Precision of systematic sampling and transitive methods

Kiên Kiêu[a]
Sandie Souchet[b]
Jacques Istas[c]
[a]Unit de biomtrie, INRA, F78026 St-Cyr, France
[b]SAMOS, Universit Paris I, F75634 Paris, France
[c]Unit de biomtrie, INRA, F78352 Jouy-en-Josas, France

April 18, 2008

Abstract

The use of the transitive methods for assessing the precision of systematic sampling is discussed. A key point of the transitive methods is the choice of a local model for the covariogram near the origin. The relationship between the regularity of the measurements and the regularity of their covariogram is given. This result is useful for choosing the appropriate covariogram model. A method for estimating the measurement regularity from discrete data is proposed for cases where it cannot be assessed a priori. Stereological applications where sampling is based on geometric probes such as serial sections, point or line grids are also discussed.

AMS Classification: 62G05, 62G20, 62M30, 60D05

Keywords: Spatial statistics; Transitive methods; Stereology; Geometric sampling; Precision assessment; Serial sections; Cavalieri volume estimator;

1 Introduction

Systematic sampling is widely used when investigating spatial structures. In this paper we consider the sampling of the structure by means of some probes, e.g. points, transects, quadrats. The parameter of interest can be expressed as an integral of some function, called the measurement function, over the probe space (or sampling space). The sampling design is systematic such that the probes form a regular pattern, e.g. a point lattice. Standard empirical estimators based on systematic samples are unbiased under uniform sampling conditions.

However precision assessment is usually not an obvious task because of correlations between data. A common approach is to model the correlation structure of the data. The so-called *transitive methods* due to Matheron (1965) are alternative non-parametric and asymptotic methods. They are based on mean squared error (MSE) approximations, involving the behaviour of the covariogram of the measurement function near the origin. In his founding work, Matheron considered mainly applications to mining problems such as the estimation of the total ore quantity from density measurements at a regular grid of points.

More recently the use of transitive methods in stereology has been discussed in a series of papers, see e.g. Thioulouse et al. (1985), Gundersen and Jensen (1987), Cruz-Orive (1989) or Cruz-Orive (1993). Stereology is concerned with statistical inference about quantitative parameters of spatial structures, based on geometric samples such as plane or line sections through the structure. In stereology, systematic sampling is often preferred to simple random sampling (uniformly distributed and independent probes) both for practical and statistical reasons. A simple example is when the parameter Q to be estimated is the volume of a bounded region B of R^3. The volume of B can be expressed as

$$Q = \int_R \text{area}(B \cap \Lambda(x)) dx,$$

where $\Lambda(x)$ is a plane with fixed orientation and position x, e.g. a horizontal plane with height x. Using a systematic set of parallel planes with constant distance T between consecutive planes, see Figure 1, the volume can be estimated by T times the sum of areas measured on the planar sections. This estimator is unbiased if the "first" section to hit the structure B is uniformly distributed in a slice of height T, without any assumption concerning the shape or the orientation of the structure. The volume estimator is known in stereology as the *Cavalieri estimator*, see e.g. Sterio (1984) or Cruz-Orive (1987).

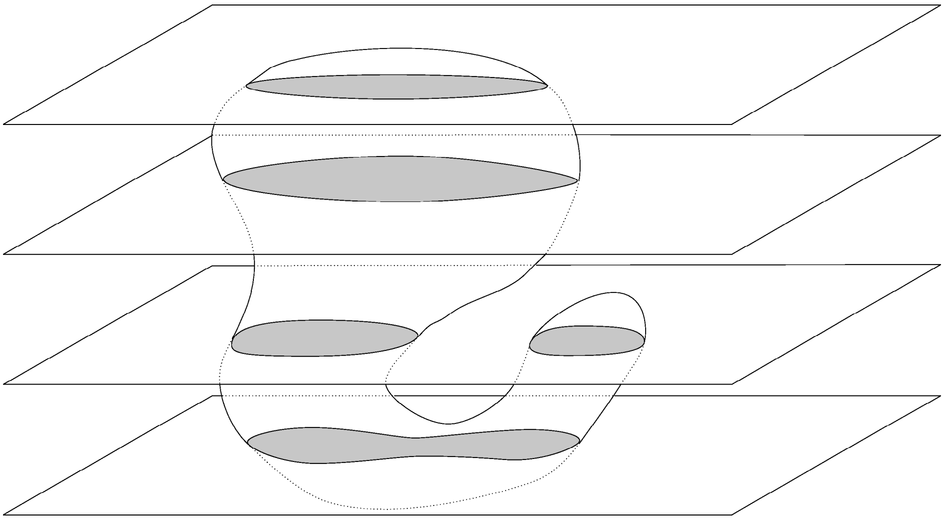

Figure 1: Cavalieri volume estimation. Areas of the structure profile on sections are to be measured. The sum of areas on the series of sections times the distance between 2 consecutive sections is the Cavalieri estimate of the volume of the structure.

In Gundersen and Jensen (1987), Cruz-Orive (1989) and Cruz-Orive (1993), it is emphasized that systematic sampling is rather efficient in stereology compared to simple random sampling. In particular, a number of cases is discussed where the MSE rate of convergence is of the order of N^{-2} where N is the sample size. This rate is to be compared to the rate usually obtained under simple random sampling which is N^{-1}.

The standard method of assessing the precision now used in stereology is based on the assumption that the covariogram has non null slope at the origin. In Cruz-Orive (1993), an example is given where the standard assumption does not hold and an alternative method is discussed. In the present paper the relationship between the smoothness of the measurement function and the behaviour of its covariogram at the origin is further investigated. A new method for MSE estimation is presented. In a first step a smoothness parameter of the measurement function is estimated from the data. In a second step a local model for the covariogram near the origin is chosen according to the estimated smoothness parameter. The final MSE estimator is constructed by fitting the model to the data.

As an illustration we use numerical examples of the Cavalieri procedure.

