mathbf{y}_3 - \mathbf{y}_1\|) - \gamma(\|\mathbf{y}_3\|) - \gamma(\|\mathbf{y}_2 - \mathbf{y}_1\|)]^2$$

so that condition (A5) is satisfied provided that the semivariogram is bounded and has a finite sill.

The idea behind assumption (A5) is to impose a condition on the covariance between $Z(\mathbf{x}_i) - Z(\mathbf{x}_j)$ and $Z(\mathbf{x}_k) - Z(\mathbf{x}_l)$ when locations $(\mathbf{x}_i, \mathbf{x}_j)$ are distant from $(\mathbf{x}_k, \mathbf{x}_l)$.

The final conditions will be referred to the convergence rates required of the bandwidth h and the neighbourhood radius δ .

- (A6) $\{h + \lambda^{-1} + \lambda^d n^{-1} + (nh)^{-1}\} \xrightarrow{n \rightarrow \infty} 0$,
- (A7) $\delta = \lambda a$, for some $a > 0$.

4. Main results

Let $\{Z(\mathbf{x}): \mathbf{x} \in D \subset \mathbb{R}^d\}$ be an intrinsic and isotropic random process. Denote by $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)$ the values of the process observed at spatial locations $\mathbf{x}_1, \dots, \mathbf{x}_n$, respectively.

The new kernel semivariogram estimator proposed in the case of clustered data is defined as follows:

$$\hat{\gamma}(u) = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{1}{\sqrt{n_i \times n_j}} \times K\left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h}\right) [Z(\mathbf{x}_i) - Z(\mathbf{x}_j)]^2}{2 \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\sqrt{n_i \times n_j}} \times K\left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h}\right)}, \quad u \geq 0, \tag{3}$$

where $n_i = \sum_k I_{\{\|\mathbf{x}_i - \mathbf{x}_k\| \leq \delta\}}$ and $n_j = \sum_k I_{\{\|\mathbf{x}_j - \mathbf{x}_k\| \leq \delta\}}$; h and δ represent the bandwidth and neighbourhood radius selectors, respectively.

Remark 1. For the edge effect correction, sometimes required in statistics of spatial analysis based on the counting of neighbours within a given distance, one can consider a local correcting factor given by $n_i/V(B)$, where $V(B) = \pi\delta^2 \cap D$ is the volume of the ball (\mathbf{x}_i, δ) inside the observation region.

The asymptotic results for the expectation and variance of above estimator are presented in the following theorems. Derivation of the these results requires the assumptions introduced in the previous section, leading us to a desirable consistent estimation.

Theorem 1

Assume that conditions (A1)–(A4) are satisfied. Additionally, suppose the convergence rates stated in (A6). Then, for $u > 0$, one has

$$E[\hat{\gamma}(u)] = \gamma(u) + \frac{1}{2} c_K \gamma''(u) h^2 + o(h^2),$$

where $c_K = \int z^2 K(z) dz$.

This theorem also shows that the proposed estimator is asymptotically unbiased. See section 6.1 for a sketch of this proof.

Remark 2. Note that previous result is strictly attained on $u \geq Ch$, although, as stated, theorem 1 would remain valid for any $u > 0$ and large n , since the bandwidth parameter h tends to 0 as n increases. To appreciate the latter, take into account that, under isotropy, the variogram domain is restricted to non-negative values; in addition, the kernel function operates on the distances $\|\mathbf{x}_i - \mathbf{x}_j\| \in [u - Ch, u + Ch]$, where interval $[u - Ch, u + Ch]$ is assumed to be wholly contained within the domain of γ . Then, according to theorem 1, the bias of $\hat{\gamma}(u)$ is of the exact order h^2 for $u \geq Ch$; however, near the endpoint 0, $u < Ch$, an order h rather than h^2 is expected, because of the boundary effect. As suggested in Garcia-Soidán et al. (2004), the adoption of a specific combination of boundary kernels is a possible solution to keep the same rate of convergence.

For the analysis of the asymptotic efficiency, we proceed with the variance result of the proposed variogram estimator. A decreasing variance estimate means a growing efficiency of the estimator, as it will tend to be more accurate.

Theorem 2

Assume the hypotheses required in theorem 1. Additionally, suppose that assumptions (A5) and (A7) are satisfied. Then, for $u > 0$, one has

$$\begin{aligned} \text{Var}[\hat{\gamma}(u)] = & \frac{B_d(u)d_K}{2u^{d-1}A_d^2} E_d(n, a)n^{-2} \lambda^d h^{-1} + \frac{C_d(u)}{A_d^2} F_d(n, a)n^{-1} \\ & + \frac{D_d(u)}{4A_d^2} G_d(n, a)\lambda^{-d} + o(n^{-2} \lambda^d h^{-1} + n^{-1} + \lambda^{-d} + h^4), \end{aligned}$$

where $d_K = \int (K(z))^2 dz$ and $A_d, B_d(u), C_d(u), D_d(u), E_d(n, a), F_d(n, a)$ and $G_d(n, a)$ are as given in (8), (12), (13), (14), (9), (10) and (11), respectively.

The proof of above theorem is sketched in section 6.2.

The fundamental idea behind our asymptotic study is the observation region may be considered expansible, in $D_n = \lambda_n D_0$ [see assumption (A1)]. As the region D_n grows, more locations are expected. As a consequence of stationarity, the spatial dependency, described by the theoretical γ , is kept unchanged. The estimate of γ may suffer some impact resulting from a larger sample size, because new distances may be used in the estimation process.

This reasoning of expansion must occur in a controlled way. A possible negative effect of adopting a too large expansion rate is to make the ‘radius of influence’, known as the range, into a negligible value when compared to the maximum distance on D_n . This would lead us to a meaningless spatial dependency on the observation region.

Remark 3. With respect to the convergence rates stated in (A6), furthermore, we can take

$$\lambda^d = c_1 n^{c_0} + o(n^{c_0}) \tag{4}$$

for some constants $c_0 > 0$ and $c_1 > 0$. The expansion rate is established by constant c_0 , i.e. a smaller value for c_0 means a slower expansion for D_n .

Bear in mind the above relationship $\lambda^d = O(n^{c_0})$ and suppose O_1, O_2 and O_3 identify the convergence rates of the first, second and third terms of the conditional variance, respectively. Then

$$\begin{aligned} O_1 &= O(n^{-2} \lambda^d h^{-1}) = O(n^{-(2-c_0)} h^{-1}), \\ O_2 &= O(n^{-1}), \\ O_3 &= O(\lambda^{-d}) = O(n^{-c_0}). \end{aligned}$$

As a result from (A6), one has $0 < c_0 < 1$, which conveys that O_2 is of a lesser order than O_1 and O_3 and, therefore, the contribution of O_2 to the variance is asymptotically negligible. As a result of this, an alternative expression for the variance may be written as below.

