Non-model based bandwidth selection for kernel estimators of spatial intensity functions

M.N.M. van Lieshout

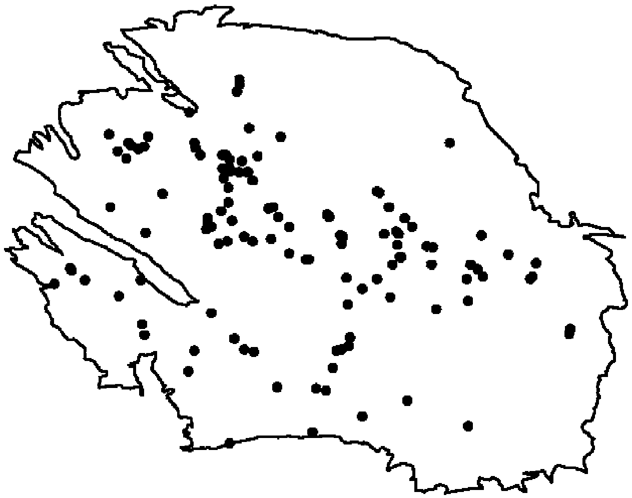
colette@cwi.nl

CWI & University of Twente
The Netherlands

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;
- may place two points at the same position.

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)

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

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;
- box kernel may be replaced by, e.g., a Gaussian kernel κ:

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

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

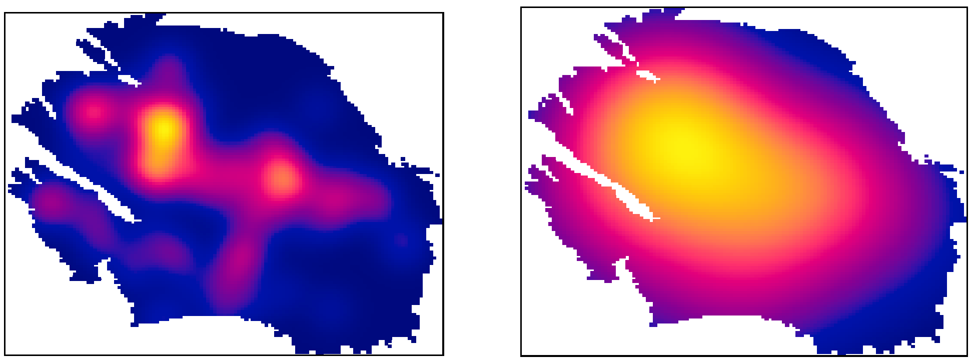
Van Lieshout (2012)

For the **local** border correction

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

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

The bandwidth $h > 0$ controls the amount of smoothing.



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

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

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

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 $\kappa(\cdot)$ 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

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

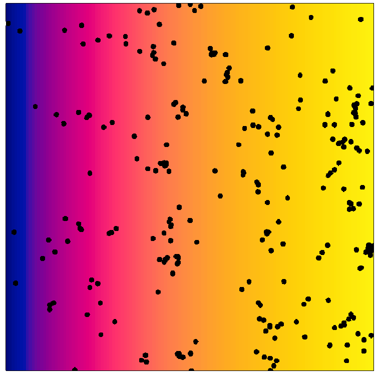
and set

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

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

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

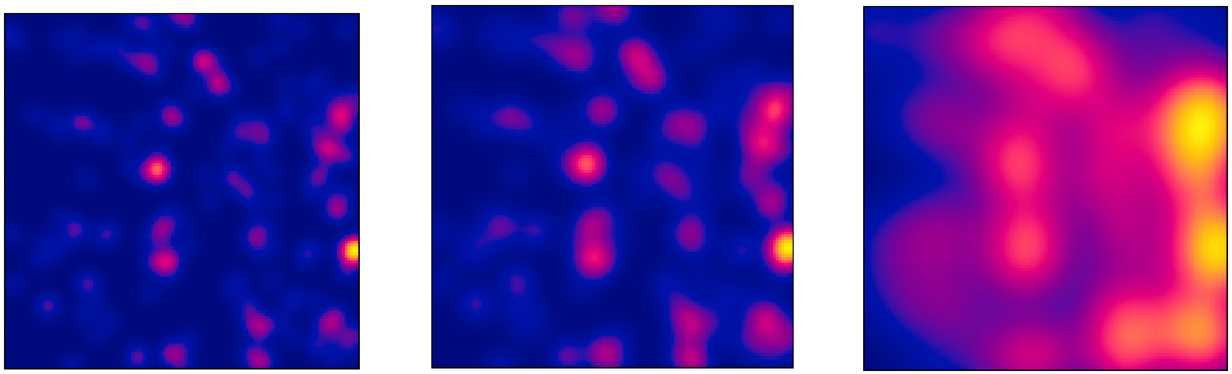
Log-Gaussian Cox process – Simulation



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

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

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.