2 Unbiased estimator of integral characteristics

We consider the general case where the parameter of interest Q can be written as the integral of some *measurement function* f over a *one-dimensional* Euclidean space, i.e.

$$Q = \int_R f(x)\ dx. \tag{1}$$

The observation is supposed to be of the form $\{(x, f(x)) : x \in \mathbf{S}\}$, $\mathbf{S}$ being a random countable subset[1] of R. We shall concentrate on the case where the sample $\mathbf{S}$ is *systematic* and *uniform random*, i.e.

$$\mathbf{S} = \{(\mathbf{U} + k)\, T : k \in Z\}, \quad T > 0,$$

where $\mathbf{U}$ is uniformly distributed in $[0, 1[$. Below, T is referred to as the *sampling period*, $F = T^{-1}$ as the *sampling frequency.* If the measurement function f has a bounded support, then the support length times F is equal to the *mean sample size* (mean number of non-null measurements) which will be denoted by N.

The standard *unbiased* estimator for Q is

$$\widehat{\mathbf{Q}}_T = T \sum_{x \in \mathbf{S}} f(x). \tag{2}$$

Note that $\widehat{\mathbf{Q}}_T$ can be seen as the standard integral approximation from discrete data. The unbiasedness results from the fact that $\mathbf{S}$ samples R uniformly.

Examples of measurement functions associated with the Cavalieri method are shown in Figure 2. The structures behind the shown measurement functions are the unions of 8 tetrahedra with a face parallel to the section planes (Figure 2a), 8 ellipsoids (Figure 2b), and 8

[1] We use bold symbols for random variables.

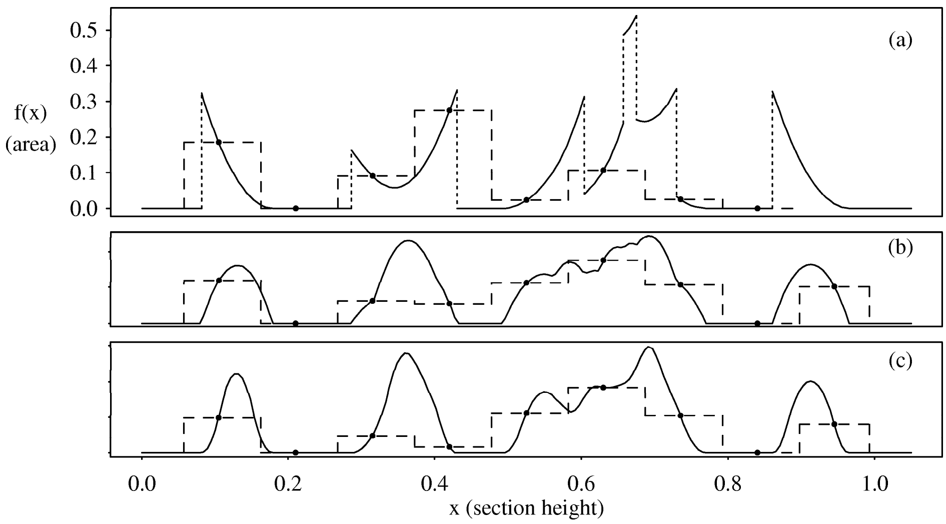

Figure 2: Examples of measurement functions of the Cavalieri method. Curves in solid lines are measurement (area measured on a section) functions. Dots are areas measured on a series of sections. Curves in dashed lines are estimated measurement functions based on the serial sections.

tetrahedra with neither face nor edge parallel to the section planes (Figure 2c), respectively. The sizes and locations of the 8 primary objects are the same through the 3 structures. The jumps observed in Figure 2a occur for sections containing a face of the structure.

The standard Euler-MacLaurin formula, see e.g. Abramowitz and Stegun (1964), provides expansions of the error involved in integral discrete approximations. However the formula cannot be applied when the measurement function is not "smooth" between the sampling points. Such a case may happen for the measurement function shown in Figure 2a which is not continuous. As a matter of fact, the randomness of **S** implies that almost surely the discontinuities occur between the sampling points. In Kiu (1997) and Souchet (1995), an alternative version of the Euler-MacLaurin formula has been derived. This latter formula can be applied when the measurement function is not "smooth" between the sampling points. The functions considered by the authors are called *piecewise smooth functions* and are defined as follows. Let m and p be non-negative integers. A function $h : R \to R$ is said to be (m, p)-

piecewise smooth if the support of h is bounded and h is $(m+p)$-times piecewise differentiable such that

- all h derivatives of order less than m are continuous on R
- the h derivatives of order from m to $m+p$ may not be continuous on a finite set, but their jumps are required to be finite.

Below the set of points where a function h is not continuous is denoted by D_h and the jump of h at a point $x \in D_h$ is denoted by $s_h(x)$. Following the terminology used in Kiu (1997), a jump of a l-th derivative is called a *transition of order l*. The primary transitions are the jumps of the first non-continuous derivative. Therefore the smoothness parameter m is called *the primary transition order.*

Observe that the class of piecewise smooth functions as defined above does not encompass unbounded functions or functions with unbounded derivatives. Such measurement functions appear in connection with the Cavalieri method e.g. if the structure is a cylinder with a revolution axis parallel to the section planes. Also very irregular functions such as Brownian motion are not piecewise smooth. Measurement functions of this type will be the object of future research.

Note that the measurement function shown in

- Figure 2a is $(0, \infty)$-piecewise smooth with transitions of order 0, 1 and 2
- Figure 2b is $(1, \infty)$-piecewise smooth with transitions of order 1 and 2
- Figure 2c is $(2, \infty)$-piecewise smooth with transitions of order 2,

see Figure 3. The transitions of the measurement function involved in the Cavalieri method are related to geometrical features of the structure under study. In the examples considered

