Anders Rønn-Nielsen, Jon Sporring and Eva B. Vedel Jensen

Estimation of sample spacing in stochastic processes

Estimation of sample spacing in stochastic processes

Anders Rønn-Nielsen[1], Jon Sporring[2] and Eva B. Vedel Jensen[3]

[1]Department of Mathematical Sciences , University of Copenhagen ,
arnielsen@math.ku.dk
[2]Department of Computer Science , University of Copenhagen , sporring@di.ku.dk
[3]Department of Mathematics , Aarhus University , eva@math.au.dk

Abstract

Motivated by applications in electron microscopy, we study the situation where a stationary and isotropic random field is observed on two parallel planes with unknown distance. We propose an estimator for this distance. Under the tractable, yet flexible class of Lévy-based random field models, we derive an approximate variance of the estimator. The estimator and the approximate variance perform well in two simulation studies.

Keywords: Lévy-based modelling, Monte Carlo methods, random fields, stereology, section distance, systematic sampling, variance, variogram.

1 Introduction

Systematic sampling is a wide-spread technique in microscopy, stereology and in the spatial sciences in general. In 3D, systematic sampling is often performed in two steps. In a first step, a stack of equally spaced, parallel plane sections through the spatial structure of interest is generated. In a second step, observations are usually made at a regular lattice of grid points within each section, see e.g. [1, Chapter 7].

Real sampling procedures may involve technically very challenging physical cutting. An example is the generation of ultra thin sections for electron microscopy where the intended distance h between neighbour sections is in the order of nanometers. In such cases, the actual section distance may deviate from the intended distance.

In [9, 10], the effect of such errors on the variance of a Cavalieri type of estimator has been studied. (For a short account of the Cavalieri estimator, see [1, Section 7.1].) Here, the aim is to estimate an integral $\Theta = \int_{\mathbb{R}} f(x) \, dx$, using the estimator $\widehat{\Theta} = h \sum_k f(x_k)$, where $x_k = u + kh \in \mathbb{R}$, $k \in \mathbb{Z}$, is a systematic set of points. A simple geometric example concerns the estimation of the volume of a bounded object in $\mathbb{R}^3$ in which case $f(x)$ may be the area of the intersection between the object and a horizontal plane at height $x \in \mathbb{R}$. Without errors, $x_k - x_{k-1} \equiv h$. In [9, 10], the

1

effect of errors in section positions is studied under different spatial point process models for $\{x_k\}$.

If the aim is to reconstruct the object rather than estimating quantitative properties of the object, it becomes important to know the actual realized distances between neighbour sections. A concrete example of this situation may be found in [8] where ultra thin electron microscopy sections are analysed.

In the present paper, we take up this problem. We study the situation where a stationary and isotropic random field $\{X_v : v \in \mathbb{R}^3\}$ is observed on two parallel planes L_1 and L_2 with unknown distance h. More specifically, we observe the random field at a regular lattice of grid points $\{v_{1i}\}$ on L_1 and, similarly, at $\{v_{2i}\}$ on L_2 where $\|v_{1i} - v_{2i}\| = h$ for all i, see the illustration in Figure 1. The idea is now to obtain information about the correlation structure of the field, using the observed values within sections. This information is then used in the estimation of the unknown distance h between sections.
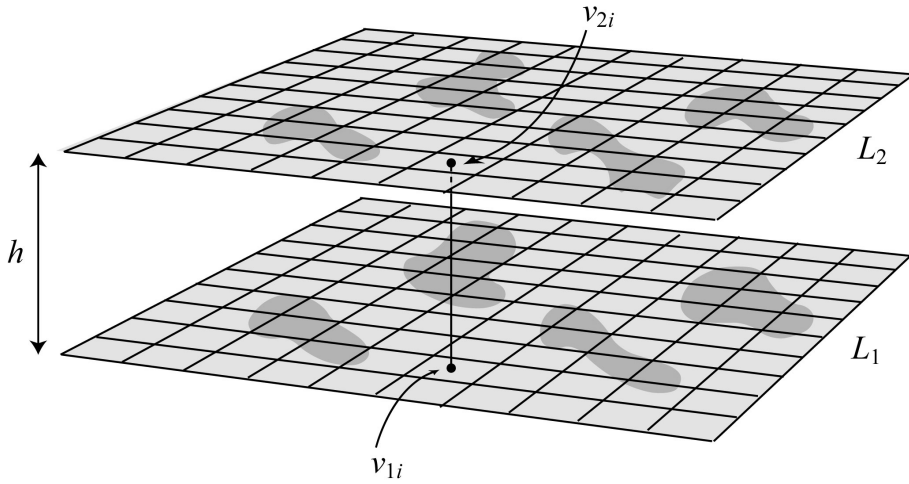

Figure 1: Illustration of the sampling of the random field on two parallel planes a distance h apart.

This procedure has been applied with success in [8]. In the present paper, we study the statistical properties of the resulting estimator of h. Under the tractable, yet flexible class of Lévy-based random field models, we derive an approximate variance of the estimator. We also examine the utility of the estimator and the variance approximation in two simulation studies.