Corollary 1

Assume the hypotheses required in theorem 2. Additionally, suppose the convergence rate announced in (4) is satisfied. Then, for $u > 0$, one has

$$\text{Var}[\hat{\gamma}(u)] = \frac{B_d(u)d_K}{2u^{d-1}A_d^2} E_d(n, a)n^{-2} \lambda^d h^{-1} + \frac{D_d(u)}{4A_d^2} G_d(n, a)\lambda^{-d} + o\left(n^{-2} \lambda^d h^{-1} + \lambda^{-d} + h^4\right),$$

where $d_K = \int (K(z))^2 dz$ and $A_d, B_d(u), D_d(u), E_d(n, a)$ and $G_d(n, a)$ are as given in (8), (12), (14), (9) and (11), respectively.

4.1. Bandwidth selector h

We now intend to use the information available in the sampled data to make guesses about the optimal kernel bandwidth h . Ideally, we would like these to be educated guesses that are likely to originate a $\hat{\gamma}$ close to the true value of γ . Of course, we need some definition of being ‘close’. One of the most common measures is the MSE, which is defined as

$$\text{MSE}[\hat{\gamma}(u)] = E [(\hat{\gamma}(u) - \gamma(u))^2].$$

Hence, we propose the selection of the optimal kernel bandwidth, as the value that minimizes the MSE function, originating a local bandwidth selector.

Then, according to the previous results, one has

$$\begin{aligned} \text{MSE}[\hat{\gamma}(u)] &= (\text{Bias}[\hat{\gamma}(u)])^2 + \text{Var}[\hat{\gamma}(u)] \\ &\simeq \frac{c_K^2 \gamma''(u)^2}{4} h^4 + \frac{B_d(u) d_K E_d(n, a)}{2u^{d-1} A_d^2} n^{-2} \lambda^d h^{-1} + \frac{D_d(u) G_d(n, a)}{4A_d^2} \lambda^{-d}. \end{aligned}$$

From here, for $u > 0$, the bandwidth parameter that asymptotically minimizes the MSE $[\hat{\gamma}(u)]$ becomes

$$h_{opt}(u) = \left[\frac{B_d(u) d_K E_d(n, a)}{2u^{d-1} A_d^2 c_K^2 \gamma''(u)^2} \right]^{1/5} n^{-2/5} \lambda^{d/5}.$$

Remark 4. Alternatively, one might deal with a global bandwidth parameter, by minimizing the Mean Integrated Square Error, or MISE, defined as

$$\text{MISE}[\hat{\gamma}(u)] = \int_R (\text{MSE}[\hat{\gamma}(u)]) du = \int_R (\text{Bias}[\hat{\gamma}(u)])^2 du + \int_R \text{Var}[\hat{\gamma}(u)] du$$

for some $R \subset [0, +\infty)$. For instance, we may take $R = [m_0, m]$, where $m = \sup\{\|\mathbf{x}_i - \mathbf{x}_j\| : \mathbf{x}_i, \mathbf{x}_j \in D\}$

and some constant $m_0, 0 < m_0 < m$. The resulting optimal bandwidth would be

$$h_{opt} = \left[\frac{\int_R \frac{B_d(u)}{u^{d-1}} du d_K E_d(n, a)}{2A_d^2 c_K^2 \int_R \gamma''(u)^2 du} \right]^{1/5} n^{-2/5} \lambda^{d/5},$$

where $c_K = \int z^2 K(z) dz$, $d_K = \int (K(z))^2 dz$ and $A_d, B_d(u)$ and $E_d(n, a)$ are as given in (8), (12) and (9), respectively.

Both derived local and global bandwidth expressions involve the unknown function $\gamma(u)$. For this purpose, a simple parametric approach, such as the first one presented in Zimmerman & Zimmermman (1991) (see Table 1 in Menezes et al., 2005), may be used to estimate $\gamma(u)$. This parametric estimation can be improved by being incorporated into an iterated non-parametric procedure.

From the convergence rate given in (4), the bandwidth parameter that asymptotically minimizes the MSE $[\hat{\gamma}(u)]$ becomes

$$h_{opt}(u) = \left[c_1 \frac{B_d(u) d_K E_d(n, a)}{2u^{d-1} A_d^2 c_K^2 \gamma''(u)^2} \right]^{1/5} n^{-(2-c_0)/5}. \tag{5}$$

With this selection of the bandwidth parameter, it follows that

$$\text{MSE}[\hat{\gamma}(u)] = O(n^{-4(2-c_0)/5} I_{\{c_0 > \frac{8}{9}\}} + n^{-c_0} I_{\{c_0 \leq \frac{8}{9}\}})$$

and the minimum order is achieved for $c_0 = 8/9$.

Remark 5. In the case of \mathbb{R}^2 the optimal bandwidth in (5) becomes

$$h_{opt}(u) = \left[\frac{B_2(u) d_K E_2(n, a)}{2uA_2^2 c_K^2 \gamma''(u)^2} \right]^{1/5} n^{-2/9}, \tag{6}$$

where we fixed $c_1 = 1$; c_K and d_K are constants dependent on the chosen kernel function; A_2 in (8) results from the isotropy assumption and it simplifies to 2π ; $B_2(u)$ in (12) is a measure of the spatial correlation and, for the Gaussian case, it can be approximated by $8A_2\gamma(u)^2$; and, finally, $E_2(n, a)$ in (9) is a measure of how the points are distributed on D_0 .

4.2. Neighbourhood radius selector δ

This section is concerned with the problem of selecting a suitable neighbourhood radius δ when estimating the semivariogram given in (3). This second smoothing parameter is to be used as a global measure all over the observation region. The estimates of the semivariogram $\gamma(u)$ seem not to be significantly affected by small deviations of the δ value, according to our experience. Bear in mind, however, that if $\delta \rightarrow 0$ then the resulting $\gamma(u)$ estimator tends to be equal to the Nadaraya–Watson one. The other extreme of too large value for δ is not interesting neither, as all locations will be treated in the same way and we aim to upweight areas with low density.

One possibility here is to take the δ quantity as the mode of a histogram of the distances. Thus, based on counts of distances between sample locations, one gets an indicator value to be used globally as the neighbourhood radius.

This approach can be implemented as follows. For a given point and for a list of equispaced distances, one must count how the remaining $n - 1$ points are spread within that list of distances. After repeating this for all n points, the partial sum organized by the distances, give us the distance for the maximum count, i.e a proposal for δ value.

Many methodologies for cluster analysis are directly motivated from those derived for kernel density estimation (see e.g. Wong & Lane, 1983; Silverman, 1986; Cuevas *et al.*, 2001). They support the natural idea that clusters correspond to peaks in the underlying density function f on \mathbb{R}^d . In our case, we are more interested on the density estimation of the distances between locations than on the density estimation of the locations themselves.