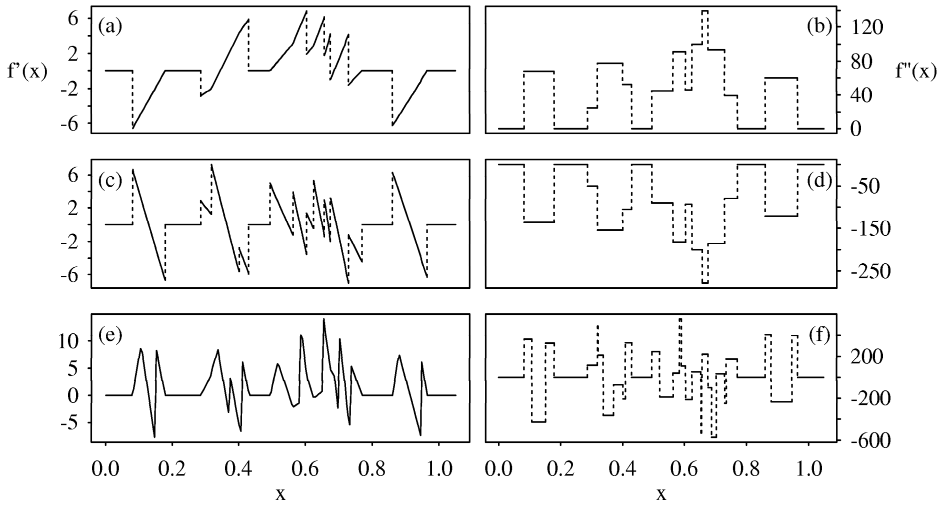

Figure 3: Derivatives of order 1 and 2 for the measurement functions shown in Figure 2. Derivatives of order larger than 2 are identically equal to 0.

in the figures, transitions of order 0 occur for section planes containing a planar face of the structure. Transitions of order 1 occur for section planes tangent to the structure. Transitions of order 2 occur for sections containing a vertex of the structure.

The alternative Euler-MacLaurin formula for piecewise smooth functions provides expansions for the error involved in integral approximations. These expansions are based on transitions as defined above. They involve the so-called Bernoulli polynomials. The Bernoulli polynomials form a sequence of polynomials indexed by non-negative integers. The polynomial with index l will be denoted by P_l and is of degree l. Explicit expressions and formal definitions can be found in e.g. Abramowitz and Stegun (1964). For any integer $l = 0, 1, 2, \ldots$ and $T > 0$, define the periodic function

$$P_{l,T} : x \mapsto P_l \left(\frac{x}{T} - \left[\frac{x}{T} \right] \right)$$

where $[\cdot]$ denotes the integer part. The alternative Euler-MacLaurin formula can now be stated as follows. Let h be a $(m, 1)$-piecewise smooth function with $m \geq 1$. Then for any

$T > 0$,

$$T\sum_{k\in Z} h(kT) - \int_R h(x)\,dx$$
$$= (-1)^m T^{m+1} \sum_{a\in D_{h^{(m)}}} s_{h^{(m)}}(a)\, P_{m+1,T}(a) + \mathrm{o}\left(T^{m+1}\right). \tag{3}$$

This result holds also for $m = 0$ under the extra condition that $D_h \cap ZT = \emptyset$. Detailed proofs are given in Souchet (1995) and Kiu (1997). The derivation of the alternative Euler-MacLaurin formula is based on the same arguments as the standard Euler-MacLaurin formula. The integral expression of the remainder is

$$(-1)^m T^{m+1} \int_R h^{(m+1)}(x)\, P_{m+1,T}(x)\,dx.$$

Now suppose that the measurement function f is $(m,1)$-piecewise smooth. The alternative Euler-MacLaurin formula then provides an expansion of $\widehat{\mathbf{Q}}_T - Q$. Note that if $m = 0$, the extra condition holds almost surely because of the randomness of $\mathbf{S}$. One gets the following expansion which holds almost surely

$$\widehat{\mathbf{Q}}_T - Q = (-1)^m T^{m+1} \sum_{a\in D_{f^{(m)}}} s_{f^{(m)}}(a)\, P_{m+1,T}(a - \mathbf{U}T) + \mathrm{o}\left(T^{m+1}\right). \tag{4}$$

Note that the remainder is stochastic.

Hence the worst rate of convergence is of the order of T and is reached when the measurement function f is not continuous on R. If f is continuous on R but its first derivative is not then the estimation error converges as T^2 etc... The convergence rate can also be expressed in terms of the mean sample size N. The worst rate of convergence is of the order of N^{-1} and is reached when the measurement function f is not continuous on R. If f is continuous on R but its first derivative is not then the estimation error converges as N^{-2} etc... Compare with simple random sampling where the convergence rate does not depend on the regularity of the measurement function and is always of the order of $N^{-1/2}$. The efficiency of systematic sampling compared to simple random sampling was already noticed in Gundersen and Jensen (1987) and Cruz-Orive (1989). These authors have considered mean squared error convergence rather that estimation error convergence.

3 Mean squared error expansion

The formula (4) can be used for approximating the MSE of $\widehat{\mathbf{Q}}_T$. Repeated integration by part yields

$$\mathrm{E}\left[P_{m+1,T}\left(a-\mathbf{U}T\right)P_{m+1,T}\left(b-\mathbf{U}T\right)\right]=\left(-1\right)^{m}P_{2m+2,T}\left(b-a\right).$$

Hence we get the following mean squared error expansion for an $(m,1)$-piecewise smooth function

$$\begin{aligned}&\mathrm{MSE}\left[\widehat{\mathbf{Q}}_T\right] && (5)\\ &= \left(-1\right)^{m}T^{2m+2}\sum_{a,b\in D_{f^{(m)}}} s_{f^{(m)}}\left(a\right)s_{f^{(m)}}\left(b\right)P_{2m+2,T}\left(b-a\right)+\mathrm{o}\left(T^{2m+2}\right).\end{aligned}$$

This expansion has been computed for the examples of measurement functions described in the previous section, see Figure 4 (dashed curves). Note that the convergence rate of the mean squared error is of the order of $T^{2m+2}\propto N^{-2m-2}$.

In Matheron (1965), Gundersen and Jensen (1987) and Cruz-Orive (1989), expansions of the mean squared error involve the so-called covariogram of the measurement function. The covariogram is defined by Matheron (1965)

$$g\left(y\right)=\int_{R}f\left(x+y\right)f\left(x\right)\,dx,\quad y\in R. \tag{6}$$

The covariogram describes the spatial variation of f. Note that the covariogram is an even function. Covariograms of the measurement functions shown in Figure 2 are presented in Figure 5.