For an introduction to Lévy-based modelling, see e.g. [5]. In relation to variance estimation, Lévy-based models have earlier been used in circular systematic sampling ([4]). Lévy-based models have also been shown to be a useful modelling tool for Cox point processes and growth, see [3] and [6].

The paper is organised as follows. First, we introduce the estimator and explore some of its properties. Under the Lévy-based random field model, we then derive an approximate variance of the estimator and study its performance in two simulation studies. The derivation of the variance formula and some integral calculations

relating to one of the simulation studies are deferred to two appendices.

2 The estimator of section distance

Consider a stationary and isotropic random field $\{X_v : v \in \mathbb{R}^3\}$. Since the random field is stationary and isotropic, the variogram $\mathbb{E}(X_{v_1} - X_{v_2})^2$ only depends on $\|v_1 - v_2\|$ and is thereby determined by $\gamma : [0, \infty) \to [0, \infty)$, where

$$\mathbb{E}(X_{v_1} - X_{v_2})^2 = \gamma(\|v_1 - v_2\|).$$

For a wide range of covariance models, γ is strictly increasing. We will make this assumption in the following.

Let L_1 and L_2 be two parallel planes in $\mathbb{R}^3$ with an unknown distance h. Assume without loss of generality that

$$L_1 = \{(x, y, 0) \mid x, y \in \mathbb{R}\}, \qquad L_2 = \{(x, y, h) \mid x, y \in \mathbb{R}\}.$$

We furthermore assume that pairs of data points $(X_{v_{1i}}, X_{v_{2i}})$, $i = 1, \ldots, n$, are observed where, for all $i = 1, \ldots, n$, $v_{1i} \in L_1$, $v_{2i} \in L_2$ and the points are on the form

$$v_{1i} = (x_i, y_i, 0), \qquad v_{2i} = (x_i, y_i, h),$$

see also Figure 1. The goal is to estimate h and to find an approximate expression for the variance of the estimator.

Since

$$S = \frac{1}{n} \sum_{i=1}^{n} (X_{v_{1i}} - X_{v_{2i}})^2$$

is an unbiased estimator of $\gamma(h)$, we propose to estimate h by

$$\hat{h} = \gamma^{-1}(S). \tag{2.1}$$

From a first–order Taylor expansion of γ^{-1} around $\gamma(h)$ we find

$$\hat{h} \approx h + (\gamma^{-1})'(\gamma(h))(S - \gamma(h)) = h + \frac{1}{\gamma'(h)}(S - \gamma(h)).$$

From this it is seen that

$$\mathbb{E}(\hat{h} - h)^2 \approx \frac{1}{\gamma'(h)^2} \mathbb{E}(S - \gamma(h))^2$$

so

$$\mathrm{Var}(\hat{h}) \approx \frac{1}{\gamma'(h)^2} \mathrm{Var}(S).$$

In most applications, the variogram γ is unknown. In order to apply the estimator (2.1), it is therefore needed to replace γ in (2.1) by an estimate, based on the available observations within sections. Since the distances between sample points within sections are known, we can estimate γ by fitting a parametric curve, induced

by one of the known covariance models, to the empirical variogram. For a recent application of this procedure, see [5, Section 4].

Note that if h is large, then the observations on the two planes may be almost uncorrelated, and it becomes difficult or impracticable to apply the estimator. More specifically, let us suppose that

$$\text{Cov}(X_{v_1}, X_{v_2}) \to 0,$$

as $\|v_1 - v_2\| \to \infty$. Then,

$$\gamma(h) \to 2\text{Var}(X_v),$$

as $h \to \infty$. If h is so large that $\gamma(h) \approx 2\text{Var}(X_v)$, then it may happen that $S > 2\text{Var}(X_v)$ and in such cases $\hat{h}$ does not exist.

3   Variance estimation under a Lévy–based model

3.1   The Lévy–based random field model

It is possible to find an explicit expression for γ and $\text{Var}(S)$ under a Lévy-based random field model. Assume that Z is an independently scattered infinitely divisible random measure on $\mathbb{R}^3$. Such a measure is called a Lévy basis, see [2] and references therein for details. The measure Z thereby satisfies that, for $B \subseteq \mathbb{R}^3$, $Z(B)$ has an infinitely divisible distribution and, for $B_1, B_2 \subseteq \mathbb{R}^3$ disjoint, $Z(B_1)$ and $Z(B_2)$ are independent.

