Kristjana Ýr Jónsdóttir and Eva B. Vedel Jensen

Lévy-based modelling in circular systematic sampling

Lévy-based modelling in circular systematic sampling

Kristjana Ýr Jónsdóttir and Eva B. Vedel Jensen

Department of Mathematics
Aarhus University
kyj@imf.au.dk, eva@imf.au.dk

Abstract

In the present paper, we develop Lévy-based error prediction in circular systematic sampling. We use a model-based statistical setting as in Hobolth and Jensen (2002), but relax the assumption that the measurement function is Gaussian. The measurement function is represented as a periodic stationary stochastic process X obtained by a kernel smoothing of a Lévy basis. The process X may have an arbitrary covariance function. The distribution of the error predictor, based on measurements in n systematic directions is derived. Statistical inference is developed for the model parameters in the case where the covariance function follows the celebrated p-order covariance model.

Keywords: Fourier series; Lévy basis; planar particles; saddlepoint approximation; stationary stochastic processes; stereology; systematic sampling

1 Introduction

A long-standing problem in stereology is variance estimation in systematic sampling. One class of problems involves estimation of an integral of the form

$$Q = \int_0^{2\pi} x(\theta)d\theta, \tag{1}$$

where $x(\theta)$ is an integrable function on $[0, 2\pi)$, called the measurement function. The estimator typically considered is based on circular systematic sampling and takes the form

$$\widehat{Q}_n = \frac{2\pi}{n} \sum_{i=0}^{n-1} x\left(\Theta + \frac{2\pi i}{n}\right), \qquad (n \geq 1),$$

where Θ is uniformly distributed in $[0, \frac{2\pi}{n})$. For instance, if Y is a bounded convex planar set containing the origin O, examples of (1) are

$$x(\theta) = \begin{cases} \frac{1}{2}r(\theta)^2 & \text{if } Q = \text{area of } Y, \\ h(\theta) & \text{if } Q = \text{boundary length of } Y, \end{cases}$$

1

where $r(\theta)$ and $h(\theta)$ are the radial function and the support function in direction θ, respectively. The geometric identity (1) is in these cases a consequence of polar decomposition in the plane and an identity for mean width, cf. e.g. Schneider (1993, (5.3.12)). If instead Y is a bounded convex spatial set containing O, the volume and the surface area of Y may be estimated by a two-step procedure which involves circular systematic sampling in a section through O and the use of the cubed radial function or the squared support function, cf. Gundersen (1988) and Cruz-Orive (2005). Yet another example is volume estimation by the so-called vertical rotator, cf. Jensen and Gundersen (1993).

In Gual-Arnau and Cruz-Orive (2000), a design-based procedure of approximating the variance of $\widehat{Q}_n$, based on modelling the covariogram of $x(\theta)$ by a polynomial model, is developed. Hobolth and Jensen (2002) consider a model-based procedure, where the measurement function is assumed to be a realization of a periodic stationary stochastic Gaussian process $X = \{X(\theta) : \theta \in [0, 2\pi]\}$. It is shown in Hobolth and Jensen (2002) that the covariogram model considered in the paper by Gual-Arnau and Cruz-Orive (2000) is a special case of a p-order covariance model for the stochastic process X in the model-based set-up.

The p-order covariance model is given by

$$c(\theta) = \mathrm{Cov}(X(\theta), X(0)) = \lambda_0 + \sum_{s=2}^{\infty} \lambda_s \cos(s\theta), \quad \lambda_s^{-1} = \alpha + \beta(s^{2p} - 2^{2p}), \quad s \geq 2, \ (2)$$

where the model parameters satisfy $p > 1/2$, $\alpha, \beta > 0$. In Hobolth et al. (2003), this parametric covariance function has been used in the modelling of the radial function of a random star-shaped planar particle. In this case, p determines the smoothness of the particle boundary while α and β determine the 'global' and the 'local' shape of the particle, respectively. (Note that in (2) λ_1 is set to zero which ensures that the reference point of the particle is approximately the centre of mass.) The model-based counterpart of the design-based methodology provided in Gual-Arnau and Cruz-Orive (2000) was further developed in Jónsdóttir et al. (2006), where the general form of the p-order covariance model (2) was used to obtain a more accurate approximation of the prediction error $\mathbb{E}(\widehat{Q}_n - Q)^2$.

In Hobolth and Jensen (2002) and Jónsdóttir et al. (2006), the process X is assumed to be Gaussian. Motivated by the fact that powers of the radial function and the support function are used in practice, we will in this paper consider non-Gaussian models, obtained as a kernel smoothing of a so-called Lévy basis. As we will show, it is possible under the Lévy-based model to derive the distribution of the error predictor $\widehat{Q}_n - Q$ which may be markedly non-Gaussian for the moderate sizes of n used in practice.