So, additionally, we propose two equivalent approaches based on the density estimation of the distances. Suppose $\{\mathbf{x}_i\}_{i=1}^n$ locations in \mathbb{R}^d , then define

$$d_j = \|\mathbf{x}_i - \mathbf{x}_k\|, \quad j = 1, \dots, \frac{n(n-1)}{2}.$$

First, a kernel estimation is applied on equispaced distances, ranging from the lowest to the largest sampled distance d_j . An alternative maybe that of applying the estimation on the sampled distances themselves. The δ quantity may then be derived from the maximum of these functions or even from, for instance, the 10 per cent highest values. The quantities obtained are analogous.

In Fig. 2, two examples of sampling designs are considered in the left panels. The first suggests the presence of a cluster and the second suggests the presence of two clusters. In the right panels, the corresponding density estimation of the distances between sample locations are plotted and their global maximum values are marked. The resulting value is a global proposal for δ , so that any point \mathbf{x}_i on the observation region is now associated to a measure n_i for the total number of neighbours within distance δ .

All these naive approaches for the derivation of the δ value present similar results. So, we argue that they all are good candidates to be used as a neighbourhood radius selector.

5. Numerical study

To analyse the performance of the proposed semivariogram estimator for clustered data, simulations of spatial data in IR^2 were carried out. Gaussian data were generated on the observation region $D \subset \mathbb{R}^2$ by selecting a theoretical variogram model to specify the spatial dependency. The region D is assumed to be equal to λD_0 , where D_0 is the bounded and fixed square unit. The new estimator is compared against the estimator of Matheron and the one using the Nadaraya–Watson kernel, both given in (1). The symmetric Epanechnikov kernel was employed in the two previous kernel-type estimators.

The optimal local bandwidth for the robust to clusters estimator can be derived as a function of lag u , as given in (6) with $c_0 = 8/9$ and $c_1 = 1$, so that the corresponding scale factor is $\lambda = n^{4/9}$. We considered a sample size $n = 100$ and a theoretical exponential variogram with a nugget effect of 0.6, a sill of 1.336 and the corresponding range equal to 5.0.

Regarding the above bandwidth expression, note that c_K identifies the variance of the Epanechnikov kernel and d_K identifies the integral of this squared kernel. As we are in \mathbb{R}^2 , the A_2 in (8) is reduced to 2π , and $B_2(u)$ in (12) becomes $16\pi\gamma(u)^2$. As the bandwidth derivation needs itself an estimation of the semivariogram $\gamma(u)$ and its second derivative, a pilot parametric estimation was used for this purpose. To estimate $E_2(n, a)$, given in (9), the existing integrals were numerically approximated to a sample average, as

$$\int g(\mathbf{w}) d\mathbf{w} = \int \frac{g(\mathbf{w})}{f_0(\mathbf{w})} f_0(\mathbf{w}) d\mathbf{w} = E_{f_0} \left[\frac{g(\mathbf{W})}{f_0(\mathbf{W})} \right] \text{ where } f_0(\mathbf{w}) \text{ is the density function of } \mathbf{W}.$$

So, \hat{E}_2 may be derived, as follows:

$$\hat{E}_2(n, \hat{a}) = \frac{\frac{1}{N} \sum_{i=1}^N \hat{f}_0(\mathbf{w}_i) / \hat{H}(\hat{a}, \mathbf{w}_i)^2}{\left(\frac{1}{N} \sum_{i=1}^N \hat{f}_0(\mathbf{w}_i) / \hat{H}(\hat{a}, \mathbf{w}_i) \right)^2},$$

where $N = 5000$ is the number of extra points generated from density f_0 ; $\hat{a} = \hat{\delta}/\lambda$, where $\hat{\delta}$ is obtained as the maximum of a kernel density estimator of the distances; $\hat{H}(\hat{a}, \mathbf{w}_i) = n_i/n$, being n_i the number of original sampled points within the circle of centre \mathbf{w}_i and radius \hat{a} ; and \hat{f}_0 results from a non-parametric density estimation of the spatial locations in D_0 .

A conclusive analysis of the behaviour of the semivariogram estimators included in our comparative study must be based on results from several independent cases. We then generate a total of 100 independent data sets and, for each one, derive the integrated square error (ISE) between the estimator and the theoretical semivariogram. The ISE, defined as $\int_{\alpha}^{\beta} [\hat{\gamma}(u) - \gamma(u)]^2 du$, was approximated numerically through the trapezoid rule. In Table 1, the mean values of the resulting ISEs are compared for two distinct sampling designs:

- A CSR design, where spatial locations are uniformly distributed on D ;
- A clustered design, where 40% of the total spatial locations are gathered together into one sub-region of D .

As the observation region D depends on λ , we decided to group the mean values of the ISEs into four classes of lags: $(0, 0.6\lambda)$, $(0, 0.3\lambda)$, $(0, 0.2\lambda)$ and $(0, 0.1\lambda)$. To easily compare columns, all ISE values were standardized by dividing them by the corresponding integral interval, $\beta - \alpha$.

According to Table 1, the new empirical estimator from (3), named ‘RobCluster’, offers a better performance in the presence of clustered data. Under a CSR model, the NW kernel estimator and the new estimator present similar results, and better than those from Matheron’s proposal.

Table 1. Mean values of the standardized ISEs, from the empirical estimators

		$u \leq 0.6\lambda$	$u \leq 0.3\lambda$	$u \leq 0.2\lambda$	$u \leq 0.1\lambda$
CSR	Matheron	1.270	0.943	0.819	0.763
	NW kernel	0.527	0.314	0.276	0.291
	RobCluster	0.500	0.307	0.276	0.298
CLUSTER	Matheron	1.519	1.141	0.889	0.568
	NW kernel	0.582	0.525	0.488	0.400
	RobCluster	0.392	0.294	0.243	0.245

The total number of replicas is 100 and in each replica the total sample size is 100. NW: Nadaraya–Waston.

Table 2. Mean values of the standardized ISEs, from valid estimators

		$u \leq 0.6\lambda$	$u \leq 0.3\lambda$	$u \leq 0.2\lambda$	$u \leq 0.1\lambda$
CSR	Matheron	0.861	0.672	0.609	0.577
	NW kernel	0.504	0.285	0.223	0.201
	RobCluster	0.479	0.279	0.231	0.214
CLUSTER	Matheron	0.925	0.848	0.764	0.519
	NW kernel	0.490	0.475	0.473	0.385
	RobCluster	0.364	0.278	0.229	0.217

The total number of replicas is 100 and in each replica the total sample size is 100.
 NW: Nadaraya–Waston.

The three previous empirical estimators are advised not to be used within the inference and prediction context, as they are not valid and may originate a negative mean square prediction error. There are several different approaches to achieve a valid estimator and their performances are compared in Menezes *et al.* (2005). In here, these empirical estimators were fitted through a reiterated WLS criterion to a class of permissible variograms following Shapiro & Botha (1991); for coding purpose, the routine QPROG from IMSL library was used to resolve the quadratic programming problem. See Christakos (1984) for more detail on the issue of permissibility.

In Table 2, we summarize the mean values of the standardized ISEs, considering the valid version of the previous three estimators. Once more, one may confirm the better behaviour of the proposed variogram under clustered data.