We assume that Z is stationary and isotropic and that the field $\{X_v : v \in \mathbb{R}^3\}$ is given by

$$X_v = \int_{\mathbb{R}^3} f(u + v)\, Z(\mathrm{d}u)\,,$$

where f is a kernel function, satisfying that $f(v) = f(\|v\|)$ and $\int_{\mathbb{R}^3} f(v)\,\mathrm{d}v < \infty$. These assumptions make the field $\{X_v : v \in \mathbb{R}^3\}$ stationary and isotropic and furthermore, it is possible to compute cumulants and covariances for the field. Under mild regularity conditions, we have

$$
\begin{aligned}
\text{Cov}(X_{v_1}, X_{v_2}) &= \text{Var}(Z') \int_{\mathbb{R}^3} f(u + v_1) f(u + v_2)\,\mathrm{d}u \\
&= \text{Var}(Z') \int_{\mathbb{R}^3} f(u) f(u + v_1 - v_2)\,\mathrm{d}u \\
&= \text{Var}(Z') K(v_1 - v_2)\,,
\end{aligned}
$$

say, where Z' is the associated spot variable, having the distribution of $Z(B)$, when $B \subseteq \mathbb{R}^3$ has volume 1. Due to stationarity and isotropy, $K(v) = K(\|v\|)$, say, and the variogram is of the form

$$\gamma(h) = 2\,\text{Var}(Z')(K(0) - K(h))\,, \qquad h \geq 0\,. \tag{3.1}$$

Furthermore, the n'th cumulant of X_v can be expressed as

$$\kappa_n(X_v) = \kappa_n(Z') \int_{\mathbb{R}^3} f(u + v)^n\,\mathrm{d}u\,.$$

In Table 1 below, the distributions and the first four cumulants of the spot variable are listed for the gamma, the inverse Gaussian and the normal inverse Gaussian bases. For more details, see [5, Section 5].

Table 1: The spot variable and its cumulants for various Lévy bases.

Basis	Gamma	Inverse Gaussian	Normal inverse Gaussian
$Z(\mathrm{d}u)$	$\Gamma(\alpha\,\mathrm{d}u, \lambda)$	$\mathrm{IG}(\delta\,\mathrm{d}u, \gamma)$	$\mathrm{NIG}(\alpha, \beta, \mu\,\mathrm{d}u, \delta\,\mathrm{d}u)$
Z'	$\Gamma(\alpha, \lambda)$	$\mathrm{IG}(\delta, \gamma)$	$\mathrm{NIG}(\alpha, \beta, \mu, \delta)$
$\kappa_1(Z')$	α/λ	δ/γ	$\mu + \delta\beta/(\alpha^2 - \beta^2)^{1/2}$
$\kappa_2(Z')$	α/λ^2	δ/γ^3	$\delta\alpha^2/(\alpha^2 - \beta^2)^{3/2}$
$\kappa_3(Z')$	$2\alpha/\lambda^3$	$3\delta/\gamma^5$	$3\delta\beta\alpha^2/(\alpha^2 - \beta^2)^{5/2}$
$\kappa_4(Z')$	$6\alpha/\lambda^4$	$15\delta/\gamma^7$	$3\delta(\alpha^2 + 4\beta^2)\alpha^2/(\alpha^2 - \beta^2)^{7/2}$

3.2   The variance of S

In Appendix A, it is shown under the Lévy–based random field model that

$$\mathrm{Var}(S) = \frac{1}{n^2} \sum_{i,j=1}^{n} \left[\kappa_4(Z')\rho_4(v_{1i}, v_{1j}) + 2\,\mathrm{Var}(Z')^2\rho_2(v_{1i}, v_{1j})^2 \right], \qquad (3.2)$$

where

$$\rho_2(v_{1i}, v_{1j}) = \int_{\mathbb{R}^3} \big(f(u + v_{1i}) - f(u + v_{2i})\big)\big(f(u + v_{1j}) - f(u + v_{2j})\big)\,\mathrm{d}u, \qquad (3.3)$$

$$\rho_4(v_{1i}, v_{1j}) = \int_{\mathbb{R}^3} \big(f(u + v_{1i}) - f(u + v_{2i})\big)^2\big(f(u + v_{1j}) - f(u + v_{2j})\big)^2\,\mathrm{d}u. \qquad (3.4)$$

Below, we indicate how this result is derived.

First, notice that

$$X_{v_{1i}} - X_{v_{2i}} = \int_{\mathbb{R}^3} \left[f(u + v_{1i}) - f(u + v_{2i}) \right] Z(\mathrm{d}u),$$

$i = 1, \ldots, n$. For ease of notation, let us write

$$Y_i = X_{v_{1i}} - X_{v_{2i}}, \qquad g_i(u) = f(u + v_{1i}) - f(u + v_{2i}).$$

We find

$$\mathrm{Var}(S) = \mathbb{E}(S^2) - (\mathbb{E}S)^2$$

$$= \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbb{E}(Y_i^2 Y_j^2) - (\mathbb{E}(Y_1^2))^2. \qquad (3.5)$$

These moments can be calculated, using a well–known result for the logarithm of the Laplace transform of an integral with respect to a Lévy basis. The result holds

5

if $\mathbb{E}\big(e^{\epsilon Z'}\big) < \infty$ for some $\epsilon > 0$. This condition is fulfilled for the three Lévy bases listed in Table 1. Let for $\lambda_1, \lambda_2 \in \mathbb{R}$ sufficiently close to 0 and $i, j \in \{1, \ldots, n\}$

$$K_{i,j}(\lambda_1, \lambda_2) = \log \mathbb{E}\big(e^{\lambda_1 Y_i + \lambda_2 Y_j}\big). \tag{3.6}$$

Then, see e.g. [3, (10)],

$$K_{i,j}(\lambda_1, \lambda_2) = \int_{\mathbb{R}^3} \log \mathbb{E}\Big(e^{(\lambda_1 g_i(u) + \lambda_2 g_j(u))Z'}\Big)\, \mathrm{d}u. \tag{3.7}$$

By differentiating (3.6) and (3.7) with respect to λ_1 and λ_2 and equating the two expressions at $\lambda_1 = \lambda_2 = 0$ with each other, we obtain the variance formula (3.2).