Lévy-based modelling has been popular in recent years, e.g. in the modelling of turbulent flows, spatio-temporal growth, spatial point processes and random fields, cf. Barndorff-Nielsen and Schmiegel (2004), Jónsdóttir et al. (2008), Hellmund et al. (2008) and Jónsdóttir et al. (2011). More specifically, we will consider stochastic processes on the form

$$X(\theta) = \mu + \int_0^{2\pi} k(\theta - \phi) Z(d\phi), \qquad \theta \in [0, 2\pi),$$

where Z is a homogeneous and factorizable Lévy basis on $[0, 2\pi)$ and k is a deterministic kernel function. In principle, any covariance model, including the p-order covariance model, can be induced under this modelling framework, by assuming a specific form of the kernel function, see Section 2 below. Under the p-order model, it is easy to control the local and global fluctuations of the stochastic process X. The Lévy-based models with p-order covariance thus constitute a flexible and tractable model class. In particular, this model class has more structure than the non-Gaussian models considered in Hobolth et al. (2003) and this allows us to derive distributional results.

Section 2 gives a theoretical background for stationary periodic processes with period 2π based on kernel smoothing of a Lévy basis. In Section 3, estimation of $\mathbb{E}(\widehat{Q}_n - Q)^2$ under the general Lévy-based model is discussed. In Section 4, we derive the distribution of the error predictor $\widehat{Q}_n - Q$ under the Lévy-based model, and show how to estimate this distribution. This section also provides an example of random particles simulated from a Lévy-based model and the distribution of the n-point area estimator of these particles. Finally, a discussion is provided in Section 5. Some technical derivations are deferred to an appendix.

2 Lévy-based stochastic processes on the circle

This section provides an overview of stationary periodic processes on $[0, 2\pi)$ based on integration with respect to a Lévy basis. For further details on the general theory on Lévy bases, in particular, the integration with respect to a Lévy basis, the reader is referred to Barndorff-Nielsen and Schmiegel (2004) and Hellmund et al. (2008).

Let $X = \{X(\theta) : \theta \in [0, 2\pi)\}$ be a 2π periodic stationary stochastic process on $[0, 2\pi)$, given by

$$X(\theta) = \mu + \int_0^{2\pi} k(\theta - \phi) Z(d\phi), \qquad \theta \in [0, 2\pi), \tag{3}$$

where Z is a homogeneous and factorizable Lévy basis on $[0, 2\pi)$ and k is a periodic kernel function with period 2π. Furthermore, we will assume that k is an even function on $[0, 2\pi)$ with a Fourier representation

$$k(\theta) = \xi_0 + \sum_{s=1}^{\infty} \xi_s \cos(s\theta). \tag{4}$$

A spatio-temporal version of (3) has previously been considered in Jónsdóttir et al. (2008).

The Lévy basis Z is extended by $Z(A + 2\pi m) = Z(A)$ for all $m \in \mathbb{Z}$ and all Borel sets $A \in \mathcal{B}([0, 2\pi))$. A Lévy basis has the property that $Z(A_1), \ldots, Z(A_n)$ are independent when $A_1, \ldots, A_n \in \mathcal{B}([0, 2\pi))$ are disjoint and $Z(A)$ is infinitely divisible for any $A \in \mathcal{B}([0, 2\pi))$. The assumption of homogeneity implies that all the finite-dimensional distributions of Z are translation invariant.

If Z is Gaussian, the integral (3) exists if k is L^2-integrable with respect to the Lebesgue measure on $[0, 2\pi)$. When Z is a so-called Lévy jump basis (e.g. Gamma or

inverse Gaussian Lévy basis), the integral exists if k is integrable with respect to the Lebesgue measure on $[0, 2\pi)$ and if $\int_{\mathbb{R}} |r| V(dr) < \infty$, where V is the Lévy measure associated with Z. These results follow from Hellmund et al. (2008, Lemma 1).

An important entity associated with a Lévy basis is its spot variable Z' which is infinitely divisible. Without loss of generality we will in what follows assume that the spot variable Z' is centered, $Z' = W - \mathbb{E}(W)$, where W is an infinitely divisible random variable.

The following theorem characterizes the distribution of the stochastic variable $X(\theta)$.

Theorem 1. *The stochastic variable $X(\theta)$ has the following cumulant (generating) function*

$$K_X(t) = \log \mathbb{E}(e^{tX(\theta)}) = t(\mu - 2\pi\xi_0 \mathbb{E}(W)) + \int_0^{2\pi} K_W(tk(\phi))d\phi, \qquad (5)$$

where $K_W(t)$ is the cumulant function of W. Moreover, the derivatives of $K_X(t)$ are given by (when they exist)

$$\begin{aligned}
K_X'(t) &= \mu - 2\pi\xi_0 \mathbb{E}(W) + \int_0^{2\pi} K_W'(tk(\phi))k(\phi)d\phi, \\
K_X^{(r)}(t) &= \int_0^{2\pi} K_W^{(r)}(tk(\phi))k(\phi)^r d\phi, \qquad r \geq 2,
\end{aligned} \qquad (6)$$

where $K_W^{(r)}(t)$ denotes the r'th derivative of $K_W(t)$.