Alternatively, a global bandwidth (see Remark 4) could have been chosen. In this case, the optimal expression for bandwidth h would not depend on lag u , as it would depend instead on some integrals of u . Bear in mind, a global bandwidth is expected to originate faster simulations when compared to a local one, as it avoids a specific estimation for each lag u . The natural drawback is to propose a less accurate solution.

6. Proofs

6.1. Proof of theorem 1

Our strategy of proof is similar to the one from Hall *et al.* (1994) and Garcia-Soidán *et al.* (2004), together with the following lemma.

Lemma 1

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

According to this lemma, as

$$\text{Bias} [\hat{\gamma}(u)] = E_{Z, P} [\hat{\gamma}(u) - \gamma(u)] = E_P [E_Z [\hat{\gamma}(u) - \gamma(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]]$$

with P identifying the random process for the spatial locations, then our target becomes the derivation of the order of $E_Z [\hat{\gamma}(u) - \gamma(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]$. Future references to $E_Z [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]$ will be simplified to $E[\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]$.

$$\begin{aligned} E[\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n] &= E \left[\frac{\sum w_{ij}(u) [Z(\mathbf{x}_i) - Z(\mathbf{x}_j)]^2}{2 \sum w_{ij}(u)} \mid \mathbf{V}_1, \dots, \mathbf{V}_n \right] \\ &= \frac{\sum w_{ij}(u) \gamma(\|\mathbf{x}_i - \mathbf{x}_j\|)}{\sum w_{ij}(u)}, \end{aligned}$$

where w_{ij} is given in (2).

Write $a_1(u) = \sum w_{ij}(u)$ and $a_2(u) = \sum w_{ij}(u)\gamma(\|\mathbf{x}_i - \mathbf{x}_j\|)$. The spatial locations have been taken as $\mathbf{x}_i = \lambda\mathbf{v}_i$, $1 \leq i \leq n$; see conditions (A1) and (A2). Then,

$$E[\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n] - \gamma(u) = \frac{a_2(u) - a_1(u)\gamma(u)}{a_1(u)}. \tag{7}$$

Below we shall show that the following orders hold for $u > 0$:

$$a_1(u) = u^{d-1} A_d n \lambda^{-d} h \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w} + o(n\lambda^{-d}h) \quad \text{a.s.},$$

$$a_2(u) - a_1(u)\gamma(u) = \frac{1}{2} c_K \gamma''(u) u^{d-1} A_d n \lambda^{-d} h^3 \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w} + o(n\lambda^{-d}h^3) \quad \text{a.s.},$$

where $H(a, \mathbf{w}) = \int_{\|\mathbf{y}\| \leq a} f_0(\mathbf{w} - \mathbf{y}) d\mathbf{y}$.

Consequently, by considering the latter relations and applying lemma 1 to (7), one obtains

$$E[\hat{\gamma}(u)] - \gamma(u) = E\left[\frac{a_2(u) - a_1(u)\gamma(u)}{a_1(u)}\right] = \frac{1}{2} c_K \gamma''(u) h^2 + o(h^2),$$

which would allow one to conclude theorem 1 is valid.

Order of $a_1(u)$: For $u > 0$, as the kernel function K is compactly supported, one has

$$a_1(u) = \sum_{i \neq j} \frac{K\left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h}\right)}{\sqrt{n_i n_j}} + nK\left(\frac{u}{h}\right) = \sum_{i \neq j} \frac{K\left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h}\right)}{\sqrt{n_i n_j}}.$$

Proceeding in a similar way as in the proof of theorem 3.1 of Garcia-Sóidan et al. (2004), the dominant term of $a_1(u)$ is given by $n^2\alpha$, where

$$\alpha = E\left[\frac{K\left(\frac{u - \lambda\|\mathbf{V}_1 - \mathbf{V}_2\|}{h}\right)}{\sqrt{\sum_{k_1, k_2} I_{\{\lambda\|\mathbf{V}_1 - \mathbf{V}_{k_1}\| \leq \delta, \lambda\|\mathbf{V}_2 - \mathbf{V}_{k_2}\| \leq \delta\}}}}\right].$$

Consider the new random variables $\mathbf{W}_1 = \mathbf{V}_1 - \mathbf{V}_2, \dots, \mathbf{W}_{n-1} = \mathbf{V}_1 - \mathbf{V}_n$.

Have also in mind that a realization of \mathbf{W}_1 obeys to $K\left(\frac{u - \lambda\|\mathbf{w}_1\|}{h}\right) \rightarrow 0$ unless $\|\mathbf{w}_1\| \rightarrow 0$. This happens as K is compactly supported, i.e. $K(z) = 0$ if $|z| > C$, meaning that $\lambda^{-1}(u - Ch) \leq \|\mathbf{w}_1\| \leq \lambda^{-1}(u + Ch)$. Then

$$\alpha = \int \dots \int \frac{K\left(\frac{u - \lambda\|\mathbf{w}_1\|}{h}\right) f_{n-1}(\mathbf{w}_1, \dots, \mathbf{w}_{n-1})}{\sqrt{1 + 2 \sum_{k_1=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq \frac{\delta}{\lambda}\}} + \sum_{k_1, k_2=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq \frac{\delta}{\lambda}, \|\mathbf{w}_{k_2} - \mathbf{w}_1\| \leq \frac{\delta}{\lambda}\}}}} d\mathbf{w}_1 \dots d\mathbf{w}_{n-1}.$$

As a marginal distribution, $f_{n-1}(\mathbf{w}_1, \dots, \mathbf{w}_{n-1})$ can be written as $\int f_n(\mathbf{w}, \mathbf{w}_1, \dots, \mathbf{w}_{n-1}) d\mathbf{w} = \int f_0(\mathbf{w}) f_0(\mathbf{w} - \mathbf{w}_1) \dots f_0(\mathbf{w} - \mathbf{w}_{n-1}) d\mathbf{w}$, and because of assumption (A7), we obtain

$$\alpha = \int \dots \int \frac{K\left(\frac{u - \lambda\|\mathbf{w}_1\|}{h}\right) f_0(\mathbf{w}) f_0(\mathbf{w} - \mathbf{w}_1) \dots f_0(\mathbf{w} - \mathbf{w}_{n-1})}{\sqrt{\left(1 + 2 \sum_{k_1=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}}\right) + \left(\sum_{k_1, k_2=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a, \|\mathbf{w}_{k_2} - \mathbf{w}_1\| \leq a\}}\right)}} d\mathbf{w} d\mathbf{w}_1 \dots d\mathbf{w}_{n-1}.$$