For the Gaussian kernel function

$$f(x) = \frac{1}{(2\pi\sigma)^{3/2}} e^{-\frac{1}{2\sigma}\|x\|^2}, \qquad x \in \mathbb{R}^3, \tag{3.8}$$

with $\sigma > 0$, it is possible to calculate ρ_2 and ρ_4 explicitly, see Appendix B. We find

$$\rho_2(v_{1i}, v_{1j}) = \frac{1}{\sqrt{2}} \frac{1}{(2\pi\sigma)^{3/2}} e^{-\frac{1}{4\sigma}\|v_{1i} - v_{1j}\|^2} \Big(1 - e^{-\frac{1}{4\sigma}h^2}\Big) \tag{3.9}$$

and

$$\rho_4(v_{1i}, v_{1j}) = \frac{1}{4(2\pi\sigma)^{9/2}} e^{-\frac{1}{2\sigma}\|v_{1i} - v_{1j}\|^2} \Big(1 - 4e^{-\frac{3}{8\sigma}h^2} + 3e^{-\frac{1}{2\sigma}h^2}\Big). \tag{3.10}$$

These expressions for ρ_2 and ρ_4 may be inserted into (3.2) to get an explicit expression for the variance of S in the case of a Gaussian kernel function.

As we shall see in the next subsection, further simplifications may be obtained by approximating the double sum in (3.2) by integrals.

3.3 Further approximations

Assume that the points $v_{1i} = (x_i, y_i, 0)$ in L_1 constitute a regular lattice of grid points. Let A be the subset of $\mathbb{R}^2$ spanned by the points (x_i, y_i), and assume that A has area a. If the area of the fundamental region of the lattice is sufficiently small, the formula for $\mathrm{Var}(S)$ can be approximated as follows

$$\mathrm{Var}(S) = \frac{1}{n^2} \sum_{i,j=1}^{n} \Big[\kappa_4(Z')\rho_4(v_{1i}, v_{1j}) + 2\,\mathrm{Var}(Z')^2 \rho_2(v_{1i}, v_{1j})^2\Big]$$

$$\approx \frac{1}{a^2} \int_A \int_A \Big[\kappa_4(Z')\rho_4(v_1, v_2) + 2\,\mathrm{Var}(Z')^2 \rho_2(v_1, v_2)^2\Big]\mathrm{d}v_1 \mathrm{d}v_2.$$

If $f(x) \approx 0$ for $\|x\| > d$, where d is substantially smaller than the width of A, we have the further approximation

$$\mathrm{Var}(S) \approx \frac{1}{a^2} \int_A \int_{\mathbb{R}^2} \Big[\kappa_4(Z')\rho_4(v_1, v_2) + 2\,\mathrm{Var}(Z')^2 \rho_2(v_1, v_2)^2\Big]\mathrm{d}v_1 \mathrm{d}v_2.$$

Since both $\rho_2(v_1, v_2)$ and $\rho_4(v_1, v_2)$ only depend on $\|v_1 - v_2\|$, the inner integral does not depend on v_2. Hence, we have

$$\mathrm{Var}(S) \approx \frac{1}{a} \int_{\mathbb{R}^2} \left[\kappa_4(Z')\rho_4(v, v_0) + 2\,\mathrm{Var}(Z')^2 \rho_2(v, v_0)^2 \right] \mathrm{d}v, \qquad (3.11)$$

where $v_0 \in A$ is a fixed point. By the same type of approximation argument, the variance can also be approximated by

$$\mathrm{Var}(S) \approx \frac{1}{a} \int_{A} \left[\kappa_4(Z')\rho_4(v, v_0) + 2\,\mathrm{Var}(Z')^2 \rho_2(v, v_0)^2 \right] \mathrm{d}v. \qquad (3.12)$$

3.4 Monte Carlo methods

If the integrals in (3.11) and (3.12) are too complicated to calculate, a solution could be to estimate them by Monte Carlo methods. Assume for convenience that A contains $v_0 = (0, 0)$, and define furthermore $\bar{h} = (0, 0, h)$ and $\bar{v} = (v_1, v_2, 0)$ for $v = (v_1, v_2) \in A$. Then,

$$\int_A \rho_4(v, v_0)\mathrm{d}v = \int_A \int_{\mathbb{R}^3} \left(f(u + \bar{v}) - f(u + \bar{v} + \bar{h}) \right)^2 \left(f(u) - f(u + \bar{h}) \right)^2 \mathrm{d}u\,\mathrm{d}v$$

and

$$\int_A \rho_2(v, v_0)^2 \mathrm{d}v = \int_A \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(f(u + \bar{v}) - f(u + \bar{v} + \bar{h}) \right) \left(f(u) - f(u + \bar{h}) \right)$$
$$\cdot \left(f(\tilde{u} + \bar{v}) - f(\tilde{u} + \bar{v} + \bar{h}) \right) \left(f(\tilde{u}) - f(\tilde{u} + \bar{h}) \right) \mathrm{d}u\,\mathrm{d}\tilde{u}\,\mathrm{d}v.$$