Proof. The result is obtained by using that the cumulant function of the integral $f \bullet Z = \int_0^{2\pi} f(\theta)Z(d\theta)$ of a function f with respect to a homogeneous and factorizable Lévy basis Z is given by

$$K_{f \bullet Z}(t) = \int_0^{2\pi} K_{Z'}(tf(\theta))d\theta, \qquad (7)$$

where Z' is the spot variable associated with Z, cf. Hellmund et al. (2008, (10)). $\square$

It follows that the cumulants of the stochastic variable $X(\theta)$ are given by

$$\kappa_1(X(\theta)) = \mu, \quad \kappa_r(X(\theta)) = \kappa_r(W) \int_0^{2\pi} k(\phi)^r d\phi, \qquad r \geq 2.$$

Possible choices of the distribution of W are the Gaussian, Gamma and inverse Gaussian distributions. When the kernel function is proportional to an indicator function, $k(\theta) = c\mathbf{1}_A(\theta)$ for $A \in \mathcal{B}([0, 2\pi))$, the marginal distribution of $X(\theta)$ will be of the same type as that of W. Otherwise, the marginal distribution will not be as simple, but the process X will inherent the name of the distribution of the underlying spot variable, e.g. when W is Gamma distributed, X is called a Gamma Lévy process, irrespectively of the choice of the kernel function. We will typically assume that $\kappa_2(W) = 1$, i.e. the skewness and kurtosis of W are equal to the third and fourth cumulant, respectively. In Table 1, we give the cumulant function, third and fourth cumulants of W for the three distributions mentioned above. Note that

4

Table 1: Examples of the distribution of W, together with the corresponding cumulant function, and the third and fourth cumulants.

distribution	Gaussian	Gamma	Inverse Gaussian
W	$N(0,1)$	$\Gamma(\eta, \sqrt{\eta})$	$IG(\eta^3, \eta)$
$K_W(t)$	$-t^2/2$	$-\eta \log(1 - t/\sqrt{\eta})$	$-\eta^4(1 - \sqrt{1 - 2t/\eta^2})$
$\kappa_3(W)$	0	$2/\sqrt{\eta}$	$3/\eta^2$
$\kappa_4(W)$	0	$6/\eta$	$15/\eta^4$

as W has unit variance, the Gamma and inverse Gaussian Lévy bases are only determined by a single parameter $\eta > 0$. The Lévy measures V of the Gamma and inverse Gaussian Lévy bases satisfy the condition $\int_{\mathbb{R}} |r| V(dr) < \infty$ for the existence of (3), cf. Hellmund et al. (2008, Example 3).

In principle, any covariance function can be modelled within this set-up. This can be seen, using the theorem below.

Theorem 2. *The stochastic process X has a mean value μ and a covariance function*

$$c(\theta) = \lambda_0 + \sum_{s=1}^{\infty} \lambda_s \cos(s\theta), \qquad \theta \in [0, 2\pi).$$

where

$$\lambda_0 = 2\pi \xi_0^2 \kappa_2(W), \quad \lambda_s = \pi \xi_s^2 \kappa_2(W), \qquad s \geq 1. \tag{8}$$

Proof. Using that

$$c(\theta) = \mathrm{Cov}(X(\theta), X(0)) = \kappa_2(W) \int_0^{2\pi} k(\theta - \phi) k(-\phi) d\phi,$$

we easily obtain equation (8). $\qquad \square$

Using Theorem 2, we can construct the candidate kernel k that induces a given covariance function c. For instance, if c follows the p-order model (2), then k is of the form (4) with

$$\xi_0 = \sqrt{\frac{\lambda_0}{2\pi\kappa_2(W)}}, \qquad \qquad \xi_1 = 0,$$

$$\xi_s = \frac{1}{\sqrt{\pi\kappa_2(W)[\alpha + \beta(s^{2p} - 2^{2p})]}}, \qquad s \geq 2.$$

This choice of kernel will for a Gaussian Lévy basis give a well-defined integral (3) if $p > 1/2$, while for a Gamma or an inverse Gaussian basis the integral is well-defined if $p > 1$.

3   Estimating $\mathbb{E}(\widehat{Q}_n - Q)^2$ under the Lévy-based model

In Hobolth and Jensen (2002), the focus was on the prediction error $\mathbb{E}(\widehat{Q}_n - Q)^2$. If the covariance function of X has the following Fourier expansion

$$c(\theta) = \lambda_0 + \sum_{s=1}^{\infty} \lambda_s \cos(s\theta), \qquad \theta \in [0, 2\pi),$$

then it was shown in Hobolth and Jensen (2002) that

$$\mathbb{E}(\widehat{Q}_n - Q)^2 = \sum_{k=1}^{\infty} \lambda_{nk}. \tag{9}$$

Note that in Hobolth and Jensen (2002), circular systematic sampling on $[0, 1)$ instead of $[0, 2\pi)$ is considered, so (9) represents an adjusted version of Hobolth and Jensen (2002, (8)).