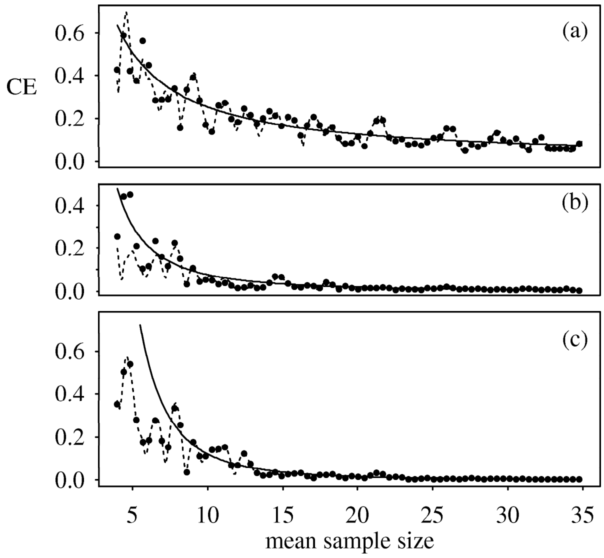

Figure 4: Coefficients of error for Cavalieri volume estimators. Exact coefficients of error (CE) calculated from simulations are represented by dots. Approximated CE based on primary transitions are represented by dashed line curves, cf. (5). Approximated CE based on primary transition amplitudes (extension term) are represented by solid line curves, cf. (9). Underlying structures are described in Section 2.

It is easy to see that

$$\mathrm{MSE}\left[\widehat{\mathbf{Q}}_T\right] = T\sum_{k\in Z} g\left(kT\right) - \int_R g\left(y\right)dy. \tag{7}$$

In order to calculate expansions for (7), Matheron assumed that the measurement function f has a bounded support $[a,b]$ and that its covariogram is non-smooth only at 0, $c = b-a$ and $-c$ (c and $-c$ are the endpoints of the covariogram's support). Expansions of (7) are obtained by applying the standard Euler-MacLaurin formula inside the covariogram's support and by modelling the covariogram at the neighbourhood of 0, c and $-c$ by polynomials. A more direct approach consists in applying the alternative Euler-MacLaurin formula. The latter method provides expansions even if the covariogram is non-smooth inside its support. In view of (3) if the covariogram is $(q,1)$-piecewise smooth, (7) can be expanded as

$$\mathrm{MSE}\left[\widehat{\mathbf{Q}}_T\right] = (-1)^q T^{q+1}\sum_{c\in D_{g^{(q)}}} s_{g^{(q)}}\left(c\right)P_{q+1,T}\left(c\right) + \mathrm{o}\left(T^{q+1}\right). \tag{8}$$

Compare the result above with (5). It is tempting to search for a relationship between the smoothness properties of a measurement function and the smoothness properties of its covariogram. In his paper, Matheron (1965) stated that if the measurement function is differentiable then its covariogram is twice differentiable. Furthermore, Matheron gives an expression of the covariogram second derivative in terms of the first derivative of the measurement function. In Souchet (1995) and Kiu (1997), it is shown that if f is a (m, p)-piecewise smooth function then its covariogram g is $(2m + 1, 2p - 1)$-piecewise smooth, see examples in Figure 5. The authors also derived general expressions of the covariogram derivative in terms of the derivatives of f. In particular it is shown that

$$D_{g^{(2m+1)}} = D_{f^{(m)}} - D_{f^{(m)}}$$

and that

$$s_{g^{(2m+1)}}(c) = -(-1)^m \sum_{\substack{a,b \in D_{f^{(m)}} \\ b-a=c}} s_{f^{(m)}}(a)\, s_{f^{(m)}}(b)\,.$$

In Matheron (1965), the quantity

$$\begin{aligned} E(T) &= -T^{2m+2} P_{2m+2}(0)\, s_{g^{(2m+1)}}(0) \\ &= -2T^{2m+2} P_{2m+2}(0)\, g^{(2m+1)}\left(0^+\right) \end{aligned} \tag{9}$$

is called the *extension term.* Note that the extension term depends only on the amplitudes of the transitions of the measurement function. The sum of the other terms involved in the mean squared error expansions is equal to

$$Z(T) = -T^{2m+2} \sum_{\substack{c \in D_{g^{(2m+1)}} \\ c \neq 0}} P_{2m+2,T}(c)\, s_{g^{(2m+1)}}(c)$$

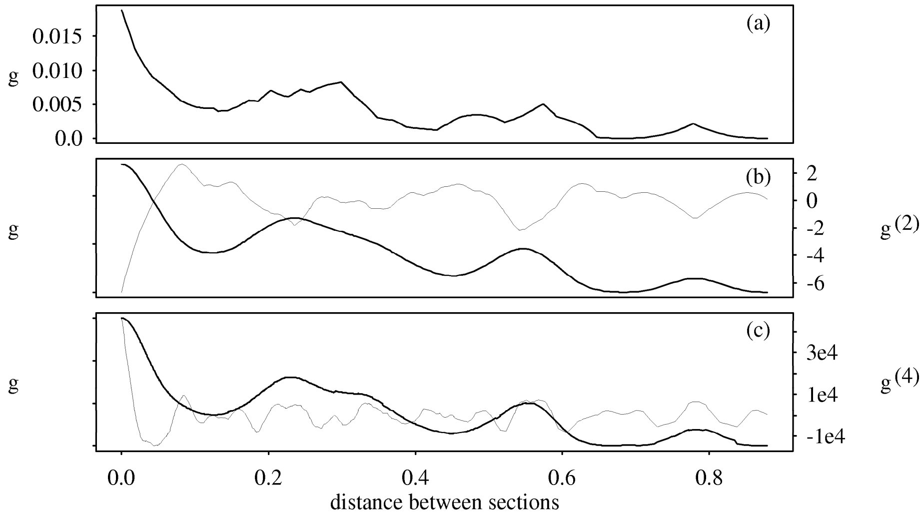

Figure 5: Covariograms (thick line curves) and their last continuous derivatives (thin line curves in (b) and (c)). The corresponding measurement functions are shown in Figure 2.

