Planar segment processes with reference mark distributions, modeling and estimation

Viktor Beneš, Jakub Večeřa, Milan Pultar

Department of Probability and Mathematical Statistics
Faculty of Mathematics and Physics, Charles University

19th Workshop
on Stochastic Geometry, Stereology and Image Analysis,
Luminy, May 15-19, 2017

Outline of the talk

Mark point processes with density w.r.t. Poisson process, having a reference mark distribution

History of the problem of reference density estimation

Theoretical solution

Example 1: Segment process with a reference directional distribution

Model formulation as a particle Gibbs process

Example 2: Segment process with a reference length distribution

Simulation study, numerical results

The problem

$B \subset \mathbb{R}^d$ bounded measurable set, $(M, \mathcal{M})$ a measurable mark space

$$(Y, \mathcal{Y}) = (B \times M, \mathcal{B}(B) \otimes \mathcal{M})$$

Poisson process η on Y with intensity measure λ,

$$\lambda(\mathrm{d}(y, r)) = \tau \mathrm{d}y g(r) \mathrm{d}r,$$

y location, r mark, $\tau > 0$ parameter

g density of reference mark distribution (unobserved)

X random process with density p w.r.t. η, space of outcomes $(N, \mathcal{N})$

hereditary Papangelou conditional intensity $\lambda^*(\mathbf{x}, u) = \frac{p(\mathbf{x} \cup u)}{p(\mathbf{x})}$, $\mathbf{x} \in N$

The aim: nonparametric estimation of g from observed data

History of the problem

Møller and Helisová (2010), Scand. J. Statist.

disc process in $\mathbb{R}^2$, mark is the radius of disc

exponential type density w.r.t. Boolean model, heather data (Diggle, 1981)

pragmatic approach: using fitted Boolean disc model as the reference process, followed by MLE of the density parameters

Dereudre, Lavancier, Staňková Helisová (2014)

the same data and reference processes

model estimation: Takacz-Fiksel method instead of MLE

Beneš, Večeřa, Eltzner, Huckemann, et al. (2017), segment process, marks length and/or direction, data of stress fibres in stem cells

fully parametric approach (poster of J. Večeřa)

Theoretical solution

$f_X^{(y)}$ density of the Palm mark distribution

of the process X at the location $y \in B$ (observed)

the reference and observed distributions related by formula

$$f_X^{(y)}(r) = \frac{\rho(y,r)}{\int_M \rho(y,r)\mathrm{d}r}$$

intensity function $\rho(u) = \tau g(r)\mathbb{E}\lambda^*(X,u),\; u = (y,r)$

ρ typically unknown - approximation

reference density transformed into p, homogeneous Poisson process

Segment process with reference directional distribution

segment length $r > 0$ is fixed, mark - direction, $u = (y, \varphi) \in Y$

Poisson segment process η with the intensity measure

$$\lambda(\mathrm{d}(y,\varphi)) = \mathrm{d}y \frac{1}{\pi} \mathrm{d}\varphi$$

$$p(\mathbf{x}) = c \exp(a\, N(\mathbf{x}))\tau^{n(\mathbf{x})} \prod_{u_i \in \mathbf{x}} g(\varphi_i), \ \mathbf{x} \in \mathbf{N},$$

$n(\mathbf{x})$ total number of segments, $N(\mathbf{x})$ total number of intersections between segments, g a reference probability density on $[0, \pi)$, φ_i direction of $i-$th segment u_i, $a \leq 0, \ \tau > 0$ are parameters, c a normalizing constant. The conditional intensity

$$\lambda^*(\mathbf{x}, u) = \tau g(\varphi) \exp(aN_{\mathbf{x}}(u)),$$

$N_{\mathbf{x}}(u)$ number of segments of $\mathbf{x} \setminus \{u\}$ hit by the segment u