In Hobolth et al. (2003), a procedure for estimating the prediction error under a Gaussian p-order model was developed, based on a Fourier expansion of X

$$X(\theta) \sim A_0 + \sum_{s=1}^{\infty} \left(A_s \cos(s\theta) + B_s \sin(s\theta)\right), \qquad \theta \in [0, 2\pi),$$

where

$$A_0 = \frac{1}{2\pi} \int_0^{2\pi} X(\theta) d\theta,$$

$$A_s = \frac{1}{\pi} \int_0^{2\pi} X(\theta) \cos(s\theta) d\theta,$$

$$B_s = \frac{1}{\pi} \int_0^{2\pi} X(\theta) \sin(s\theta) d\theta.$$

When X is a periodic stationary Gaussian process, the Fourier coefficients of X become independent and normally distributed, $A_s \sim B_s \sim N(0, \lambda_s)$. As suggested in Hobolth et al. (2003), the parameters α and β in the p-order model can then be estimated using maximum likelihood estimation based on the first S Fourier coefficients,

$$L_{0,S}(\alpha, \beta) = \prod_{s=2}^{S} \frac{1}{2\pi \lambda_s(\alpha, \beta)} \exp\left(-\frac{(a_s^2 + b_s^2)}{2\lambda_s(\alpha, \beta)}\right), \tag{10}$$

where $\lambda_s(\alpha, \beta)$, $s = 2, \ldots, S$, satisfy (2) and a_s and b_s, $s = 2, \ldots, S$, denote discretized Fourier coefficients of X.

In this section, we will show that this procedure can be used under the general Lévy-based model. The following theorem gives the distribution of the Fourier coefficients and their relations under the general Lévy-based model. In the Gaussian case, the theorem can be found e.g. in Dufour and Roy (1976).

Theorem 3. *The stationary Lévy-based stochastic process X can be written in terms of its Fourier coefficients as*

$$X(\theta) = A_0 + \sum_{s=1}^{\infty} \left(A_s \cos(s\theta) + B_s \sin(s\theta) \right), \qquad \theta \in [0, 2\pi),$$

where $A_0 = \mu + \xi_0 Z([0, 2\pi))$,

$$A_s = \xi_s \int_0^{2\pi} \cos(s\phi) Z(d\phi) \quad and \quad B_s = \xi_s \int_0^{2\pi} \sin(s\phi) Z(d\phi). \qquad (11)$$

Moreover, the Fourier coefficients are pairwise uncorrelated and the Fourier coefficients of order s have the same distribution which is characterized by the cumulant function $K_{A_s}(t) = K_{B_s}(t) = K_U(t\xi_s)$, where

$$K_U(t) = \int_0^{2\pi} K_W(t \cos(\theta)) d\theta.$$

Proof. Writing the kernel function in terms of its Fourier representation and then calculate the Fourier coefficients of X gives (11). The Fourier coefficients are uncorrelated as

$$\text{Cov}(A_s, B_r) = \kappa_2(W) \int_0^{2\pi} \cos(s\phi) \sin(r\phi) d\phi = 0, \qquad \text{for all } r, s \geq 1$$

$$\text{Cov}(A_s, A_r) = \kappa_2(W) \int_0^{2\pi} \cos(s\phi) \cos(r\phi) d\phi = 0, \qquad \text{for all } r, s \geq 1, r \neq s$$

$$\text{Cov}(B_s, B_r) = \kappa_2(W) \int_0^{2\pi} \sin(s\phi) \sin(r\phi) d\phi = 0, \qquad \text{for all } r, s \geq 1, r \neq s.$$

The cumulant function of A_s is given by

$$K_{A_s}(t) = \int_0^{2\pi} K_{Z'}(t\xi_s \cos(s\phi)) d\phi = \frac{1}{s} \int_0^{2\pi s} K_{Z'}(t\xi_s \cos(\phi)) d\phi = K_U(\xi_s t),$$

and a similar argument shows that $K_{B_s}(t) = K_U(\xi_s t)$. $\qquad \square$

The cumulant function of A_s and B_s yield simple expressions for their cumulants, which are given by

$$\kappa_1(A_0) = \mu, \quad \kappa_r(A_0) = 2\pi\xi_0\kappa_r(W), \qquad\qquad r \geq 2,$$

$$\kappa_r(A_s) = \kappa_r(B_s) = 2\pi\xi_s^r \frac{(r-1)!!}{r!!}\kappa_r(W)\,\mathbf{1}(r \text{ even}), \qquad s \geq 1, \ r \geq 1,$$

where $\kappa_r(W)$ denotes the r-th cumulant of the stochastic variable W. This means that A_s and B_s have mean, variance, skewness and kurtosis of the following form

$$\kappa_1(A_s) = 0, \quad \kappa_2(A_s) = \pi\xi_s^2\kappa_2(W), \quad \gamma_1(A_s) = 0, \quad \gamma_2(A_s) = \frac{3}{4\pi}\gamma_2(W),$$

where $\gamma_2(W)$ is the kurtosis of W. Moreover, the normalized Fourier coefficients of order $s = 1, 2, \ldots$, obtained by $\tilde{A}_s = A_s/\xi_s$ and $\tilde{B}_s = B_s/\xi_s$, will all have the same distribution characterized by the cumulant function $K_U(t)$.

From Theorems 2 and 3, it follows that for a non-Gaussian Lévy basis Z the Fourier coefficients will be uncorrelated with variance $\lambda_s(\alpha, \beta)$. Furthermore, the distribution of the Fourier coefficients in the non-Gaussian and Gaussian model only differs in even cumulants of order four and higher. Therefore, (10) can be regarded as a pseudo-likelihood function for (α, β) also in the non-Gaussian case.