It follows that $\int_A \rho_4(v, v_0)\mathrm{d}v$ can be estimated from the empirical mean of repeated independent simulations of

$$\frac{\left(f(U + \bar{V}) - f(U + \bar{V} + \bar{h}) \right)^2 \left(f(U) - f(U + \bar{h}) \right)^2}{g(U)h(V)}$$

where $U = (U_1, U_2, U_3)$ and $V = (V_1, V_2)$ are independent, U is simulated according to some probability density g on $\mathbb{R}^3$ and V is simulated according to some probability density h on A. The value of $\int_A \rho_2(v, v_0)^2 \mathrm{d}v$ can be estimated similarly.

4 Simulation studies

4.1 Gaussian kernel function

Assume that f is the Gaussian kernel function (3.8), implying a Gaussian covariance. Then, using (3.9) and (3.10), the approximation (3.11) becomes

$$\text{Var}(S) \approx \frac{1}{a} \int_{\mathbb{R}^2} \left[\kappa_4(Z')\rho_4(v, v_0) + 2\,\text{Var}(Z')^2 \rho_2(v, v_0)^2 \right] \mathrm{d}v$$

$$= \left(\kappa_4(Z') \frac{1}{(2\pi\sigma)^{9/2}} \left(\tfrac{1}{4} + \tfrac{3}{4} e^{-\frac{1}{2\sigma}h^2} - e^{-\frac{3}{8\sigma}h^2} \right) \right.$$

$$\left. + 2\,\text{Var}(Z')^2 \frac{1}{2} \frac{1}{(2\pi\sigma)^3} \left(1 - e^{-\frac{1}{4\sigma}h^2} \right)^2 \right) \cdot \frac{1}{a} \int_{\mathbb{R}^2} e^{-\frac{1}{2\sigma}(v_1^2 + v_2^2)} \, \mathrm{d}v_1 \mathrm{d}v_2$$

$$= \frac{1}{a} \frac{1}{(2\pi\sigma)^2} \left(\kappa_4(Z') \frac{1}{(2\pi\sigma)^{3/2}} \left(\tfrac{1}{4} + \tfrac{3}{4} e^{-\frac{1}{2\sigma}h^2} - e^{-\frac{3}{8\sigma}h^2} \right) \right.$$

$$\left. + \text{Var}(Z')^2 \left(1 - e^{-\frac{1}{4\sigma}h^2} \right)^2 \right).$$

Furthermore, cf. (3.1) and Appendix B, we get with $\bar{h} = (0, 0, h)$

$$\gamma(h) = 2\,\text{Var}(Z') \left(\int_{\mathbb{R}^3} f(u)^2 \, \mathrm{d}u - \int_{\mathbb{R}^3} f(u) f(u + \bar{h}) \, \mathrm{d}u \right)$$

$$= 2\,\text{Var}(Z') \frac{1}{(4\pi\sigma)^{3/2}} \left(1 - e^{-\frac{1}{4\sigma}h^2} \right),$$

from which we find

$$\gamma'(h) = \text{Var}(Z') \frac{1}{(4\pi\sigma)^{3/2}} \frac{h}{\sigma} e^{-\frac{1}{4\sigma}h^2}.$$

Combining these results, we find

$$\text{Var}(\hat{h}) \approx \frac{1}{\gamma'(h)^2} \text{Var}(S)$$

$$\approx \frac{16\pi\sigma^3}{ah^2} \left(\frac{\kappa_4(Z')}{\text{Var}(Z')^2} \frac{1}{(2\pi\sigma)^{3/2}} \left(\tfrac{1}{4} e^{\frac{1}{2\sigma}h^2} + \tfrac{3}{4} - e^{\frac{1}{8\sigma}h^2} \right) + \left(e^{\frac{1}{4\sigma}h^2} - 1 \right)^2 \right). \quad (4.1)$$

In this example, we combine the Gaussian kernel function with a normal inverse Gaussian (NIG) Lévy basis for which the spot variable Z' is $\text{NIG}(\alpha, \beta, \mu, \delta)$-distributed. For such a Lévy basis,

$$\frac{\kappa_4(Z')}{\text{Var}(Z')^2} = \frac{3(\alpha^2 + 4\beta^2)}{\delta\alpha^2\sqrt{\alpha^2 - \beta^2}},$$

see Table 1. In Figure 2, the resulting approximate variance (4.1) of $\hat{h}$ is plotted as a function of h for a concrete choice of parameter values. The chosen values of the parameters give a covariance structure similar to that obtained in a concrete analysis presented in [5]. In Figure 2, the set A has been chosen such that the effective support of the kernel function is much smaller than the width of A.