and is called the *Zitterbewegung*. The extension term represents the central tendency of the mean squared error while the Zitterbewegung is an oscillating function of T. A detailed discussion can be found in Kiu (1997), see also the coefficient of error approximations based only on the extension term in Figure 4 (solid line curves). In Matheron (1965), Gundersen and Jensen (1987) and Cruz-Orive (1989), the extension term is expressed as

$$E(T) = -T^{2m+2}\frac{B_{2m+2}}{m+1}\beta_{2m+1}$$

where $B_{2m+2} = (2m+2)!P_{2m+2}(0)$ is the $(2m+2)$-th Bernoulli number and

$$\beta_{2m+1} = \frac{g^{(2m+1)}(0^+)}{(2m+1)!}.$$

In particular

$$E(T) = \begin{cases} -T^2\beta_1/6 & \text{if } m = 0 \\ T^4\beta_3/60 & \text{if } m = 1 \\ -T^6\beta_5/126 & \text{if } m = 2 \\ T^8\beta_7/120 & \text{if } m = 3. \end{cases}$$

Usually the mean squared error estimation focus on the extension term, see Matheron (1965), Gundersen and Jensen (1987) or Cruz-Orive (1989). These authors argue that when the measurement function is stochastic, the Zitterbewegung is expected to be small in the mean compared to the extension term. In Kiu (1997) the particular case where the measurement function is the indicator function of an interval with random length is considered in details. It is shown that if the interval length density function is Riemann integrable then the mean Zitterbewegung can be neglected when T is sufficiently small.

4 Discrete approximations of the extension term

The extension term as expressed in (9) depends on the covariogram derivatives at the origin. However in general estimates of the covariogram are available only on a discrete set of values. By approximating the covariogram by a polynomial near the origin, the covariogram derivatives at the origin can be interpolated from discrete covariogram data. This yields approximations of the MSE extension term based on discrete covariogram data.

Below we review some of the methods discussed in Matheron (1965), Gundersen and Jensen (1987), Cruz-Orive (1989) and Cruz-Orive (1993). A more general presentation can be found in Kiu (1997). An important feature of these methods is that the primary transition order of the measurement function (or equivalently of its covariogram) is assessed a priori.

Let us start with the case where the covariogram g of the measurement function f is assessed to be $(1,2)$-piecewise smooth. We have seen above that such a covariogram can be obtained from a $(0,2)$-piecewise smooth measurement function[2]. Then by a Taylor expansion in a right-hand side neighbourhood of 0, we get

$$g\left(y\right)=\beta_0+y\beta_1+y^2\beta_2+\mathrm{O}\left(y^3\right), \tag{10}$$

where $\beta_1=g'\left(0^+\right)$. This expansion holds for y "small enough". Now if T is "small enough",

[2] As stated in the previous section, if a function is $(0,2)$-piecewise smooth then its covariogram is $(1,3)$-piecewise smooth. Observe that $(1,3)$-piecewise smoothness is stronger than $(1,2)$-piecewise smoothness.

we can apply the above expansion for $y = 0, T, 2T$ in order to get an approximation of β_1 based on $g(0), g(T)$ and $g(2T)$. Finally one finds

$$E(T) \simeq E_0(T) = \frac{T}{12}\left(3g(0) - 4g(T) + g(2T)\right). \tag{11}$$

This approximation is suggested in Gundersen and Jensen (1987) and Cruz-Orive (1989) for use in geometric sampling including precision assessment of the Cavalieri method. The approximation error is of order T^4 for T "small enough". As a matter of fact this order holds if T is so small that g is 3 times continuously differentiable in $]0, 2T[$. In particular, this latter condition does not hold if

$$D_{g'} \cap]0, 2T[\neq \emptyset,$$

i.e. there are any transitions of the measurement function with order 0 (jumps) separated by a distance less than $2T$. In Kiu (1997) an alternative Taylor formula is derived for piecewise smooth functions. In this case it can be written as

$$g(y) = \beta_0 + y[\beta_1 + \sum_{\substack{c \in D_{g'} \\ 0<c<y}} (1 - \frac{c}{y}) s_{g'}(c)] + y^2\beta_2 + \mathrm{O}\left(y^3\right). \tag{12}$$

Using the latter formula rather than (10) when evaluating the error approximation involved in (11), one finds

$$\begin{aligned} & E_0(T) - E(T) \\ = \; & \frac{T^2}{3}[\sum_{\substack{c \in D_{g'} \\ 0<c<T}} (1 - \frac{c}{T}) s_{g'}(c) - \sum_{\substack{c \in D_{g'} \\ T \le c < 2T}} (1 - \frac{c}{T}) s_{g'}(c)] + \mathrm{O}\left(T^3\right). \end{aligned}$$

According to this result, the approximation error in (11) is of the order of T^2 if there are any transitions of the measurement function with order 0 separated by a distance less than $2T$. Note that then the approximation error is of the same order as the extension term.

The discrete approximation of the extension term (11) can be extended to piecewise smooth functions with arbitrary primary transition order. For instance if the measurement function is $(1,1)$-piecewise smooth, then for positive y "small enough", the following expansion of its covariogram holds

$$g(y) = \beta_0 + y^2\beta_2 + y^3\beta_3 + \mathrm{O}\left(y^4\right),$$

where $\beta_3 = g^{(3)}\left(0^+\right)/6$. There is no linear term in the expansion above because the continuity of the first derivative of the covariogram implies that $g'(0) = 0$. If T is "small enough", we can apply the above expansion for $y = 0, T, 2T$ in order to get an approximation of β_3 based on $g(0), g(T)$ and $g(2T)$. Finally one finds

$$E(T) \simeq E_1(T) = \frac{T}{240}\left(3g(0) - 4g(T) + g(2T)\right). \tag{13}$$

This approximation was first proposed by Cruz-Orive (1993) when assessing the precision of the Cavalieri method for "very regular, quasi-ellipsoidal" structures. As a matter of fact it may be used for any $(1,1)$-piecewise smooth measurement function. For a $(2,1)$-piecewise smooth measurement function, one gets the approximation

$$E(T) \simeq E_2(T) = \frac{T}{8316}\left(10g(0) - 15g(T) + 6g(2T) - g(3T)\right).$$

For a $(3,1)$-piecewise smooth measurement function, one gets the approximation

$$E(T) \simeq E_3(T) = \frac{T}{289920}\left(35g(0) - 56g(T) + 28g(2T) - 8g(3T) + g(4T)\right).$$

A general formula for the discrete approximation of $E(T)$ for a (m,p)-piecewise smooth function is given in Kiu (1997) together with a general discussion of the approximation error.

