Bandwidth selection for kernel estimators of spatial intensity functions

M.N.M. van Lieshout

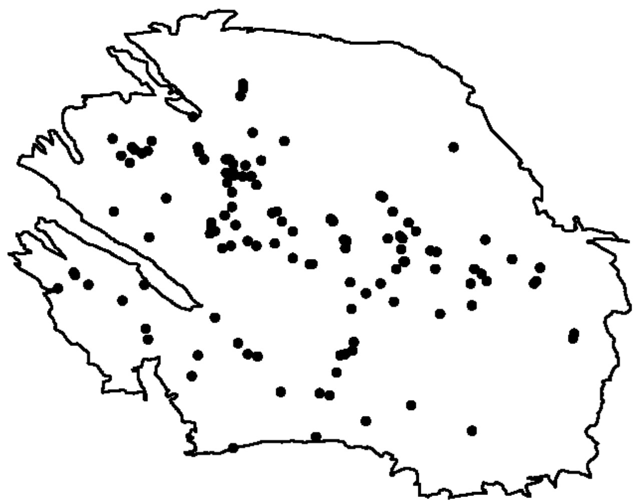
colette@cwi.nl

CWI & University of Twente
The Netherlands

includes joint work with Ottmar Cronie

Point processes

A realisation of a point process Φ on $\mathbb{R}^d$ is a (spatial) pattern, i.e. an **unordered set** of points such that any **bounded** set $A \subset \mathbb{R}^d$ contains only **finitely many** of them.



Consequently, Φ

- contains at most countably many points;
- has no accumulation points.

Intensity function

Let $N(A)$ be the number of points of Φ in set $A \subset \mathbb{R}^d$ and define the set function

$$M(A) = \mathbb{E}N(A),$$

the **expected number of points in** A.

Often

$$M(A) = \int_A \lambda(x)\, dx$$

for some function $\lambda(x) \geq 0$, the **intensity function** of Φ.

Goal: estimate λ based on a realisation $\Phi \cap W$ in a bounded Borel set W (assumed to be open and non-empty).

Kernel estimation

For $x_0 \in W$, set (Berman and Diggle, 1985, 1989)

$$\lambda_{BD}(\widehat{x_0; h, \Phi}, W) := \frac{N(b(x_0, h) \cap W)}{|b(x_0, h) \cap W|}$$

where $b(x_0, h)$ is the closed ball around x_0 with radius h and $|\cdot|$ denotes area.

Remarks:

- **bandwidth** parameter $h > 0$ determines smoothness;
- the box kernel may be replaced by any kernel (symmetric pdf) κ:

$$\lambda(\widehat{x_0; h, \Phi}, W) := h^{-d} \sum_{x \in \Phi \cap W} \kappa\left(\frac{x_0 - x}{h}\right) w_h(x_0, x)^{-1}$$

  with

$$w_h(x_0, x) = w_h(x_0) = h^{-d} \int_W \kappa\left(\frac{x_0 - w}{h}\right) dw.$$