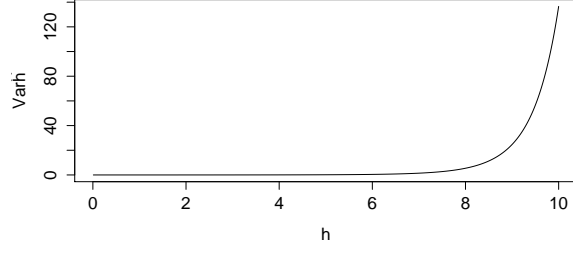Note that the variance increases substantially as a function of h.

Figure 2: A plot of the approximate expression for $\text{Var}(\hat{h})$ as a function of h. The kernel function is Gaussian with $\sigma = 5$. The parameters of the NIG distribution associated with the Lévy basis are $\alpha = 0.6$, $\beta = 0.4$, $\mu = 2.4$ and $\delta = 2$. Furthermore, $A = [0, 100]^2$.

We have simulated from the model described above for four values of h: 0.2, 0.4, 1, 2. In each case, we made 1000 simulated values of $\hat{h}$ each based on simulation of two parallel fields with distance h. From the simulated values of $\hat{h}$ we calculated the empirical mean $\widehat{\mathbb{E}}\hat{h}$ and the empirical variance $\widehat{\text{Var}}(\hat{h})$. The results of this can be seen in Table 2 together with the approximate variance $\overline{\text{Var}}(\hat{h})$, using the formula (4.1), and estimates $\widetilde{\text{Var}}(\hat{h})$ of the approximate variance, based on Monte Carlo simulation. Figure 3 shows four jointly simulated parallel fields from the model such that the last three fields are at distances 0.4, 1 and 2 from the first.

Table 2: For four values of h, $\widehat{\mathbb{E}}\hat{h}$ is the empirical mean of 1000 values of $\hat{h}$ simulated with a Gaussian kernel function and a NIG Lévy basis. The empirical variance is denoted $\widehat{\text{Var}}(\hat{h})$. Furthermore, $\overline{\text{Var}}(\hat{h})$ is the approximate variance, and $\widetilde{\text{Var}}(\hat{h})$ is the Monte Carlo estimate of the approximate variance. The parameters of the model are as specified in Figure 2.

h	0.2	0.4	1	2
$\widehat{\mathbb{E}}\hat{h}$	0.1999647	0.3997637	1.001492	1.994434
$\widehat{\text{Var}}(\hat{h})$	$6.512 \cdot 10^{-5}$	$2.638 \cdot 10^{-4}$	$1.640 \cdot 10^{-3}$	$7.692 \cdot 10^{-3}$
$\overline{\text{Var}}(\hat{h})$	$6.298 \cdot 10^{-5}$	$2.534 \cdot 10^{-4}$	$1.652 \cdot 10^{-3}$	$7.703 \cdot 10^{-3}$
$\widetilde{\text{Var}}(\hat{h})$	$6.313 \cdot 10^{-5}$	$2.587 \cdot 10^{-4}$	$1.675 \cdot 10^{-3}$	$7.951 \cdot 10^{-3}$

It is seen from Table 2 that the bias of $\hat{h}$ is negligible. Furthermore, the variance approximation (4.1) and its evaluation by Monte Carlo methods perform well.