Figure 1 shows the small difference in the saddlepoint densities of the normalized Fourier coefficients and the Gaussian density for different values of η in the case of a Gamma Lévy basis. Furthermore, a simulation study indicated that the estimates of (α, β) are robust against deviations from Gaussianity in the underlying distribution, but the mean square error of the estimates increases somewhat as η decreases.
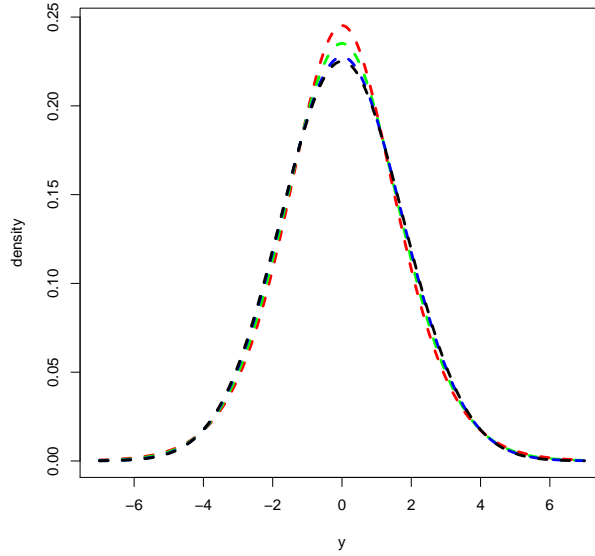

Figure 1: The saddlepoint density of the normalized Fourier coefficients for $\eta = 2$ (red line), $\eta = 4$ (green line) and $\eta = 16$ (blue line). The density in a Gaussian model is shown for comparison (black line).

4 The distribution of $\widehat{Q}_n - Q$ under the Lévy-based model

In the previous section, we have seen that the method developed in Hobolth et al. (2003) for estimating $\mathbb{E}(\widehat{Q}_n - Q)^2$ based on a Gaussian process is robust against departures from the distributional assumption. In this section, we will derive the distribution of the error predictor $\widehat{Q}_n - Q$ which may be markedly non-Gaussian.

Theorem 4. *Under the Lévy-based model, the error predictor is distributed as*

$$\widehat{Q}_n - Q \sim 2\pi \int_0^{2\pi} k_n(\phi) Z(d\phi), \tag{12}$$

where

$$k_n(\phi) = \sum_{s=1}^{\infty} \xi_{sn} \cos(sn\phi).$$

8

The distribution of $\widehat{Q}_n - Q$ is characterized by its cumulant function

$$K_{\widehat{Q}_n - Q}(t) = \int_0^{2\pi} K_W(2t\pi k_n(\phi))d\phi.$$

Proof. Recall that

$$\widehat{Q}_n = \frac{2\pi}{n} \sum_{i=0}^{n-1} X\left(\Theta + \frac{2\pi i}{n}\right), \qquad (n \geq 1),$$

where Θ is uniformly distributed in $[0, \frac{2\pi}{n})$. Without loss of generality we can assume that $\Theta = 0$ as the distribution of $\widehat{Q}_n$ does not depend on Θ. The result (12) is obtained by observing that the mean of the kernel functions is given by

$$\frac{1}{n} \sum_{j=0}^{n-1} k\left(\frac{2\pi j}{n} - \phi\right) = \xi_0 + \sum_{s=1}^{\infty} \xi_{sn} \cos(sn\phi).$$

The expression for the cumulant function of $\widehat{Q}_n - Q$ is a consequence of (7). $\qquad \square$

As the cumulant function of $\widehat{Q}_n - Q$ has a simple form, its cumulants are easily available. In particular, it enables us to obtain a saddlepoint approximation of its density. An alternative is to use Theorem 4 for simulating the distribution of $\widehat{Q}_n - Q$.

Example 5. Let us consider a Lévy-based model (3) for X with a Gamma Lévy basis Z and k chosen such that the covariance function of X follows a p-order model. Under this model the Fourier coefficients A_s, and B_s, $s \geq 1$, of $X(\theta)$ have mean, variance, skewness and kurtosis,

$$\kappa_1(A_s) = 0, \quad \kappa_2(A_s) = \pi\xi_s^2 = \lambda_s(\alpha, \beta), \quad \gamma_1(A_s) = 0, \quad \gamma_2(A_s) = \frac{9}{2\pi\eta}.$$

This model may, for instance, be used to model the squared radial function of random star-shaped planar particles containing the origin. Figure 2 shows examples of particles simulated from such a model, using different values of η. The value of p, α and β was $p = 2$, $\log \alpha = 6$ and $\log \beta = -3$. We used $\eta = 2, 4, 16$; a low value of η corresponds to an underlying distribution with a heavy tail. The value of η controls the frequency and size of the irregularities of the boundary of the particles. Small values of η will produce particles with few large fluctuations on the particle boundary and less smaller fluctuations. Higher values of η will produce particles with more frequently occurring moderate fluctuations across the boundary.

Figure 3 shows the corresponding saddlepoint densities of the estimated area of the particles for two different values of n. $\qquad \square$

In applications, it is needed to estimate the parameter η of the underlying Lévy basis Z, i.e. the parameter of the distribution of W. For a given kernel function k, we have a simple expression for the cumulant function of X and its derivatives (when they exist), cf. Theorem 1. We suggest estimating the parameter η determining the Lévy basis by considering the saddlepoint approximation of the density of $X(\theta)$.