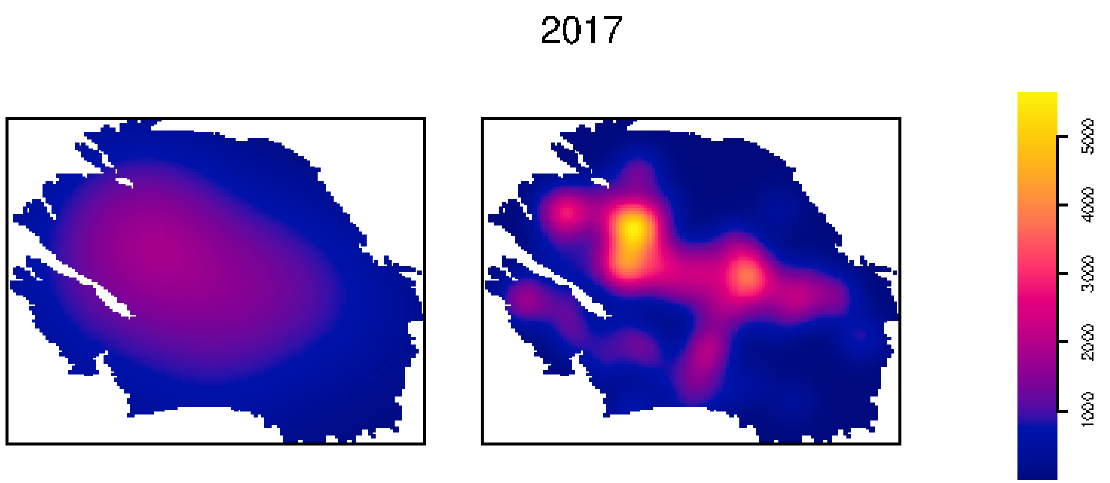
Mass preserving local border correction

Van Lieshout (2012)

For the **local** border correction

$$w_h(x_0, x) = w_h(x) = h^{-d} \int_W \kappa \left(\frac{w - x}{h} \right) dw,$$

$$\int_W \lambda(x; \widehat{h, \Phi}, W) dx = N(W).$$



Left: $h = 0.07$. Right: $h = 0.02$.

Selecting the bandwidth I: Diggle (1985)

Let Φ be a **stationary, isotropic Cox process** with random intensity function Λ. In other words, the distribution of Λ is translation and rotation invariant and given $\Lambda = \lambda$, Φ is an **inhomogeneous Poisson process**:

- the number of points in set A follows a Poisson distribution with mean

$$\int_A \lambda(x) dx;$$

- the points are scattered independently with probability density

$$\lambda(x) / \int_A \lambda(x) dx.$$

To select the bandwidth, minimise (over h) the **mean squared error**

$$\mathbb{E}\left[\{\widehat{\lambda}(0; h, \Phi, W) - \Lambda(0)\}^2 \right].$$

Selecting the bandwidth I (ctd)

For the box kernel in $\mathbb{R}^2$ and $w_h \equiv 1$, minimise

$$\frac{\lambda^2}{\pi^2 h^4} \int_0^{2h} \left\{ 2h^2 \arccos\left(\frac{t}{2h}\right) - \frac{t}{2}(4h^2 - t^2)^{1/2} \right\} dK(t) + \lambda \frac{1 - 2\lambda K(h)}{\pi h^2}$$

over h where

$$\lambda K(h) = \mathbb{E}\left[N(b(0,h)|0 \in \Phi\right].$$

The implementation requires

- an estimator of the constant intensity $\lambda > 0$ of Φ,
- an estimator $\widehat{K}$ of Ripley's K-function (quadratic in the number of points),
- and a Riemann integral over the bandwidth range.

The data must contain at least two points.

Selecting the bandwidth II: Loader (1999)

Let Φ be an **inhomogeneous Poisson process** and maximise the **leave-one-out cross-validation log likelihood**

$$\sum_{x \in \Phi \cap W} \log \widehat{\lambda}(x; h, \Phi \setminus \{x\}, W) - \int_W \widehat{\lambda}(u; h, \Phi, W)\, du.$$

The implementation requires

- discretisation of the observation window into a lattice,
- and at each lattice point, a kernel estimator for every h.

The data pattern must consist of at least two points.

Non-parametric bandwidth selection

The following equation holds:

$$\mathbb{E}\left\{\sum_{x\in\Phi\cap W}\frac{1}{\lambda(x)}\right\} = \int_W \frac{1}{\lambda(x)}\lambda(x)\,dx = |W|.$$

Idea: minimise the discrepancy between $|W|$ and

$$T_\kappa(h;\Phi,W) = \begin{cases} \sum_{x\in\Phi\cap W}\frac{1}{\widehat{\lambda}(x;h,\Phi,W)}, & \Phi\cap W\neq\emptyset, \\ |W|, & \text{otherwise,}\end{cases}$$

to select an appropriate bandwidth h.

No model assumptions required!

Formal justification

Theorem Let ϕ be a locally finite point pattern of distinct points in $\mathbb{R}^d$, observed in some non-empty open and bounded window W, and exclude the trivial case that $\phi \cap W = \emptyset$.