4.2 Exponential kernel function

Now assume that f is an exponential kernel function

$$f(x) = \frac{\sigma^3}{8\pi} e^{-\sigma \|x\|}$$

with $\sigma > 0$. As e.g. shown in [5, Section 3], this kernel function induces a 3rd order autoregressive covariance function for the Lévy-based random field. For this kernel function, expressions for ρ_2 and ρ_4 in the variance approximations (3.11) and (3.12) are not available. Accordingly, the approximate variances need to be evaluated, using the Monte Carlo method. An example is shown in Table 3 for a NIG Lévy basis.
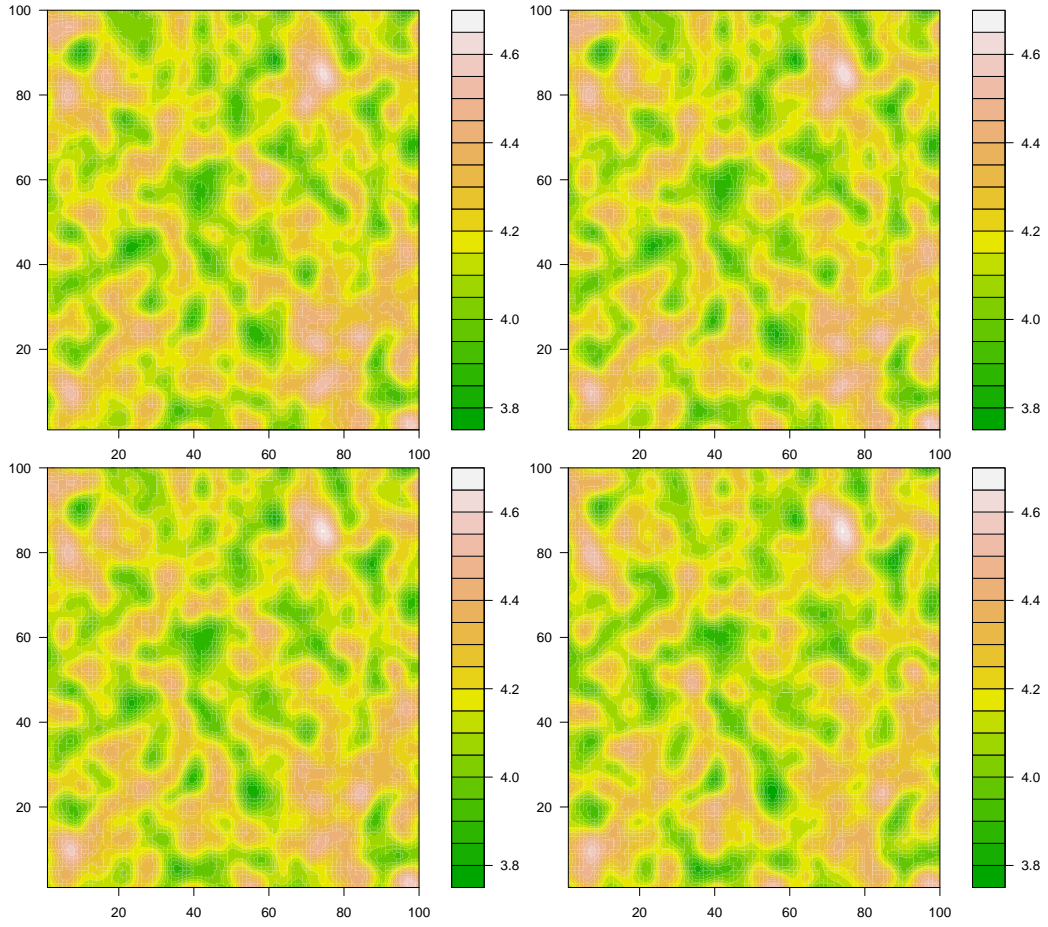
Figure 3: Four jointly simulated parallel fields with NIG Lévy basis and a Gaussian kernel function. The three last fields are at distance 0.4, 1 and 2 from the first. The parameters of the model are as specified in Figure 2.

10

Again, we have simulated 1000 values of $\hat{h}$ for the four choices of h: 0.2, 0.4, 1 and 2.

Table 3: For four values of h, $\widehat{\mathbb{E}}\hat{h}$ is the empirical mean of 1000 values of $\hat{h}$ simulated with an exponential kernel function and a NIG Lévy basis. The empirical variance is denoted $\widehat{\mathrm{Var}}(\hat{h})$. Furthermore, $\widetilde{\mathrm{Var}}(\hat{h})$ is the Monte Carlo estimate of the approximate variance. The parameter of the kernel function is $\sigma = 0.56$, while the parameters of the NIG distribution are as in Figure 2.

h	0.2	0.4	1	2
$\widehat{\mathbb{E}}\hat{h}$	0.200	0.400	0.998	1.996
$\widehat{\mathrm{Var}}(\hat{h})$	$4.732 \cdot 10^{-5}$	$1.813 \cdot 10^{-4}$	$1.260 \cdot 10^{-3}$	$7.352 \cdot 10^{-3}$
$\widetilde{\mathrm{Var}}(\hat{h})$	$4.736 \cdot 10^{-5}$	$1.925 \cdot 10^{-4}$	$1.358 \cdot 10^{-3}$	$7.434 \cdot 10^{-3}$

Acknowledgements

References

[1] Baddeley A and Jensen E B V (2005). Stereology for Statisticians. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC Press.

[2] Barndorff-Nielsen O. and Schmiegel J (2004). Lévy-based tempo-spatial modelling with applications to turbulence. Uspekhi Mat. Nauk. 159:63–90.

[3] Hellmund G, Prokešová M and Jensen E B V (2008). Lévy-based Cox point processes. Adv. Appl. Prob. 40:603–629.

[4] Jónsdóttir K Ý and Jensen E B V (2013). Lévy-based error prediction in circular systematic sampling. Image Anal. Stereol. 32:117-125.

[5] Jónsdóttir K Ý, Rønn-Nielsen A, Mouridsen K and Jensen E B V (2013). Lévy-based modelling in brain imaging. Scand. J. Stat. 40(3):511–529.

[6] Jónsdóttir K Ý, Schmiegel J and Jensen E B V (2008). Lévy-based growth models. Bernoulli 14(1):62–90.

[7] Kendall M G and Stuart A (1977). The Advanced Theory of Statistics, Vol. 1. Distribution Theory, 4th edn. Griffin, London.

[8] Sporring J, Khanmohammadi M, Darkner S, Nava N, Nyengaard J R and Jensen E B V (2014). Estimating the thickness of ultra thin sections for electron microscopy by image statistics. Proceedings of the 2014 IEEE International Symposium on Biomedical Imaging, Beijing, 29 April - 2 May 2014, 157-160.

[9] Ziegel J, Baddeley A, Dorph-Petersen K-A and Jensen E B V (2010). Systematic sampling with errors in sample locations. Biometrika 97:1-13.