5 Extension term estimation

When the primary transition order of the measurement function is assessed a priori, the methods for estimating the extension term are straightforward. They consist in replacing the covariogram values involved in the discrete approximations of the extension term by estimates. The *empirical covariogram* is the sequence indexed by Z defined by

$$\widehat{\mathbf{g}}_l = T \sum_{k \in Z} f\left(\left(\mathbf{U}+k\right)T\right) f\left(\left(\mathbf{U}+k+l\right)T\right), \quad l \in Z.$$

The extension term estimators are as follows

$$\text{if } m = 0, \quad \widehat{\mathbf{E}}_0\left(T\right) = \frac{T}{12}\left(3\widehat{\mathbf{g}}_0 - 4\widehat{\mathbf{g}}_1 + \widehat{\mathbf{g}}_2\right) \tag{14}$$

$$\text{if } m = 1, \quad \widehat{\mathbf{E}}_1\left(T\right) = \frac{T}{240}\left(3\widehat{\mathbf{g}}_0 - 4\widehat{\mathbf{g}}_1 + \widehat{\mathbf{g}}_2\right) \tag{15}$$

$$\text{if } m = 2, \quad \widehat{\mathbf{E}}_2\left(T\right) = \frac{T}{8316}\left(10\widehat{\mathbf{g}}_0 - 15\widehat{\mathbf{g}}_1 + 6\widehat{\mathbf{g}}_2 - \widehat{\mathbf{g}}_3\right) \tag{16}$$

$$\text{if } m = 3, \quad \widehat{\mathbf{E}}_3\left(T\right) = \frac{T}{289920}\left(35\widehat{\mathbf{g}}_0 - 56\widehat{\mathbf{g}}_1 + 28\widehat{\mathbf{g}}_2 - 8\widehat{\mathbf{g}}_3 + \widehat{\mathbf{g}}_4\right) \tag{17}$$

These estimators are asymptotically unbiased. For instance consider the estimator $\widehat{\mathbf{E}}_0\left(T\right)$. Since the empirical covariogram is unbiased, we have

$$\mathrm{E}\left[\widehat{\mathbf{E}}_0\left(T\right)\right] = E_0\left(T\right).$$

We have seen in the previous sections that $E_0\left(T\right)$ differs from $E\left(T\right)$ only by an order of T^4 for T small enough and that $E\left(T\right)$ is of the order of T^2. Accordingly we have

$$\lim_{T \to 0} \frac{\mathrm{E}\left[\widehat{\mathbf{E}}_0\left(T\right) - E\left(T\right)\right]}{E\left(T\right)} = 0.$$

This result extends to the other estimators above. Figure 6 shows biases for the examples of measurement functions presented in Section 2. The estimator $\widehat{\mathbf{E}}_0\left(T\right)$ has been used for Figure 6a, estimator $\widehat{\mathbf{E}}_1\left(T\right)$ for Figure 6b, estimator $\widehat{\mathbf{E}}_2\left(T\right)$ for Figure 6c. Note that the bias is rather important for small mean sample sizes.

The critical point of the methods above concerns the assessment of the primary transition order m. As noticed by Cruz-Orive (1993) when using the estimator $\widehat{\mathbf{E}}_0\left(T\right)$ while $m = 1$, one gets in the mean an extension term estimator 20 times greater than the true extension term!

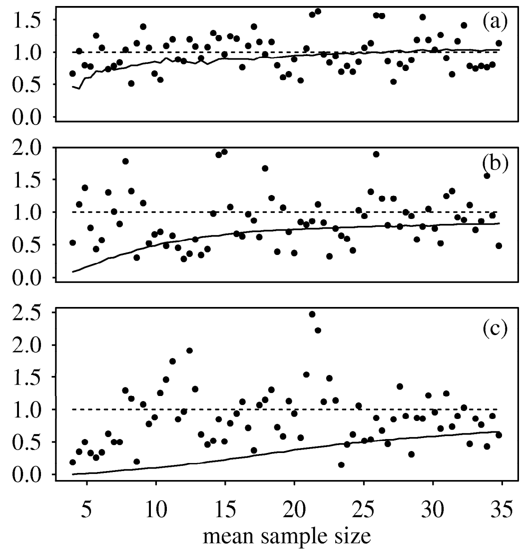

Figure 6: Bias of the extension term estimators. For consistency with Figure 4 showing coefficients of errors, the bias is measured by $\overline{\mathbf{E}}(T)^{\frac{1}{2}}/E(T)^{\frac{1}{2}}$ (solid line curve) where $\overline{\mathbf{E}}(T)$ is the average extension term estimator computed from simulations. The departures (due to the Zitterbewegung) of the exact coefficient of error from the extension term are measured by $\mathrm{CE}\left[\widehat{\mathbf{Q}}_T\right]/E(T)^{\frac{1}{2}}$ (dots). Corresponding measurement functions are shown in Figure 2.

Conversely if $m = 0$ and estimator $\widehat{\mathbf{E}}_1(T)$ is used, one gets in the mean an extension term estimator 20 times less than the true extension term. Assessing the primary transition order of a measurement function is subject to uncertainty when measurements are "complicated". Even for the Cavalieri method, there are no general results relating the smoothness properties of the measurement function to geometrical characteristics of the underlying structure.