Approximation

Saddle-point approximation (Baddeley and Nair, 2012)

$$\mathbb{E}_X e^{aN_{\mathbf{x}}(u)} \approx \mathbb{E}_{\eta(\rho)} e^{aN_{\mathbf{x}}(u)},$$

where $\eta(\rho)$ is a Poisson process with intensity function ρ. Then

$$\mathbb{E}_{\eta(\rho)} e^{aN_{\mathbf{x}}(u)} = \exp\left((e^a - 1)r^2 I_y(\varphi)\right), \;\; u = (y, \varphi),$$

where

$$I_y(\varphi) = \int_0^\pi |\sin(\varphi - \beta)|\rho(y, \beta)\mathrm{d}\beta.$$

$\rho(y, \varphi) = C_y f_X^{(y)}(\varphi)$. Given a stationary extension of X

$$g(\varphi) \approx \frac{Cf_X(\varphi)}{\tau \exp((e^a - 1)Cr^2 J(\varphi))},$$

where $J(\varphi) = \int_0^\pi |\sin(\varphi - \beta)| f_X(\varphi)\mathrm{d}\varphi$.

Takacz-Fiksel estimation

estimation of parameters C, a, τ and the density g

(i) density f_X estimated using a kernel estimator for directional data
(ii) denote $\beta(a, r, C, \varphi) = \exp((e^a - 1)r^2 CJ(\varphi))$, estimate C, a from equations:

$$\sum_{u \in \mathbf{x}} N_{\mathbf{x}}(u) = \frac{\pi|B|C}{J} \sum_{i=1}^{J} \frac{f_X(\varphi_i) N_{\mathbf{x}}(u_i) e^{aN_{\mathbf{x}}(u_i)}}{\beta(a, r, C, \varphi_i)},$$

$$n(\mathbf{x}) = \frac{\pi|B|C}{J} \sum_{i=1}^{J} \frac{f_X(\varphi_i) e^{aN_{\mathbf{x}}(u_i)}}{\beta(a, r, C, \varphi_i)}.$$

on the right integrals are evaluated by Monte Carlo method

(iii) plugg the estimators of C, a in

$$\tau = \frac{\pi C}{J} \sum_{i=1}^{J} \frac{f_X(\varphi_i)}{\beta(a, r, C, \varphi_i)}$$

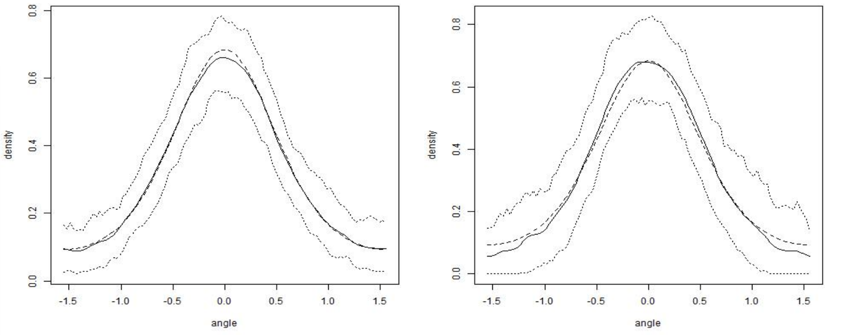
(iv) $g(\varphi) \approx \frac{Cf_X(\varphi)}{\tau \exp((e^a - 1)Cr^2 J(\varphi))}$

Simulation study, numerical results

	true	mean	sd
a	-0.5	-0.496	0.071
τ	1000	1011	154.7
C		361.0	14.5
	true	mean	sd
a	-3	-3.03	0.356
τ	1000	976	141.0
C		235.1	13.0

Tabulka: Means and standard deviations (sd) (based on 100 simulations) of Takacz-Fiksel estimates of scalar parameters in the model having the density with reference directional distribution. The two cases correspond to $a = -0.5,\ a = -3$.

Estimation of reference density



Obrázek: Semiparametric estimation based on 100 simulations of the segment process X, $a = -0.5$ (left), $a = -3$ (right). The average estimator of the reference density (full line) compared to the true reference density (dashed line) of von Mises distribution with parameters $\nu = 0$, $\kappa = 1$. The envelopes (dotted lines) correspond to empirical 90% confidence interval for the estimated reference density, pointwise in 100 points on horizontal axis.