Let κ be a Gaussian kernel. Then $T_\kappa(h; \phi, W)$ is a continuous function of h on $(0, \infty)$. For the box kernel, $T_\kappa(h; \phi, W)$ is piecewise continuous in h.

In either case, with $w_h \equiv 1$,

$$\lim_{h \to 0} T_\kappa(h; \phi, W) = 0$$

and

$$\lim_{h \to \infty} T_\kappa(h; \phi, W) = \infty.$$

Example: Log-Gaussian Cox process

Coles and Jones (1991)

Let Z be a Gaussian random field on W with mean zero and covariance function

$$\sigma^2 \exp\left(-\beta\|x-y\|\right), \quad \sigma^2, \beta > 0,$$

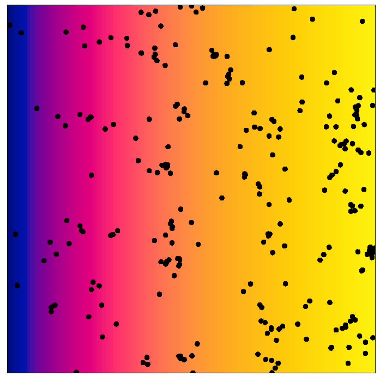
and set

$$\Lambda(x) = \eta(x)\exp\{Z(x)\}.$$

Then the intensity function of the Cox process Φ driven by Λ is

$$\lambda(x) = \eta(x)\exp(\sigma^2/2).$$

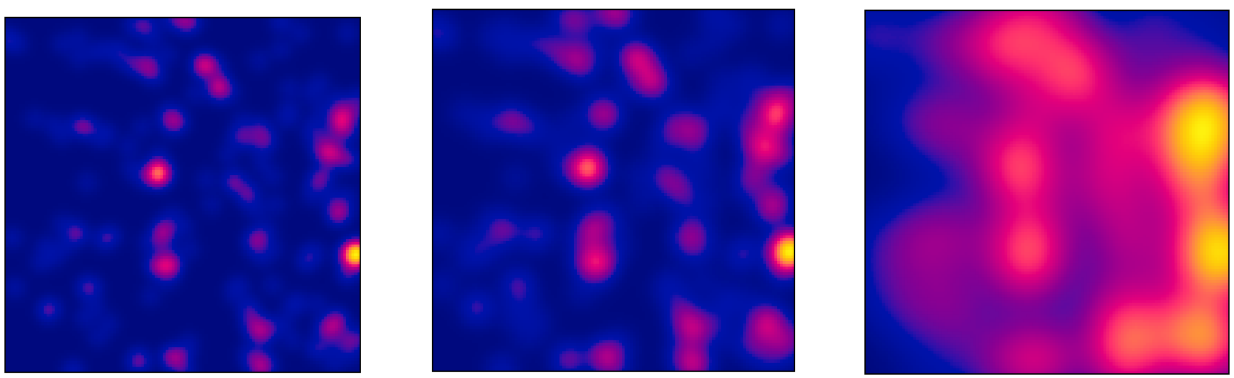
Log-Gaussian Cox process – Simulation



Linear trend

$$\eta(x, y) = 10 + 80x, \quad (x, y) \in [0, 1]^2,$$

$\beta = 50$ and $\sigma^2 = 2 \log 5$, so on average 250 points.



From left to right: State estimation $h = 0.02$, cross-validation $h = 0.03$ and new method $h = 0.08$.

Log-Gaussian Cox process – Results

The quality of a kernel estimator is measured by

$$
\begin{aligned}
MISE(\widehat{\lambda}(\cdot;h)) &= \mathbb{E}\left[\int_W \left(\widehat{\lambda}(x;h,\Phi,W) - \lambda(x)\right)^2 dx\right] \\
&= \int_W \left[\mathrm{Var}(\widehat{\lambda}(x;h,\Phi,W)) + \mathrm{bias}^2(\widehat{\lambda}(x;h,\Phi,W))\right] dx.
\end{aligned}
$$

Based on 100 simulations, the average MISE is given below.