6 Robust extension term estimation

In many stereological contexts, *a priori* assessment of the measurement function smoothness is not possible. Either because very little is known *a priori* about the geometry of the structures under study or because the measurement function is so complicated that predicting its smoothness is out of reach. In Istas and Lang (1995) a method for estimating the smoothness of a function from discrete data is presented. The functions considered by the authors are

Gaussian processes. However the method described in Istas and Lang (1995) can be used without modification to piecewise smooth functions. The estimation of the extension term is a two-step procedure: first the primary transition order of the measurement function is estimated from the data, then the extension term is estimated as described in the previous section assuming that the estimated primary transition order is correct.

Let us start with the case where m is not known exactly but can be assumed to be either 0 or 1. Let us consider the estimator $\widehat{\mathbf{E}}_1(T)$. We have

$$\mathrm{E}\left[\widehat{\mathbf{E}}_1(T)\right] \simeq \begin{cases} E(T)/20 = -T^2\beta_1/120 & \text{if } m=0 \\ E(T) = T^4\beta_3/60 & \text{if } m=1. \end{cases}$$

Now suppose that 2 measurement series are available. The sampling period for the first series is T, for the second series αT, $\alpha \neq 1$. In view of the formula given just above, we have

$$\mathrm{E}\left[\widehat{\mathbf{E}}_1(\alpha T)\right] \simeq \alpha^{2m+2}\,\mathrm{E}\left[\widehat{\mathbf{E}}_1(T)\right]$$

in both cases $m=0$ or $m=1$. This relation suggests the following procedure for estimating m from 2 measurement series. Consider the extension term estimators $\widehat{\mathbf{E}}_1(T)$ and $\widehat{\mathbf{E}}_1(\alpha T)$ associated with the 2 measurement series and estimate the primary transition order m by $\widehat{\mathbf{m}}$ such that

$$\widehat{\mathbf{E}}_1(\alpha T) \simeq \alpha^{2\widehat{\mathbf{m}}+2}\widehat{\mathbf{E}}_1(T). \tag{18}$$

In case where only one measurement series is available, measures with another sampling period can be obtained by subsampling. In particular for $\alpha=2$, 2 series with sampling period $2T$ can be extracted from the initial series:

$$\mathbf{S}_{\mathrm{o}} = \{(\mathbf{U}+k)\,T : k \text{ odd}\}, \quad \mathbf{S}_{\mathrm{e}} = \{(\mathbf{U}+k)\,T : k \text{ even}\}.$$

Let $\widehat{\mathbf{E}}_{1,\mathrm{o}}(2T)$, $\widehat{\mathbf{E}}_{1,\mathrm{e}}(2T)$ be the extension term estimators based on the series $\mathbf{S}_{\mathrm{o}}$, $\mathbf{S}_{\mathrm{e}}$ respectively. The mean of $\widehat{\mathbf{E}}_1(2T)$ can be estimated by the average of $\widehat{\mathbf{E}}_{1,\mathrm{o}}(2T)$ and $\widehat{\mathbf{E}}_{1,\mathrm{e}}(2T)$. Then the estimation equation (18) writes

$$\frac{\widehat{\mathbf{E}}_{1,\mathrm{o}}\left(2T\right)+\widehat{\mathbf{E}}_{1,\mathrm{e}}\left(2T\right)}{2}\simeq 2^{2\widehat{\mathbf{m}}+2}\widehat{\mathbf{E}}_{1}\left(T\right).$$

It is easy to see that this latter equation can be rewritten as

$$3\widehat{\mathbf{g}}_0-4\widehat{\mathbf{g}}_2+\widehat{\mathbf{g}}_4\simeq 2^{2\widehat{\mathbf{m}}+1}\left(3\widehat{\mathbf{g}}_0-4\widehat{\mathbf{g}}_1+\widehat{\mathbf{g}}_2\right),$$

i.e.

$$\widehat{\mathbf{m}}=\mathrm{round}\left(\frac{\log\dfrac{3\widehat{\mathbf{g}}_0-4\widehat{\mathbf{g}}_2+\widehat{\mathbf{g}}_4}{3\widehat{\mathbf{g}}_0-4\widehat{\mathbf{g}}_1+\widehat{\mathbf{g}}_4}}{2\log 2}-\frac{1}{2}\right).$$

This principle can be extended to the general case where the primary transition order m is not known exactly but can be assumed to be less than or equal to some value M assessed a priori. The key point is that the extension term estimator obtained when assuming that $m=M$ has a mean equal to

$$c_m E\left(T\right)\simeq c_m T^{2m+2}\beta_{2m+1},$$

where m is the true primary transition order and c_m is a constant depending only on m. The relation above holds only if $m\leq M$. The calculation of the constants c_m is straightforward. For instance let us consider the extension term estimator $\widehat{\mathbf{E}}_3\left(T\right)$ obtained by assuming that $m=3$. If the true value of m is 2, the expansion of the covariogram writes

$$g\left(iT\right)\simeq\beta_0+i^2T^2\beta_2+i^4T^4\beta_4+i^5T^5\beta_5+i^6T^6\beta_6+i^7T^7\beta_7,$$

for $i=0,\ldots,4$. Calculations yield

$$\begin{aligned}\mathrm{E}\left[\widehat{\mathbf{E}}_3\left(T\right)\right]=E_3\left(T\right)&\simeq-\frac{2}{723}T^6\beta_5\\&\simeq\frac{91100}{723}E\left(T\right),\end{aligned}$$

i.e. $c_2 = 91100/723 \simeq 126.0$. Extending the approach described above for the case $m = 0$ or 1, one constructs general estimators of m when it is assumed that $m \leq M$, M being fixed a priori. In particular, if m is assumed to be less than or equal to 3, then m is estimated by

$$\widehat{\mathbf{m}} = \mathrm{round}\left(\frac{\log \dfrac{35\widehat{\mathbf{g}}_0 - 56\widehat{\mathbf{g}}_2 + 28\widehat{\mathbf{g}}_4 - 8\widehat{\mathbf{g}}_6 + \widehat{\mathbf{g}}_8}{35\widehat{\mathbf{g}}_0 - 56\widehat{\mathbf{g}}_1 + 28\widehat{\mathbf{g}}_2 - 8\widehat{\mathbf{g}}_3 + \widehat{\mathbf{g}}_4}}{2\log 2} - \frac{1}{2} \right). \tag{19}$$