Alternative model formulation

$\mathcal{C}^{(d)}$ space of compact sets in $\mathbb{R}^d$ with Hausdorff metric, $\mathcal{C}_0 \subset \mathcal{C}^{(d)}$ sets with circumcentre at origin

stationary particle process X (Schneider, Weil, 2008), intensity measure θ

$\theta(A) = \gamma \int_{\mathcal{C}_0} \int_{\mathbb{R}^d} \mathbf{1}_A(x+K)\mathrm{d}x\mathrm{d}Q(K), \quad \lambda = \theta/\gamma,$

X is a Gibbs process with Papangelou conditional intensity κ, activity $\tau > 0$, if

$$\mathbb{E}\int_{\mathcal{C}^{(d)}} f(K, X - \delta_K)\,\xi(dK) = \tau\,\mathbb{E}\int_{\mathcal{C}^{(d)}} f(K, X)\kappa(K, X)\,\lambda(dK)$$

for all measurable $f\colon \mathcal{C}^d \times \mathbf{N}^d \to [0, \infty)$, Ruelle (1970)

$\kappa(K, \mu) := \exp\left[-\int_{\mathcal{C}^{(d)}} U(K \cap L)\,\mu(dL)\right], \quad (K, \mu) \in \mathcal{C}^{(d)} \times \mathbf{N}^d$

Stationary segment process X in $\mathbb{R}^2$, intensity measure concentrated on the system of segments $S \subset \mathcal{C}^{(2)}$

potential $U(K) := a\mathbf{1}\{K \neq \emptyset\},\ a > 0,\ K \in \mathcal{C}^{(2)}$ (poster D. Novotná)

The segment process with reference length distribution

Circle $B = b(0, e_a) \subset \mathbb{R}^2$, $L_o = [0, e_a]$ interval of segment lengths

$Y = B \times L_o \times [0, \pi)$ space of segments, $u = (y, r, \varphi) \in Y$, centre y, length r, axial direction φ

Poisson segment process η, intensity measure $\lambda(\mathrm{d}u) = \mathrm{d}y \frac{1}{e_a} \mathrm{d}r \frac{1}{\pi} \mathrm{d}\varphi$ on Y

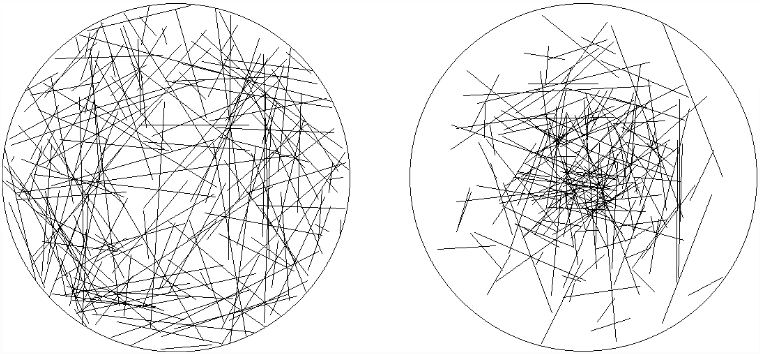
Segment process X, density p w.r.t. η :

$$p(\mathbf{x}) = c 1_{[\mathbf{x} \subset B]} \exp(b\, D(\mathbf{x})) \tau^{n(\mathbf{x})} \prod_{u_i \in \mathbf{x}} g(r_i)$$

$b \in \mathbb{R}$, $\tau > 0$ parameters, r_i length of $i-$th segment u_i, g reference probability density on L_o

$$D(\mathbf{x}) = \sum_{u \in \mathbf{x}} d(u), \quad d(u) = \max_{z \in u} \frac{||z||}{e_a}$$

Segment process simulation



Obrázek: Simulated realizations of the segment process with $e_a = 1$ having parameters $b = 10, \ \tau = 3$ (left) and $b = -10, \ \tau = 4000$ (right), g is beta distribution with parameters $\alpha = 2, \ \beta = 4$.