	New	State estimation	Cross-validation
$(\sigma^2, \beta) = (2\log(5), 50)$	89.6	1,477.2	536.0
$(\sigma^2, \beta) = (2\log(2), 10)$	57.5	136.9	112.6
$(\sigma^2, \beta) = (2\log(5), 10)$	335.3	2,960.6	2,251.2

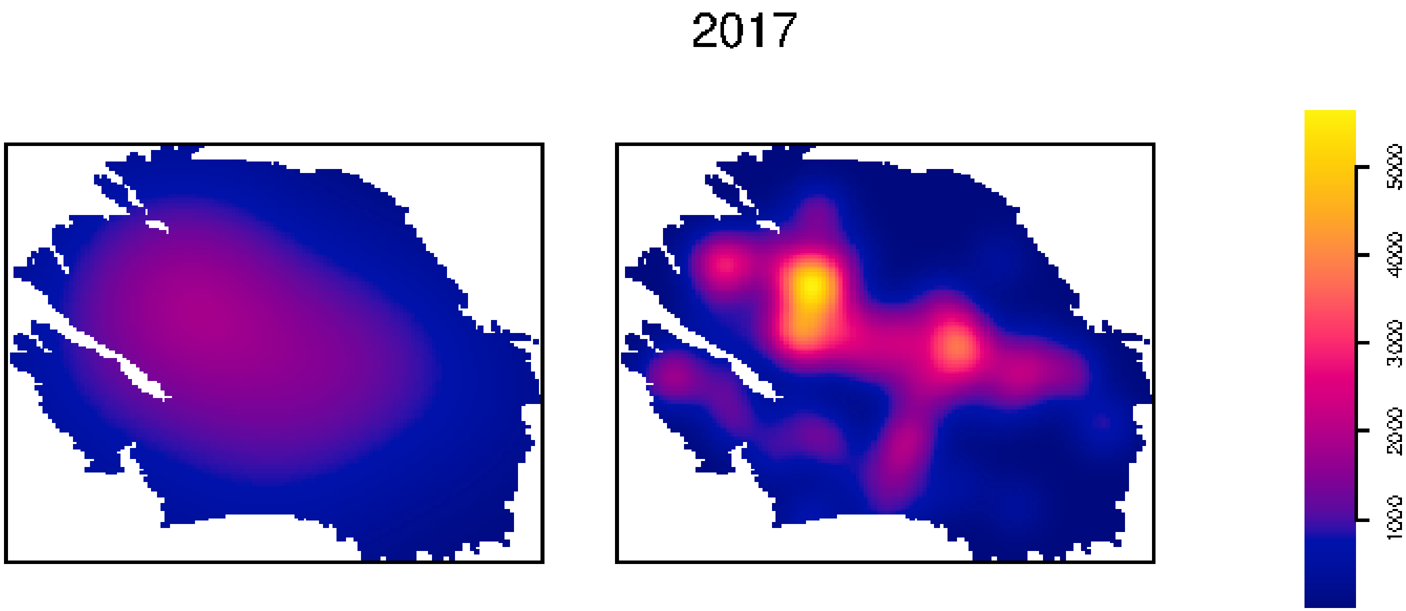
Conclusions

Based on a simulation study, we reach the following conclusions.

- For **clustered** patterns with a moderate number of points, the new method performs the best.
- For **Poisson** processes with a moderate number of points, likelihood based cross-validation performs the best.
- For **regular** patterns with a moderate number of points, the new and the likelihood-based methods give good results.
- For large patterns, the Diggle method seems best.

For details:
O. Cronie and M.N.M. van Lieshout. A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* 105:455–462, 2018.