The expression under the square root of α may be simplified as follows

$$\begin{aligned} & \left(1 + 2 \sum_{k_1=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}}\right) + \left(\sum_{k_1, k_2=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a, \|\mathbf{w}_{k_2} - \mathbf{w}_1\| \leq a\}}\right) \\ &= \left(3 + 2 \sum_{k_1=2}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}}\right) + \left(\sum_{k_1=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{\substack{k_1, k_2=1 \\ k_1 \neq k_2}}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a, \|\mathbf{w}_{k_2}\| \leq a\}}\right) \end{aligned}$$

$$\begin{aligned}
 &= \left(3 + 2 \sum_{k_1=2}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} \right) + \left(1 + 3 \sum_{k_1=2}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{\substack{k_1, k_2=2 \\ k_1 \neq k_2}}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} I_{\{\|\mathbf{w}_{k_2}\| \leq a\}} \right) \\
 &= 4 + 5 \sum_{k_1=2}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{\substack{k_1, k_2=2 \\ k_1 \neq k_2}}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} I_{\{\|\mathbf{w}_{k_2}\| \leq a\}}.
 \end{aligned}$$

Then

$$\alpha = \int \dots \int \frac{K\left(\frac{u-\lambda\|\mathbf{w}_1\|}{h}\right) f_0(\mathbf{w})^2 f_0(\mathbf{w}-\mathbf{w}_2) \dots f_0(\mathbf{w}-\mathbf{w}_{n-1})}{\sqrt{4 + 5 \sum_{k_1=2}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{\substack{k_1, k_2=2 \\ k_1 \neq k_2}}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} I_{\{\|\mathbf{w}_{k_2}\| \leq a\}}}} d\mathbf{w} \dots d\mathbf{w}_{n-1}.$$

Now, convert $\mathbf{w}_1 = (w^{(1)}, \dots, w^{(d)})$ to spherical polar coordinates with the transformation

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

where $\sin \theta_0 = \cos \theta_d = 1$, $0 \leq \theta_{d-1} < 2\pi$ and $0 \leq \theta_i < \pi$, for $i = 1, \dots, d-2$. The corresponding Jacobian transformation is given by

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

Furthermore, suppose $\sum_{k_1 \geq 2} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} = k$ and apply some basic combinatory rules to obtain

$$\begin{aligned}
 \alpha &= \left(\int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^{m_0} r^{d-1} J_d(\theta_1, \dots, \theta_{d-1}) K\left(\frac{u-\lambda r}{h}\right) dr d\theta_1 \dots d\theta_{d-1} \right) \\
 &\quad \times \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} \int f_0(\mathbf{w})^2 H(a, \mathbf{w})^k (1 - H(a, \mathbf{w}))^{n-2-k} d\mathbf{w}}{\sqrt{4 + 5k + k(k-1)}},
 \end{aligned}$$

where $m_0 = \sup\{\|x\| : x \in D_0\}$ and $H(a, \mathbf{w}) = \int_{\|y\| \leq a} f_0(\mathbf{w}-\mathbf{y}) dy$.

Finally, with the following change of variable

$$t = h^{-1}(u - \lambda r) \Rightarrow r = \lambda^{-1}(u - th) \Rightarrow dr = -\lambda^{-1} h dt,$$

and as K is compactly supported, it becomes

$$\begin{aligned}
 \alpha &= \left(\int \dots \int J_d(\theta_1, \dots, \theta_{d-1}) d\theta_1 \dots d\theta_{d-1} \right) \left(\int_{\frac{u-m_0\lambda}{h}}^{\frac{u}{h}} (\lambda^{-1}(u-th))^{d-1} K(t)\lambda^{-1} h dt \right) \\
 &\quad \times \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} \int f_0(\mathbf{w})^2 H(a, \mathbf{w})^k (1 - H(a, \mathbf{w}))^{n-2-k} d\mathbf{w}}{k+2} \\
 &\simeq u^{d-1} A_d \lambda^{-d} h \int f_0(\mathbf{w})^2 \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} H(a, \mathbf{w})^k (1 - H(a, \mathbf{w}))^{n-2-k}}{k+2} d\mathbf{w},
 \end{aligned}$$

where

$$A_d = \int \dots \int J_d(\theta_1, \dots, \theta_{d-1}) d\theta_1 \dots d\theta_{d-1}. \tag{8}$$

The foregoing expression for α tends to

$$u^{d-1} A_d \lambda^{-d} h \frac{1}{n} \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w}$$

by applying known properties of the binomial theorem and the binomial distribution as sketched next

$$\begin{aligned} \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} H^k (1-H)^{n-2-k}}{k+2} &= \frac{1}{n(n-1)H^2} \sum_{k=0}^{n-2} \binom{n}{k+2} (k+1) H^{k+2} (1-H)^{n-(k+2)} \\ &= \frac{1}{n(n-1)H^2} \left\{ \sum_{k=0}^n \binom{n}{k} k H^k (1-H)^{n-k} - \sum_{k=0}^n \binom{n}{k} H^k (1-H)^{n-k} + (1-H)^n \right\} \simeq \frac{1}{nH}. \end{aligned}$$

So

$$\int f_0(\mathbf{w})^2 \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} H(a, \mathbf{w})^k (1-H(a, \mathbf{w}))^{n-2-k}}{k+2} d\mathbf{w} = \frac{1}{n} \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} + o\left(\frac{1}{n}\right)$$

because of condition (A3), which asserts that the integral performed over variable \mathbf{w} is bounded.

Order of $a_2(u) - a_1(u)\gamma(u)$: Similarly, for $u > 0$, one has

$$a_2(u) - a_1(u)\gamma(u) = \sum_{i \neq j} \frac{K\left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h}\right)}{\sqrt{n_i n_j}} (\gamma(\|\mathbf{x}_i - \mathbf{x}_j\|) - \gamma(u)),$$

and one may prove that the dominant term in here is given by $n^2\beta$, where

$$\beta = E \left[\frac{K\left(\frac{u - \lambda \|\mathbf{V}_1 - \mathbf{V}_2\|}{h}\right)}{\sqrt{\sum_{k_1, k_2} I_{\{\lambda \|\mathbf{V}_1 - \mathbf{V}_{k_1}\| \leq \delta, \lambda \|\mathbf{V}_2 - \mathbf{V}_{k_2}\| \leq \delta\}}} (\gamma(\lambda \|\mathbf{V}_1 - \mathbf{V}_2\|) - \gamma(u)) \right].$$

Let us again convert $\mathbf{w}_1 = (w^{(1)}, \dots, w^{(d)})$ to spherical polar coordinates and perform the change of variable $t = h^{-1}(u - \lambda r)$, to obtain that

$$\begin{aligned} \beta &= A_d \left(\int_{\frac{u-m_0\lambda}{h}}^{\frac{u}{h}} (\lambda^{-1}(u - th))^{d-1} K(t) (\gamma(u - th) - \gamma(u)) \lambda^{-1} h dt \right) \\ &\times \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} \int f_0(\mathbf{w})^2 H(a, \mathbf{w})^k (1-H(a, \mathbf{w}))^{n-2-k} d\mathbf{w}}{k+2}. \end{aligned}$$

Asymptotically, by using condition (A4), the expression $(\gamma(u - th) - \gamma(u))$ may be reduced to the second term of its Taylor expansion, i.e. $0.5\gamma''(u)(-th)^2$. Then, the dominant term of β becomes

$$\frac{1}{2} c_K \gamma''(u) u^{d-1} A_d \lambda^{-d} h^3 \frac{1}{n} \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w},$$

where $c_K = \int z^2 K(z) dz$.