Properties of the model

Conditional intensity

$$\lambda^*(\mathbf{x}, u) = 1_{[\mathbf{x}\cup\{u\}\subset B]}\tau g(r)\exp(bd(u))$$

Intensity function

$$\begin{aligned} \rho(u) &= \mathbb{E}\lambda^*(X, u) = \tau g(r)\exp(bd(u)), & u \subset B, \\ \rho(u) &= 0, & u \cap (\mathbb{R}^2 \setminus B) \neq \emptyset \end{aligned}$$

X inhomogeneous but isotropic

$f_X^{(y)}$ bivariate (Palm) density of length and direction, segment centered at y. Basic relation:

$$g(r) = \frac{C_y f_X^{(y)}(r, \varphi)}{e^{bd(u)}}$$

From the isotropy $f_X^{(y)}(r, \varphi) = f_X^{(Ay)}(r, A\varphi)$

normalizing constants $C_y = C_{Ay}$ for all $y \in B$ and all rotations A

Semiparametric estimation using maximum pseudolikelihood

estimate parameters b, τ and reference density g

maximize logarithmic pseudolikelihood

$$\log \mathcal{L} = \log(\tau) n(\mathbf{x}) + bD(\mathbf{x}) + \sum_{u \in \mathbf{x}} \log g(r) - \tau \int_{u \subset B} g(r) e^{bd(u)} \mathrm{d}u$$

$$\frac{\partial \log \mathcal{L}}{\partial \tau} = \frac{n(\mathbf{x})}{\tau} - \int_{u \subset B} g(r) e^{bd(u)} \mathrm{d}u,$$

$$\frac{\partial \log \mathcal{L}}{\partial b} = D(\mathbf{x}) - \tau \int_{u \subset B} d(u) g(r) e^{bd(u)} \mathrm{d}u.$$

using the basic relation

$$n(\mathbf{x}) = \tau \int_{u \subset B} C_y f_X^{(y)}(r, \varphi) \mathrm{d}u,$$

$$D(\mathbf{x}) = \tau \int_{u \subset B} C_y d(u) f_X^{(y)}(r, \varphi) \mathrm{d}u.$$

Estimation procedure

(i) discretize the data to a system of k annuli,

kernel density esimator of $f_X^{(y_j)}(r,\varphi)$ in each annulus using rotated data, $j = 1, \ldots, k$

product kernel, length: system of beta kernels (Chen, 1999), directions: circular data kernel

(ii) numerical integration of the basic relation to obtain C_{y_j} as a function of b

(iii) Monte-Carlo integration on the right of PLE equations

(iv) numerical solution of equation for $\frac{D(\mathbf{x})}{n(\mathbf{x})}$ w.r.t. unknown b

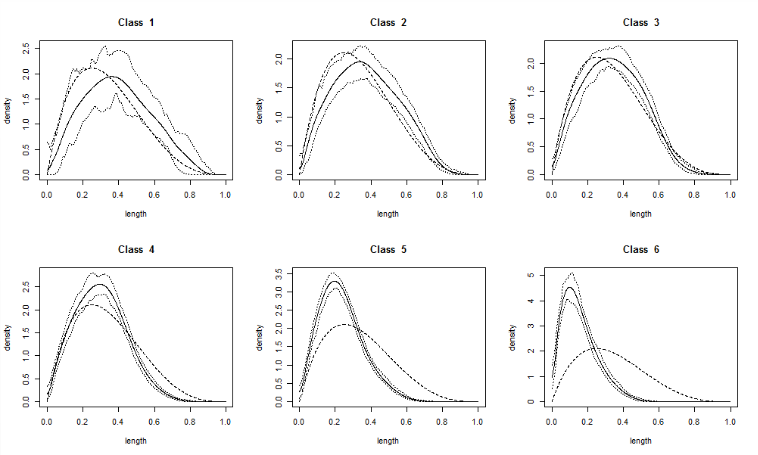
(v) subsequent plugging in, estimation of C_j, τ and finally g