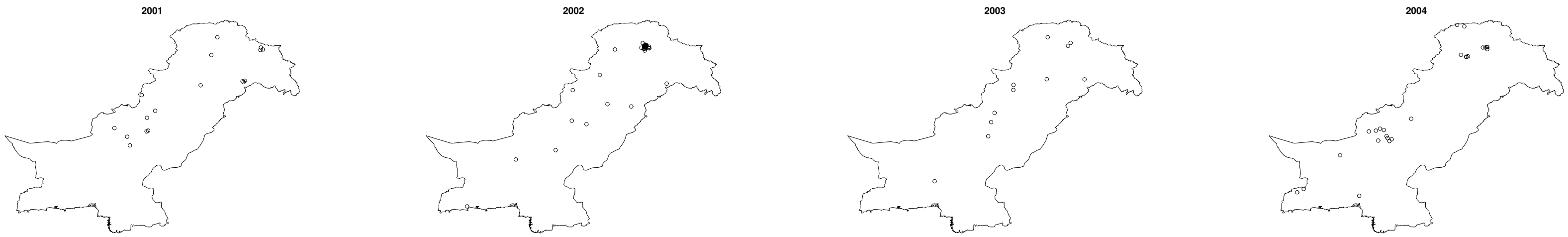
Back to Groningen



Left: $h = 0.07$ (new method). Right: $h = 0.02$ (cross validation).

Asymptotic theory: Which way to infinity?

- Increasing domain: $W_n \to \mathbb{R}^d$. Not applicable
  - when the point process is defined on a fixed domain;
  - unless strong ergodicity conditions are imposed such as stationarity.
- Infill asymptotics: replicated patterns in the same window.



Natural earthquakes in Pakistan, 2001–2004.

Complications

Classic **probability density estimation** results apply for i.i.d. point processes containing exactly one point each (Lo, 2017).

In general, however,

- λ is not normalised,
- the number of points is random,
- and their locations are not necessarily independent.

Pair correlation function

For $A, B \subset \mathbb{R}^d$ define the set function

$$M_2(A, B) = \mathbb{E}\left[\sum_{x\in\Phi}\sum_{y\in\Phi\setminus\{x\}} 1_A(x)\, 1_B(y)\right].$$

Often

$$M_2(A \times B) = \int_A \int_B \rho^{(2)}(x, y)\, dxdy$$

for some function $\rho^2(x, y) \geq 0$, that – provided $\lambda > 0$ – is often scaled to get the **pair correlation function**

$$g(x, y) := \frac{\rho^{(2)}(x, y)}{\lambda(x)\lambda(y)}, \quad x, y \in W.$$

Interpretation: $g(x, y)$ compares the joint probability that x, y belong to Φ to the product of the marginal probabilities.

Beta class of kernels

$$\kappa^{\gamma}(x) := \frac{1}{c(d,\gamma)}(1 - x^T x)^{\gamma}\, 1\{x \in b(0,1)\}, \quad x \in \mathbb{R}^d,$$

for $\gamma \geq 0$, where

$$c(d,\gamma) = \int_{b(0,1)} (1 - x^T x)^{\gamma} dx = \frac{\pi^{d/2}\Gamma(\gamma+1)}{\Gamma(d/2+\gamma+1)}, \quad d \in \mathbb{N}, \gamma \geq 0.$$

Beta kernels are **compactly supported**. For $\gamma > k$, κ^{γ} is k times **continuously differentiable** on $\mathbb{R}^d$. Furthermore, the

$$Q(d,\gamma) := \int_{\mathbb{R}^d} \kappa^{\gamma}(x)^2 dx = \frac{c(d,2\gamma)}{c(d,\gamma)^2} \quad \text{with} \quad Q(2,\gamma) = \frac{(\gamma+1)^2}{(2\gamma+1)\pi}$$

are finite and so are, for all $i = 1, \ldots, d$,

$$V(d,\gamma) := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i^2 \kappa^{\gamma}(x) dx_1 \cdots dx_d = \frac{1}{d+2\gamma+2}.$$

Infill asymptotics regime

Let $\Phi_1, \Phi_2, \ldots$ be i.i.d. simple point processes observed in a bounded open subset $\emptyset \neq W \subset \mathbb{R}^d$ with intensity function $\lambda : W \to (0, \infty)$ and well-defined pair correlation function g.