6.2. Proof of theorem 2

Let us start by considering that

$$\text{Var} [\hat{\gamma}(u)] = \text{Var} [E [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]] + E [\text{Var} [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]].$$

By using theorem 1 and lemma 1, it is straightforward to see that for $u > 0$

$$\text{Var} [E [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]] = o(h^4).$$

We need now to check that $\text{Var} [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n] = O(n^{-2} \lambda^d h^{-1} + n^{-1} + \lambda^{-d})$, and again by lemma 1, it will lead us to the convergence rate of $E [\text{var} [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n]]$. Consequently, the convergence rate stated in theorem 2 will be proved to be valid.

About the detailed expression obtained for the conditional variance, we have,

$$\begin{aligned} \text{Var} [\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n] &= E \left[(\hat{\gamma}(u) - E[\hat{\gamma}(u) | \mathbf{V}_1, \dots, \mathbf{V}_n] | \mathbf{V}_1, \dots, \mathbf{V}_n)^2 \right] \\ &= E \left[\left(\frac{\sum_{i \neq j} w_{ij}(u) ((Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2 - E[(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2])}{2 \sum_{i \neq j} w_{ij}(u)} \right) \right. \\ &\quad \times \left. \left(\frac{\sum_{k \neq l} w_{kl}(u) ((Z(\mathbf{x}_k) - Z(\mathbf{x}_l))^2 - E[(Z(\mathbf{x}_k) - Z(\mathbf{x}_l))^2])}{2 \sum_{k \neq l} w_{kl}(u)} \right) \right] \\ &= \frac{\sum_{i \neq j} w_{ij}(u) \sum_{k \neq l} w_{kl}(u) \text{Cov} [(Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2, (Z(\mathbf{x}_k) - Z(\mathbf{x}_l))^2]}{(2 \sum_{i \neq j} w_{ij}(u))^2} \\ &= (2a_1(u))^{-2} \sum_{\substack{i \neq j \\ k \neq l}} K \left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right) K \left(\frac{u - \|\mathbf{x}_k - \mathbf{x}_l\|}{h} \right) \\ &\quad \times \frac{1}{\sqrt{n_i n_j}} \frac{1}{\sqrt{n_k n_l}} g(\mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_k, \mathbf{x}_i - \mathbf{x}_l) = \frac{2e_1(u) + 4e_2(u) + e_3(u)}{4(a_1(u))^2}, \end{aligned}$$

where function $g(\cdot)$ is defined according to condition (A5) and

$$\begin{aligned} e_1(u) &= \sum_{i \neq j} \frac{K \left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right)^2 g(\mathbf{x}_i - \mathbf{x}_j, 0, \mathbf{x}_i - \mathbf{x}_j)}{n_i n_j} \quad \Leftarrow (i = k \wedge j = l), \\ e_2(u) &= \sum_{\substack{i \neq j \\ j \neq l}} \frac{K \left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right) K \left(\frac{u - \|\mathbf{x}_j - \mathbf{x}_l\|}{h} \right) g(\mathbf{x}_i - \mathbf{x}_j, 0, \mathbf{x}_i - \mathbf{x}_l)}{\sqrt{n_i n_j} \sqrt{n_j n_l}} \quad \Leftarrow (i = k), \\ e_3(u) &= \sum_{\substack{i \neq j, k, l \\ j \neq k, l \\ k \neq l}} \frac{K \left(\frac{u - \|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right) K \left(\frac{u - \|\mathbf{x}_k - \mathbf{x}_l\|}{h} \right) g(\mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_k, \mathbf{x}_i - \mathbf{x}_l)}{\sqrt{n_i n_j} \sqrt{n_k n_l}}. \end{aligned}$$

Then, according to the results from section 6.1. about $a_1(u)$, and those from the following sections about $e_1(u)$, $e_2(u)$ and $e_3(u)$, we obtain:

$$\frac{e_1(u)}{2(a_1(u))^2} = \frac{B_d(u)d_K}{2u^{d-1}A_d^2} E_d(n, a)n^{-2}\lambda^d h^{-1} + o(n^{-2}\lambda^d h^{-1}) \quad \text{a.s.},$$

$$\frac{e_2(u)}{(a_1(u))^2} = \frac{C_d(u)}{A_d^2} F_d(n, a)n^{-1} + o(n^{-1}) \quad \text{a.s.},$$

$$\frac{e_3(u)}{4(a_1(u))^2} = \frac{D_d(u)}{4A_d^2} G_d(n, a)\lambda^{-d} + o(\lambda^{-d}) \quad \text{a.s.},$$

where

$$E_d(n, a) = \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})^2} d\mathbf{w} \bigg/ \left(\int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w} \right)^2 = O(1), \tag{9}$$

$$F_d(n, a) = \int \frac{f_0(\mathbf{w})^3}{H(a, \mathbf{w})^2} d\mathbf{w} \bigg/ \left(\int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w} \right)^2 = O(1), \tag{10}$$

$$G_d(n, a) = \int \frac{f_0(\mathbf{w})^4}{H(a, \mathbf{w})^2} d\mathbf{w} \bigg/ \left(\int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})} d\mathbf{w} \right)^2 = O(1). \tag{11}$$

The binomial theorem and some properties of the binomial distribution are used to prove the validity of the latter expressions. The corresponding orders result from condition (A3).

Order of $e_1(u)$: The dominant term of $e_1(u)$ is given by $n^2\alpha_1$, where

$$\begin{aligned} \alpha_1 &= \mathbb{E} \left[\frac{K \left(\frac{u-\lambda\|\mathbf{V}_1-\mathbf{V}_2\|}{h} \right)^2 g(\lambda(\mathbf{V}_1 - \mathbf{V}_2), 0, \lambda(\mathbf{V}_1 - \mathbf{V}_2))}{\sum_{k_1, k_2} I_{\{\|\mathbf{V}_1-\mathbf{v}_{k_1}\| \leq a, \|\mathbf{V}_2-\mathbf{v}_{k_2}\| \leq a\}}} \right] \\ &= \int \dots \int \frac{K \left(\frac{u-\lambda\|\mathbf{w}_1\|}{h} \right)^2 g(\lambda\mathbf{w}_1, 0, \lambda\mathbf{w}_1) f_0(\mathbf{w})^2 f_0(\mathbf{w} - \mathbf{w}_2) \dots f_0(\mathbf{w} - \mathbf{w}_{n-1})}{4 + 5 \sum_{k_1=2}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{\substack{k_1, k_2=2 \\ k_1 \neq k_2}}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} I_{\{\|\mathbf{w}_{k_2}\| \leq a\}}} d\mathbf{w} d\mathbf{w}_1 \dots d\mathbf{w}_{n-1}. \end{aligned}$$