Once the transition primary order m is estimated, the extension term is estimated as described in the previous section assuming that $m = \widehat{\mathbf{m}}$. Note that this two-step method is quite unstable. For instance if $\widehat{\mathbf{m}} = 1$, one finds an extension term estimate 20 times less than if $\widehat{\mathbf{m}} = 0$. Therefore a sensible way of measuring the accuracy of the estimator $\widehat{\mathbf{m}}$ is to consider the probability that $\widehat{\mathbf{m}}$ is equal to the true primary transition order. Simulations have been carried out with the measurement functions shown in Figure 2. The estimator (19) has been used: we have assumed that $m \leq 3$ while the true primary transition order is 0 (Figure 2a), 1 (Figure 2b), 2 (Figure 2c). For mean sample sizes larger than 32.6 (measurement function of Figure 2a), 19.6 (measurement function of Figure 2b), 47.0 (measurement function of Figure 2c), the probability that $\widehat{\mathbf{m}} = m$ is close to 1.

7 Discussion

In this paper the use of the transitive mehods for evaluating the precision of estimators based on systematic sampling designs is considered. The problem of choosing a local model of the covariogram at the origin is discussed in details. The described relationship between the smoothness of the measurement function and the smoothness of its covariogram seems to be a useful tool for choosing the appropriate model. For applications in stereology, an important open problem is now to relate geometrical features of the structure of interest to smoothness characteristics of the considered measurement function.

A method for estimating the smoothness of the measurement function from discrete data is proposed. The numerical studies of this method show that it requires rather large samples compared to the usual size of samples in practical stereology. Note that in Kiu (1997), an alternative method is proposed which is expected to be less demanding in terms of sample size. The main idea is to use grids of points with 2 sampling periods, a large one and a small one. Pairs of points close to each other are used to estimate the primary transition order. Numerical investigations of this alternative methods are planned.

In this paper, we have not considered errors either on measurements or on the location of the sampling points. In Kiu (1997), measurement errors are also considered. The error scheme is as follows. It is assumed that the errors are centered, that their variance only depends on the measurement location and that they are uncorrelated given the location of the sampling points. Such a model is appropriate when the measurement at a given location involve some local subsampling. A common example related to the Cavalieri method is when areas on sections are estimated by countings at point grids superimposed on the sections.

Finally note that although we have concentrated in this paper on the case where the sampling space is one-dimensional, the transitive methods can also be used for sampling spaces with higher dimensions, see e.g. Matheron (1965), Gundersen and Jensen (1987), Cruz-Orive (1989) or Chadœuf et al. (1997). Extending the approach described in this paper to higher-dimensional sampling spaces will be the object of future research. As a matter of fact, in stereology sampling spaces may also be special kinds of manifolds. For instance some stereological methods, see e.g. Weibel (1979) or Weibel (1980), involve sampling by lines with various locations and orientations. The space of lines can be seen as a cylinder, see e.g. Stoyan et al. (1987). Therefore using a grid of lines with systematically distributed locations and orientations may be seen as systematic sampling on a cylinder.

Acknowledgements

The interest that Eva B. Vedel Jensen and Hans Jørgen G. Gundersen took in this work has been an encouraging support. The authors are also indebted to Eva B. Vedel Jensen for reading and improving the manuscript.

References

Abramowitz, M. & Stegun, I. A. (1964) *Handbook of Mathematical Functions.* Dover Publications, New-York.

Chadœuf, J., Moran, C. & Goulard, M. (1997) A note on extension variances in R^2. To appear in *Math. geol.*

Cruz-Orive, L. M. (1987) Stereology: historical notes and recent evolution. In: *The commemorative-Memorial Volume. Twenty-Five Years of Stereology* (ed.by R.E. Miles). *Acta Stereol.*, **6**, Suppl. II, 43-56.

Cruz-Orive, L. M. (1989) On the precision of systematic sampling: a review of Matheron's transitive methods. *J. Microsc.*, **153**, 3, 315-33.

Cruz-Orive, L. M. (1993) Systematic sampling in stereology. *Bulletin de l'institut international de statistique*, Actes de la 47ième session, livraison **2**, 451-468.

Gundersen, H. J. G. & Jensen, E. B. (1987) The efficiency of systematic sampling in stereology and its prediction. *J. Microsc.*, **147**, 3, 229-263.

Istas, J. & Lang, G. (1997) Quadratic variations and estimation of the local Hölder index of a Gaussian process. To appear in *Annales de l'institut Henri Poincaré.*

Kiêu, K. (1997) Three lectures on systematic geometric sampling. Memoirs **13**, Department of Theoretical Statistics, University of Aarhus.

Matheron, G. (1965) *Les variables régionalisées et leur estimation.* Masson, Paris.

Sandau, K. (1987) How to estimate the area of a surface using the spatial grid. *Acta Stereol.*, **6**/III, 31-36.

Santaló, L. A. (1976) *Integral Geometry and Geometric Probability.* Encyclopedia of Mathematics and Its Applications. Addison-Wesley, Reading.

Souchet, S. (1995) Précision de l'estimateur de Cavalieri. Rapport de stage, DEA de statistiques et modèles aléatoires appliqués à la finance, Université Paris-VII. Laboratoire de biométrie, INRA-Versailles.

Sterio, D.C. (1984) The unbiased estimation of number and sizes of arbitrary particles using the disector. *J. Microsc.*, **134**, 127-136.

Stoyan, D., Kendall, W.S. & Mecke, J. (1987) *Stochastic Geometry and Its Applications.* Wiley & Sons, Chichester.

Thioulouse, J., Mathy, B. & Ploye, H. (1985) Estimation d'un volume à partir de coupes sériées : sous-échantillonnage, covariogramme transitif et calcul de précision. *Mikroscopie*, **42**, 215-224.

Weibel, E. R. (1979) *Stereological Methods, volume 1, Practical Methods for Biological Morphometry.* Academic Press, London.

Weibel, E. R. (1980) *Stereological Methods, volume 2, Theoretical Fundations.* Academic Press, London.