Write

$$Y_n = \bigcup_{i=1}^{n} \Phi_i.$$

Let the bandwidths $h_n > 0$ be such that, as $n \to \infty$, $h_n \to 0$ and $nh_n^d \to \infty$, and set

$$\widehat{\lambda_n(x_0)} := \frac{\lambda(x_0; \widehat{h_n, Y_n}, W)}{n} = \frac{1}{n} \sum_{i=1}^{n} \lambda(x_0; \widehat{h_n, \Phi_i}, W)$$

Infill asymptotics – Mean squared error

Under the **technical conditions**

- $g : W \times W \to \mathbb{R}$ is bounded;
- $\lambda : W \to (0, \infty)$ is C^2 with $\lambda_{ij} = D_{ij}\lambda$, $i, j = 1, \ldots, d$, Hőlder continuous with index $\alpha > 0$ on W, that is, $\exists C > 0$ such that $\forall i, j = 1, \ldots, d$:

$$|\lambda_{ij}(x) - \lambda_{ij}(y)| \leq C||x - y||^{\alpha}, \quad x, y \in W;$$

the **mean squared error** of $\widehat{\lambda_n(x_0)}$ can be expanded as

$$h_n^4 \frac{V(d,\gamma)^2}{4} \left(\sum_{i=1}^{d} \lambda_{ii}(x_0) \right)^2 + \frac{\lambda(x_0) Q(d,\gamma)}{n h_n^d} + O\left(h_n^{4+\alpha}\right) + O\left(\frac{1}{n h_n^{d-1}}\right)$$

so – provided $\sum_i \lambda_{ii}(x_0) \neq 0$ – the asymptotically optimal bandwidth is

$$h_n^*(x_0) = \frac{1}{n^{1/(d+4)}} \left(\frac{d\lambda(x_0) Q(d,\gamma)}{V(d,\gamma)^2 \left(\sum_{i=1}^{d} \lambda_{ii}(x_0) \right)^2} \right)^{1/(d+4)} .$$

Abramson principle

Idea: in sparse regions more smoothing is necessary then in regions that are rich in points (Abramson, 1982, probability density estimation).

Definition

$$\widehat{\lambda(x_0; h, \Phi}, W) = \sum_{y \in \Phi} \frac{c(y)^d}{h^d} \kappa \left(\frac{x_0 - y}{h} c(y) \right)$$

based on a **weight function** $c : W \to (0, \infty)$ on W. In our context,

$$c(x) = \sqrt{\lambda(x)/\lambda(x_0)}.$$

Extra technical conditions:

- λ is C^5;
- $\gamma > 5$ so the Beta kernel is also C^5.

Infill asymptotics for Abramson estimator

The **mean squared error** of $\widehat{\lambda_n(x_0)}$ can be expanded as

$$h_n^8 \lambda(x_0)^2 \left(\int_{\mathbb{R}^d} A(u; x_0) du \right)^2 + \frac{\lambda(x_0) Q(d, \gamma)}{n h_n^d} + o(h_n^8) + O\left(\frac{1}{n h_n^{d-1}} \right)$$

so the asymptotically optimal bandwidth is

$$h_n^*(x_0) = \frac{1}{n^{1/(d+8)}} \left(\frac{d Q(d, \gamma)}{8 \lambda(x_0) \left(\int_{\mathbb{R}^d} A(u; x_0) du \right)^2} \right)^{1/(d+8)} .$$

An explicit expression for $A(u; x_0)$ depends on partial derivatives of λ up to fourth order.

Remark: taking an adaptive bandwidth reduces the squared bias from $O(h_n^4)$ to $O(h_n^8)$.

Conclusions

Based on asymptotic expansions, we reach the following conclusions.

- For local bandwidth estimation the asymptotically optimal bandwidth in an infill regime is given by

$$h_n^*(x_0) = \frac{1}{n^{1/(d+4)}} \left(\frac{d\lambda(x_0) Q(d,\gamma)}{V(d,\gamma)^2 \left(\sum_{i=1}^d \lambda_{ii}(x_0) \right)^2} \right)^{1/(d+4)}.$$

- For adaptive local bandwidth estimation,

$$h_n^*(x_0) = \frac{1}{n^{1/(d+8)}} \left(\frac{d Q(d,\gamma)}{8\lambda(x_0) \left(\int_{\mathbb{R}^d} A(u; x_0) du \right)^2} \right)^{1/(d+8)}.$$

For details:

M.N.M. van Lieshout. Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. ArXiv 1904.05095, April 2019.