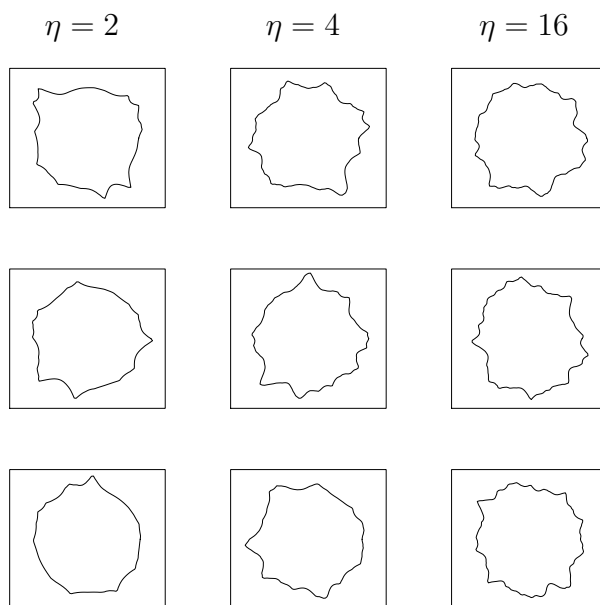$\eta = 2$ $\eta = 4$ $\eta = 16$

Figure 2: Realizations of particles obtained by assuming that the squared radial function is given by a Gamma Lévy process with a p-order covariance function. The values of p, α and β were $p = 2$, $\log \alpha = 6$ and $\log \beta = -3$. Each column corresponds to realizations for a fixed value of η ($\eta = 2, 4, 16$).
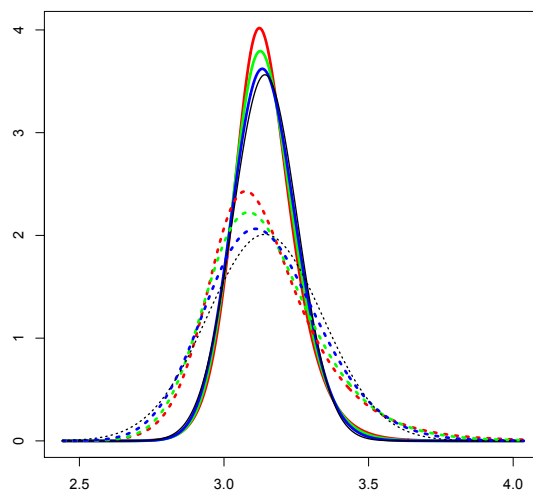

Figure 3: The saddlepoint density of the n-point area estimate for $n = 5$ (stippled) and $n = 10$ (full line). The different colours represent densities for the three particles considered: $\eta = 2$ (red lines), $\eta = 4$ (green lines) and $\eta = 16$ (blue lines). Densities for a Gaussian model is shown for comparison (black lines).

10

For more details on saddlepoint approximations, cf. Jensen (1995). The first order saddlepoint approximation of the density is given by

$$\tilde{f}(x) = \frac{1}{\sqrt{2\pi K_X''(\hat{t}(x))}} e^{K_X(\hat{t}(x)) - \hat{t}(x)x},$$

where $\hat{t}(x)$ is the solution to the saddlepoint equation

$$K_X'(t) = x.$$

Note that the saddlepoint equation is non-linear when Z is non-Gaussian, but can be solved numerically, using a Newton method for a given kernel function k. A better approximation of the density of $X(\theta)$ is obtained by multiplying the density with the correction factor

$$c(\hat{t}(x)) = 1 + \frac{K_X^{(4)}(\hat{t}(x))}{8K_X''(\hat{t}(x))^2} - \frac{5}{24}\left(\frac{K_X^{(3)}(\hat{t}(x))}{K_X''(\hat{t}(x))^{3/2}}\right)^2.$$

Given an estimation of the kernel function k we can establish a pseudo-likelihood function based on the observations of the stochastic process X given by

$$L(\eta) = \prod_{i=1}^{n} \tilde{f}(x(\theta_i)).$$

Here, $\tilde{f}(x(\theta_i))$ is calculated using the approximated kernel function

$$\hat{k}_S(\theta) = \sum_{s=2}^{S} \hat{\xi}_s \cos(s\theta),$$

where $\hat{\xi}_s = \sqrt{\lambda_s(\hat{\alpha}, \hat{\beta})/\pi}$ is obtained using the estimates of (α, β). Note that we need to normalize the densities $\tilde{f}(x_i)$ for each value of η, when maximizing the likelihood function $L(\eta)$. If the estimated likelihood function is an increasing function of η, this suggests that the underlying Lévy basis is Gaussian.