Simulation study, numerical results

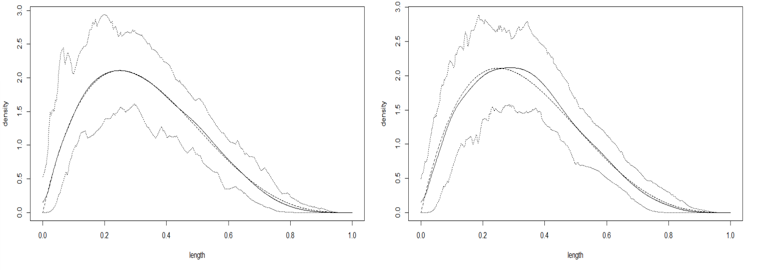
Simulation of the process with parameters
$b = 3, \tau = 900, \alpha = 2, \beta = 4, e_a = 1, k = 6$

	true	mean	sd
b	3	3.059	0.484
τ	900	938.8	189.9
	true	mean	sd
b	3	3.048	0.486
τ	900	954	186

Tabulka: Means and standard deviations (sd) of the maximum pseudolikelihood estimates of scalar parameters in the model with the reference length density. Upper part - sample I (20 simulations), lower part - sample II (40 simulations)



Obrázek: Kernel estimation of the observed length density based on 40 simulations of the inhomogeneous segment process X, $b = 3, \tau = 900$. In each of six classes the average kernel estimator of the observed length density (full line) is compared to the true reference density (dashed line) of beta distribution with parameters $\alpha = 2$, $\beta = 4$. The envelopes (dotted lines)



Obrázek: Semiparametric estimation of reference length density based on sample I (left), II (right) of 20, 40 simulations, respectively, of the segment process X, $b = 3, \tau = 900$. The average (from first four classes) estimator of the reference density (full line) compared to the true reference density (dashed line) of beta distribution with parameters $\alpha = 2$, $\beta = 4$. The envelopes (dotted lines) correspond to empirical 90% confidence interval for the estimated reference density, pointwise in 100 points on horizontal axis.

Some references related to asymptotics

Classical Jensen-Kuensch approach

Mase S. (2000). Marked Gibbs processes and asymptotic normality of maximum pseudolikelihood estimators. Mathematische Nachrichten. 209, 151-169.

Coeurjolly J.F. (2015). Almost sure behavior of functionals of stationary Gibbs point processes. Statist. Probab. Letters 97, 241–246.

Malliavin-Stein method

Torrisi G. L. (2016). Probability approximation of point processes with Papangelou conditional intensity. Bernoulli. In print.

Decorrelation techniques

Blaszczyszyn B., Yogeshwaran D., Yukich J.E. (2016) Limit theory for geometric statistics of clustering point processes. arXiv:1606.03988 [math.PR].

Basic references

Baddeley A, Nair G (2012) Fast approximation of the intensity of Gibbs point processes. Electron J Statist 6:1155–1169.

Beneš V, Večeřa J, Eltzner B, Wollnik C, Rehfeldt F, Králová V, Huckemann S (2017) Estimation of parameters in a planar segment process with a biological application. Im Anal Stereol 36,1, 25–34.

Chen SX (1999) Beta kernel estimators for density functions. Comput Statist Data Anal 31:131–145.

Dereudre D, Lavancier F, Helisova K (2014) Estimation of the intensity parameter of the germ-grain Quermass-interaction model when the number of germs is not observed. Scand J Statist 41:809–829.

Last, G., Penrose, M. (2017) Lectures on the Poisson Process. Cambridge University Press. To appear.

Møller J, Helisova K (2010) Likelihood inference for unions of interacting discs. Scand J Statist 37:365–381.

Ruelle D (1970) Superstable interactions in classical statistical mechanics. Comm Math Phys 18:127–159.

Schneider, R., Weil, W. (2008) Stochastic and Integral Geometry. Springer, Berlin.