[10] Ziegel J, Jensen E B V and Dorph-Petersen K-A (2011). Variance estimation for generalized Cavalieri estimators. Biometrika 98:187-198.

A   Derivation of $\mathrm{Var}(S)$

In this Appendix, we derive the variance formula (3.2) by differentiating (3.6) and (3.7), and equating the two obtained expressions at $\lambda_1 = \lambda_2 = 0$ with each other. Recall that

$$Y_i = X_{v_{1i}} - X_{v_{2i}}, \qquad g_i(u) = f(u + v_{1i}) - f(u + v_{2i}).$$

It is easy to differentiate (3.7) with respect to λ_1 and λ_2. Notice that

$$K_{i,j}(\lambda_1, \lambda_2) = \int_{\mathbb{R}^3} \log \mathbb{E}\big(e^{(\lambda_1 g_i(u) + \lambda_2 g_j(u))Z'}\big)\,\mathrm{d}u$$

$$= \int_{\mathbb{R}^3} K\big(\lambda_1 g_i(u) + \lambda_2 g_j(u)\big)\,\mathrm{d}u\,,$$

where K is the cumulant generating function of Z'. It follows that

$$\frac{\partial^{k+\ell}}{\partial\lambda_1^k \partial\lambda_2^\ell} K_{i,j}(\lambda_1, \lambda_2)\Big|_{\lambda_1=\lambda_2=0} = \kappa_{k+\ell}(Z') \int_{\mathbb{R}^3} g_i(u)^k g_j(u)^\ell\,\mathrm{d}u\,, \qquad \text{(A.1)}$$

where $\kappa_k(Z')$ is the k'th cumulant of Z'.

In order to differentiate (3.6), we use the theory of joint cumulants, see [7, Section 3.14 & Section 3.29]. Since $\mathbb{E}(Y_i) = \mathbb{E}(Y_j) = 0$, we get

$$\frac{\partial^2}{\partial\lambda_1^2} K_{ij}(\lambda_1, \lambda_2)\Big|_{\lambda_1=\lambda_2=0} = \mathbb{E}(Y_i^2)$$

$$\frac{\partial^2}{\partial\lambda_1 \partial\lambda_2} K_{i,j}(\lambda_1, \lambda_2)\Big|_{\lambda_1=\lambda_2=0} = \mathbb{E}(Y_i Y_j)$$

$$\frac{\partial^4}{\partial\lambda_1^2 \partial\lambda_2^2} K_{i,j}(\lambda_1, \lambda_2)\Big|_{\lambda_1=\lambda_2=0} = \mathbb{E}(Y_i^2 Y_j^2) - \mathbb{E}(Y_i^2)\mathbb{E}(Y_j^2) - 2(\mathbb{E}(Y_i Y_j))^2\,.$$

Combining these equations with (A.1), we find

$$\mathbb{E}(Y_i^2) = \kappa_2(Z') \int_{\mathbb{R}^3} g_i(u)^2\,\mathrm{d}u\,,$$

$$\mathbb{E}(Y_i^2 Y_j^2) = \kappa_4(Z') \int_{\mathbb{R}^3} g_i(u)^2 g_j(u)^2\,\mathrm{d}u + \kappa_2(Z')^2\Big[\int_{\mathbb{R}^3} g_i(u)^2\,\mathrm{d}u\Big]^2$$

$$+ 2\kappa_2(Z')^2\Big[\int_{\mathbb{R}^3} g_i(u)g_j(u)\,\mathrm{d}u\Big]^2\,.$$

Inserting this into (3.5), we obtain (3.2).

B Derivations for the Gaussian kernel function

In order to derive (3.9) and (3.10), we use that for $v \in \mathbb{R}^3$

$$\int_{\mathbb{R}^3} f(u)f(u+v) \, \mathrm{d}u = \frac{1}{(4\pi\sigma)^{3/2}} e^{-\frac{1}{4\sigma}\|v\|^2}$$

and

$$\int_{\mathbb{R}^3} f(u)^2 f(u+v)^2 \, \mathrm{d}u = \frac{1}{8(2\pi\sigma)^{9/2}} e^{-\frac{1}{2\sigma}\|v\|^2} .$$

Furthermore, for $v_1, v_2 \in \mathbb{R}^3$ such that $v_1 \perp v_2$, we have

$$\int_{\mathbb{R}^3} f(u)^2 f(u+v_1)f(u+v_1+v_2) \, \mathrm{d}u = \frac{1}{8(2\pi\sigma)^{9/2}} e^{-\frac{1}{2\sigma}\|v_1\|^2 - \frac{3}{8\sigma}\|v_2\|^2}$$

and

$$\int_{\mathbb{R}^3} f(u)f(u+v_1)f(u+v_2)f(u+v_1+v_2) \, \mathrm{d}u = \frac{1}{8(2\pi\sigma)^{9/2}} e^{-\frac{1}{2\sigma}\left(\|v_1\|^2 + \|v_2\|^2\right)} .$$

Using these expressions in (3.3) and (3.4) yield (3.9) and (3.10).