As in section 6.1, we may convert \mathbf{w}_1 to spherical polar coordinates and make a change of variable to obtain

$$\begin{aligned} \alpha_1 &= \int_{\frac{u-m_0\lambda}{h}}^{\frac{u}{h}} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} J_d(\theta_1, \dots, \theta_{d-1}) \left(\lambda^{-1}(u - th) \right)^{d-1} K(t)^2 \lambda^{-1} h \\ &\quad \times g \left((u - th) \left(\cos \theta_1, \dots, \prod_{j=0}^{d-1} \sin \theta_j \right), 0, (u - th) \left(\cos \theta_1, \dots, \prod_{j=0}^{d-1} \sin \theta_j \right) \right) dt d\theta_1 \dots d\theta_{d-1} \\ &\quad \times \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} \int f_0(\mathbf{w})^2 H(a, \mathbf{w})^k (1 - H(a, \mathbf{w}))^{n-2-k} d\mathbf{w}}{(k + 2)^2}. \end{aligned}$$

The dominant term of α_1 is given by

$$u^{d-1} B_d(u) d_K \lambda^{-d} h \int f_0(\mathbf{w})^2 \sum_{k=0}^{n-2} \frac{\binom{n-2}{k} H(a, \mathbf{w})^k (1 - H(a, \mathbf{w}))^{n-2-k}}{(k + 2)^2} d\mathbf{w},$$

where $d_K = \int (K(z))^2 dz$ and

$$B_d(u) = \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} J_d(\theta_1, \dots, \theta_{d-1}) \times g \left(u \left(\cos \theta_1, \dots, \prod_{j=0}^{d-1} \sin \theta_j \right), 0, u \left(\cos \theta_1, \dots, \prod_{j=0}^{d-1} \sin \theta_j \right) \right) d\theta_1 \dots d\theta_{d-1}. \tag{12}$$

By the binomial theorem, we can approximate $\sum_{k=0}^{n-2} \binom{n-2}{k} H^k (1-H)^{n-2-k} / (k+2)^2$ to $1/(n^2 H(a, \mathbf{w})^2)$, so the foregoing expression for α_1 simplifies to

$$u^{d-1} B_d(u) d_K \lambda^{-d} h \frac{1}{n^2} \int \frac{f_0(\mathbf{w})^2}{H(a, \mathbf{w})^2} d\mathbf{w}.$$

Order of $e_2(u)$: In here, three distinct indices i, j, l , are involved, thus the dominant term of $e_2(u)$ is given by $n^3 \alpha_2$, where

$$\alpha_2 = E \left[\frac{K \left(\frac{u-\lambda \|\mathbf{V}_1 - \mathbf{V}_2\|}{h} \right) K \left(\frac{u-\lambda \|\mathbf{V}_1 - \mathbf{V}_3\|}{h} \right) g(\lambda(\mathbf{V}_1 - \mathbf{V}_2), 0, \lambda(\mathbf{V}_1 - \mathbf{V}_3))}{\sqrt{\sum_{k_1, k_2} I_{\{\|\mathbf{V}_1 - \mathbf{V}_{k_1}\| \leq a, \|\mathbf{V}_2 - \mathbf{V}_{k_2}\| \leq a\}}} \sqrt{\sum_{k_1, k_2} I_{\{\|\mathbf{V}_1 - \mathbf{V}_{k_1}\| \leq a, \|\mathbf{V}_3 - \mathbf{V}_{k_2}\| \leq a\}}} \right].$$

As K is compactly supported, we shall have that the dominant term of the expectation above can be reduced to those values $\|\mathbf{w}_1\|$ and $\|\mathbf{w}_2\|$ tending to 0. Then, it becomes

$$\alpha_2 = \int \dots \int \frac{K \left(\frac{u-\lambda \|\mathbf{w}_1\|}{h} \right) K \left(\frac{u-\lambda \|\mathbf{w}_2\|}{h} \right) g(\lambda \mathbf{w}_1, 0, \lambda \mathbf{w}_2)}{\sqrt{1 + 2 \sum_{k_1=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{k_1, k_2=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a, \|\mathbf{w}_{k_2} - \mathbf{w}_1\| \leq a\}}}} \times \frac{f_0(\mathbf{w}) f_0(\mathbf{w} - \mathbf{w}_1) \dots f_0(\mathbf{w} - \mathbf{w}_{n-1})}{\sqrt{1 + 2 \sum_{k_1=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{k_1, k_2=1}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a, \|\mathbf{w}_{k_2} - \mathbf{w}_2\| \leq a\}}}} d\mathbf{w} d\mathbf{w}_1 \dots d\mathbf{w}_{n-1}.$$

After the denominator of the previous expression being simplified, we have:

$$\alpha_2 = \int \dots \int \frac{K \left(\frac{u-\lambda \|\mathbf{w}_1\|}{h} \right) K \left(\frac{u-\lambda \|\mathbf{w}_2\|}{h} \right) g(\lambda \mathbf{w}_1, 0, \lambda \mathbf{w}_2)}{9 + 7 \sum_{k_1=3}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} + \sum_{\substack{k_1, k_2=3, \\ k_1 \neq k_2}}^{n-1} I_{\{\|\mathbf{w}_{k_1}\| \leq a\}} I_{\{\|\mathbf{w}_{k_2}\| \leq a\}}}} \times f_0(\mathbf{w})^3 f_0(\mathbf{w} - \mathbf{w}_3) \dots f_0(\mathbf{w} - \mathbf{w}_{n-1}) d\mathbf{w} \dots d\mathbf{w}_{n-1}.$$

Now, submitting \mathbf{w}_1 and \mathbf{w}_2 to the usual conversion of coordinates and changes of variable, the dominant term becomes

$$\alpha_2 = \lambda^{-2} h^2 \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} J_d(\theta_{1,1}, \dots, \theta_{d-1,1}) J_d(\theta_{1,2}, \dots, \theta_{d-1,2}) \times g \left(u \left(\cos \theta_{1,1}, \dots, \prod_{j=0}^{d-1} \sin \theta_{j,1} \right), 0, u \left(\cos \theta_{1,2}, \dots, \prod_{j=0}^{d-1} \sin \theta_{j,2} \right) \right) d\theta_{1,1} \dots d\theta_{d-1,1} d\theta_{1,2} \dots d\theta_{d-1,2} \times \left(\int_{\frac{u-\lambda m_0}{h}}^{\frac{u}{h}} \int_{\frac{u-\lambda m_0}{h}}^{\frac{u}{h}} (\lambda^{-1}(u-t_1 h))^{d-1} (\lambda^{-1}(u-t_2 h))^{d-1} K(t_1) K(t_2) dt_1 dt_2 \right) \times \sum_{k=0}^{n-3} \frac{\binom{n-3}{k} \int f_0(\mathbf{w})^3 H(a, \mathbf{w})^k (1-H(a, \mathbf{w}))^{n-3-k} d\mathbf{w}}{9 + 7k + k(k-1)}$$

$$\simeq u^{2(d-1)} C_d(u) \lambda^{-2d} h^2 \int f_0(\mathbf{w})^3 \sum_{k=0}^{n-3} \frac{\binom{n-3}{k} H(a, \mathbf{w})^k (1 - H(a, \mathbf{w}))^{n-3-k}}{(k+3)^2} d\mathbf{w},$$

where

$$C_d(u) = \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} J_d(\theta_{1,1}, \dots, \theta_{d-1,1}) J_d(\theta_{1,2}, \dots, \theta_{d-1,2}) \\ \times g\left(u\left(\cos \theta_{1,1}, \dots, \prod_{j=0}^{d-1} \sin \theta_{j,1}\right), 0, u\left(\cos \theta_{1,2}, \dots, \prod_{j=0}^{d-1} \sin \theta_{j,2}\right)\right) d\theta_{1,1} \dots d\theta_{d-1,1} d\theta_{1,2} \dots d\theta_{d-1,2}. \quad (13)$$