The cumulant function and its derivatives have a simple analytic expression, when the underlying Lévy basis is a Gamma basis or Inverse Gaussian basis. For a Gamma Lévy process the cumulant function of $X(\theta)$ and its derivatives are given by

$$K_X(t) = -\eta \int_0^{2\pi} \log\left(1 - \frac{tk(\phi)}{\sqrt{\eta}}\right) d\phi,$$

$$K_X^{(r)}(t) = (r-1)!\eta \int_0^{2\pi} \frac{k(\phi)^r}{(\sqrt{\eta} - tk(\phi))^r} d\phi.$$

For an inverse Gaussian Lévy process the cumulant function of $X(\theta)$ and its derivatives are given by

$$K_X(t) = -\eta^4 \int_0^{2\pi}\left(1 - \sqrt{1 - \frac{2tk(\phi)}{\eta^2}}\right) d\phi,$$

$$K_X^{(r)}(t) = (2r-3)!!\eta^3 \int_0^{2\pi} \frac{k(\phi)^r}{(\eta^2 - 2tk(\phi))^{r-\frac{1}{2}}} d\phi.$$

In both cases, the saddlepoint approximation of the density of $X(\theta)$ are easily obtainable using numerical integration and a Newton algorithm for finding the saddlepoint. Note that the saddlepoint approximation of the density function for an arbitrary η can be written in terms of the saddlepoint approximation of the density, the cumulant derivatives and saddlepoint solution for $\eta = 1$.

5 Discussion

We have developed a Lévy-based error prediction in circular systematic sampling. In contrast to previous model-based methods, we consider a flexible class of non-Gaussian measurement functions based on kernel smoothing of a homogeneous Lévy basis. In particular, we have derived the distribution of the error predictor in circular systematic sampling. The modelling framework allows us to consider in principle any given covariance structure of the measurement function, in particular the popular p-order covariance model which enables controlling the local and global fluctuations of the measurement function.

Note that as the Lévy-based process X is strictly stationary, it can be shown that X has a polar expansion of the form

$$X(\theta) = \mu + \sqrt{2} \sum_{s=0}^{\infty} \sqrt{C_s} \cos(s(\theta - D_s)),$$

where the random variables

$$C_s = \tfrac{1}{2}(A_s^2 + B_s^2) = \lambda_s Z_s, \quad D_s \sim U[0, \tfrac{2\pi}{s}], \qquad s \geq 1,$$

are independent, cf. the Appendix, and $\mathbb{E}(Z_s) = 1$. Here, the variable Z_s can be expressed as

$$Z_s = \frac{1}{2\pi\kappa_2(W)} \int_0^{2\pi} \int_0^{2\pi} \cos(s(\theta - \phi)) Z(d\theta) Z(d\phi).$$

This shows that the Lévy-based models are closely related to the generalized p-order models proposed in Hobolth et al. (2003), but the Lévy-based models have more structure that allows for derivation of distributional results.

The model (3) can be used directly to model the shape of featureless two-dimensional particles by assuming that a particle Y is a stochastic deformation of a template particle Y_0. If r_0 is a radial function of a template particle, we let the radial function of Y be of the form $R(\theta) = r_0(\theta) + X(\theta)$, where X is a zero mean Lévy-based stochastic process. The strength of this technique is two-folded. Firstly, the global and local fluctuations of the Lévy-based stochastic process are controlled by the variance of the Fourier coefficients which are determined by the kernel function. Secondly, the underlying Lévy basis determines the frequency and size of the irregularities of the process. Finally, the methodology presented can be extended to model the shape of three-dimensional featureless particles, by considering Lévy-based processes on the unit sphere $\mathbb{S}_2$. Hansen et al. (2011) consider three-dimensional Lévy particles using different covariance models. The focus is here on

the Hausdorff dimension of the boundary of particles obtained using a Gaussian basis.

The saddlepoint approximation was applied here to obtain an approximation of the density of $X(\theta)$ in the Lévy-based stochastic model. As the cumulant function is easily obtainable for stochastic convolutions of the type

$$X(\xi) = \int f(\eta - \xi) Z(d\eta), \qquad \xi \in \mathbb{R}^n,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and Z is a homogeneous Lévy basis on $\mathbb{R}^n$, the saddlepoint approximation of the density is an attractive tool for studying Lévy-based convolution models in general. For the stochastic processes considered here, other types of approximations of densities can also be considered based on approximating A_s and B_s by differences of variables from the same family of distributions as W. As an example one could consider approximating the density of the Fourier coefficients by a type II McKay distribution (Holm and Alouini, 2004), when the underlying Lévy basis is a Gamma Lévy basis.

Finally, it should be emphasized that assuming that the kernel function is even does not affect the flexibility of the induced covariance model. When k is not necessarily even,

$$k(\theta) = \xi_0 + \sum_{s=1}^{\infty} \left(\xi_{s,1} \cos(s\theta) + \xi_{s,2} \sin(s\theta) \right),$$

the covariance function of X is given by

$$C(\theta) = 2\pi\xi_0^2 + \pi \sum_{s=1}^{\infty} (\xi_{s,1}^2 + \xi_{s,2}^2) \cos(s\theta), \qquad \theta \in [0, 2\pi).$$

Moreover, the Fourier coefficients of the process X will still be uncorrelated and have the same distribution described by the cumulant function

$$K_{A_s}(t) = \int_0^{2\pi} K_W(t\xi_{s,1} \cos\phi - t\xi_{s,2} \sin\phi) d\phi,$$

and cumulants given by

$$\kappa_r(A_s) = \kappa_r(B_s) = 2\pi(\xi_{s,1}^2 + \xi_{s,2}^2)^{r/2} \frac{(r-1)!!}{r!!} \kappa_r(W) \, \mathbf{1}(r \text{ even}).$$

Appendix

Consider a stationary process $X = \{X(\theta) : \theta \in [0, 2\pi)\}$ with Fourier expansion

$$X(\theta) = A_0 + \sum_{s=1}^{\infty} (A_s \cos(s\theta) + B_s \sin(s\theta)),$$

and Polar expansion

$$X(\theta) = \mu + \sqrt{2} \sum_{s=0}^{\infty} \sqrt{C_s} \cos(s(\theta - D_s)),$$

where $C_s = \frac{1}{2}(A_s^2 + B_s^2)$ and $sD_s = \arctan(B_s/A_s)$, $s \geq 1$. It is easily seen that

$$X(\theta + h) = A_0 + \sum_{s=1}^{\infty}(A_s(h)\cos(s\theta) + B_s(h)\sin(s\theta)),$$

where

$$\begin{pmatrix} A_s(h) \\ B_s(h) \end{pmatrix} = \begin{pmatrix} \cos(sh) & \sin(sh) \\ -\sin(sh) & \cos(sh) \end{pmatrix} \begin{pmatrix} A_s \\ B_s \end{pmatrix} = V \begin{pmatrix} A_s \\ B_s \end{pmatrix},$$

and V is a rotation matrix. As for each $h \in [0, 2\pi)$, $\{X(\theta) : \theta \in [0, 2\pi)\}$ and $\{X(\theta + h) : \theta \in [0, 2\pi)\}$ have the same distribution,

$$\begin{pmatrix} A_s \\ B_s \end{pmatrix} \sim \begin{pmatrix} A_s(h) \\ B_s(h) \end{pmatrix} = V \begin{pmatrix} A_s \\ B_s \end{pmatrix}.$$

We now have that

$$E_s := \begin{pmatrix} A_s/\sqrt{A_s^2 + B_s^2} \\ B_s/\sqrt{A_s^2 + B_s^2} \end{pmatrix} \sim V \begin{pmatrix} A_s/\sqrt{A_s^2 + B_s^2} \\ B_s/\sqrt{A_s^2 + B_s^2} \end{pmatrix} = V E_s,$$

and consequently E_s is uniformly distributed on the unit circle, i.e. $\arctan(B_s/A_s)$ is uniformly distributed on $[0, 2\pi)$ and D_s is uniformly distributed on $[0, 2\pi/s)$. Now consider the conditional distribution of $\sqrt{2C_s} = \sqrt{A_s^2 + B_s^2}$ given E_s. As

$$\sqrt{2C_s} \mid E_s = e_s \sim \sqrt{2C_s} \mid V E_s = e_s,$$

the conditional distribution does not depend on e_s and hence $\sqrt{2C_s}$ is independent of E_s and D_s.

Acknowledgements

References

O. E. Barndorff-Nielsen and J. Schmiegel. Lévy based tempo-spatial modeling; with applications to turbulence. *Uspekhi Mat. Nauk*, 159:63–90, 2004.

L. M. Cruz-Orive. A new stereological principle for test lines in three-dimensional space. *J. Microsc.*, 219:18–28, 2005.

J.-M. Dufour and R. Roy. On spectral estimation for a homogeneous random process on the circle. *Stoch. Proc. Appl.*, 4:107–120, 1976.

X. Gual-Arnau and L. M. Cruz-Orive. Systematic sampling on the circle and the sphere. *Adv. Appl. Prob. (SGSA)*, 32:628–647, 2000.

H. J. G. Gundersen. The nucleator. *J. Microsc.*, 151:3–21, 1988.

L. V. Hansen, T. L. Thorarinsdottir, and T. Gneiting. LŐvy particles: Modelling and simulating star-shaped random sets. Research report 4, CSGB, Department of Mathematical Sciences, Aarhus University. Submitted., 2011.

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

A. Hobolth and E. B. V. Jensen. A note on design-based versus model-based variance estimation in stereology. *Adv. Appl. Prob. (SGSA)*, 34:484–490, 2002.

A. Hobolth, J. Pedersen, and E. B. V. Jensen. A continuous parametric shape model. *Ann. Inst. Statist. Math.*, 55:227–242, 2003.

H. Holm and M.-S. Alouini. Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution. *IEEE Trans. Commun.*, 52: 1367–1376, 2004.

E. B. V. Jensen and H. J. G. Gundersen. The rotator. *J. Microsc.*, 170:35–44, 1993.

J. L. Jensen. *Saddlepoint Approximations*. Clarendon Press, Oxford, 1995.

K. Y. Jónsdóttir, L. M. Hoffmann, A. Hobolth, and E. B. V. Jensen. On variance estimation in circular systematic sampling. *J. Microsc.*, 222:212–216, 2006.

K. Y. Jónsdóttir, J. Schmiegel, and E. B. V. Jensen. Lévy-based growth models. *Bernoulli*, 14:62–90, 2008.

K. Y. Jónsdóttir, A. Rønn-Nielsen, K. Mouridsen, and E. B. V. Jensen. Lévy based modelling in brain imaging. Research report 2, CSGB, Department of Mathematical Sciences, Aarhus University. Submitted., 2011.

R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge University Press, Cambridge, 1993.