By the binomial theorem, the sum $\sum_{k=0}^{n-3} \left(\binom{n-3}{k} H^k (1 - H)^{n-3-k}\right) / (k+3)^2$ can also be approximated to $1/(n^2 H(a, \mathbf{w})^2)$, so we have that the dominant term of α_2 is given by

$$u^{2(d-1)} C_d(u) \lambda^{-2d} h^2 \frac{1}{n^2} \int \frac{f_0(\mathbf{w})^3}{H(a, \mathbf{w})^2} d\mathbf{w}.$$

Order of $e_3(u)$: In here, four distinct indices i, j, K, l , are involved, thus the dominant term of $e_3(u)$ is given by $n^4 \alpha_3$, where

$$\alpha_3 = E \left[\frac{K\left(\frac{u-\lambda\|\mathbf{V}_1-\mathbf{V}_2\|}{h}\right) K\left(\frac{u-\lambda\|\mathbf{V}_3-\mathbf{V}_4\|}{h}\right) g(\lambda(\mathbf{V}_1-\mathbf{V}_2), \lambda(\mathbf{V}_1-\mathbf{V}_3), \lambda(\mathbf{V}_1-\mathbf{V}_4))}{\sqrt{\sum_{k_1, k_2} I_{\{\|\mathbf{V}_1-\mathbf{V}_{k_1}\| \leq a, \|\mathbf{V}_2-\mathbf{V}_{k_2}\| \leq a\}}} \sqrt{\sum_{k_1, k_2} I_{\{\|\mathbf{V}_3-\mathbf{V}_{k_1}\| \leq a, \|\mathbf{V}_4-\mathbf{V}_{k_2}\| \leq a\}}}} \right].$$

Proceeding in a similar way to the previous sections and considering condition (A5) imposed on the covariance between $Z(\mathbf{x}_i) - Z(\mathbf{x}_j)$ and $Z(\mathbf{x}_k) - Z(\mathbf{x}_l)$, then it follows

$$\alpha_3 \simeq u^{2(d-1)} D_d(u) \lambda^{-3d} h^2 \frac{1}{n^2} \int \frac{f_0(\mathbf{w})^4}{H(a, \mathbf{w})^2} d\mathbf{w},$$

where

$$D_d(u) = \int_0^\delta \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} J_d(\theta_{1,1}, \dots, \theta_{d-1,1}) \\ \times J_d(\theta_{1,2}, \dots, \theta_{d-1,2}) J_d(\theta_{1,3}, \dots, \theta_{d-1,3}) t^{d-1} g\left(u(\cos \theta_{1,1}, \dots, \prod_{j=0}^{d-1} \sin \theta_{j,1}), \right. \\ \left. \left(t \cos \theta_{1,3} - u \cos \theta_{1,2}, \dots, t \prod_{j=0}^{d-1} \sin \theta_{j,3} - u \prod_{j=0}^{d-1} \sin \theta_{j,2}\right), t\left(\cos \theta_{1,3}, \dots, \prod_{j=0}^{d-1} \sin \theta_{j,3}\right)\right) \\ \times dt d\theta_{1,1} \dots d\theta_{d-1,1} d\theta_{1,2} \dots d\theta_{d-1,2} d\theta_{1,3} \dots d\theta_{d-1,3}. \quad (14)$$

Acknowledgements

The authors thank the anonymous referees for helpful suggestions and comments that improved this article. The research was partially supported by grants from the Portuguese and the Spanish research councils.

References

Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research* **20**, 251–265.

- Cressie, N. (1993). *Statistics for spatial data*. John Wiley and Sons Inc., New York.
- Cuevas, A., Febrero, M. & Fraiman, R. (2001). Cluster analysis: a further approach based on density estimation. *Comput. Statist. Data Anal.* **36**, 441–459.
- García-Soidán, P., Febrero-Bande, M. & González-Manteiga, W. (2004). Nonparametric kernel estimation of an isotropic variogram. *J. Statist. Plann. Inference* **121**, 65–92.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press, New York.
- Hall, P., Fisher, I. & Hoffmann, B. (1994). On the parametric estimation of covariance functions. *Ann. Statist.* **22**, 2115–2134.
- Isaaks, E. H. & Srivastava, R. M. (1989). *An introduction to applied geostatistics*. Oxford University Press, New York.
- Journel, A. G. & Huijbregts, C. J. (1978). *Mining geostatistics*. Academic Press, London.
- Kovitz, J. L. & Christakos, G. (2004). Spatial statistics of clustered data. *Stoch. Environ. Res. and Risk Assess.* **18**, 147–166.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246–1266.
- Menezes, R. (2005). *Assessing spatial dependence under non-standard sampling*. PhD Thesis, University of Santiago de Compostela (Spain), ISBN 84-9750-595-6.
- Menezes, R. & Tawn, J. (2003). Análise da presença de clusters na estimação de um semivariograma. *Proc. XI Annual Congress of the Portuguese Society of Statistics, Faro*, 451–457.
- Menezes, R., García-Soidán, P. & Febrero-Bande, M. (2005). A comparison of approaches for valid variogram achievement. *Comput. Statist.* **20**, 623–642.
- Ribeiro, Jr, P. J. & Diggle, P. J. (2001). geoR: A package for geostatistical analysis. *R-News* **1**, 15–18.
- Shapiro, A. & Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Comput. Statist. Data Anal.* **11**, 87–96.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London.
- Stein, M. (1999). *Interpolation of spatial data: some theory for kriging*. Springer, New York.
- Wong, M. A. & Lane, T. (1983). A kth nearest neighbour clustering procedure. *J. Roy. Statist. Soc. Ser. B* **45**, 362–368.
- Zimmerman, D. L. & Zimmerman, M. B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics* **33**, 77–91.

Received August 2005, in final form April 2007

Raquel Menezes, Department of Mathematics for Science and Technology, University of Minho, 4800-058 Guimarães, Portugal.

E-mail: rmenezes@mct.uminho.pt