Biometrika (...), **...**, ..., *pp.* 1–4
Printed in Great Britain Advance Access publication on ...

Supplementary material for 'A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions'

BY O. CRONIE
Department of Mathematics and Mathematical Statistics, Umeå University, 901 87 Umeå, Sweden
ottmar.cronie@umu.se

AND M. N. M. VAN LIESHOUT
Centrum Wiskunde en Informatica, University of Twente, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
marie-colette.van.lieshout@cwi.nl

1. EXTENDED SIMULATION STUDY

1·1. *Setup*

To compare the performance of our new bandwidth selection approach with two established approaches, Poisson process likelihood cross-validation and state estimation, we carried out a simulation study. The point process models used in the study were selected because they allow explicit formulas for their intensity functions and span a range of inhomogeneity and inter-point interaction behaviours.

The simulation study was set up as follows. Given a set of parameters, we generated 100 realizations of each model in the window $W = [0, 1]^2$. For each of the three bandwidth selection approaches, we estimated the bandwidth using no edge correction, with a discretization of 128 values in the range $[0.01, 1.5]$. For the Poisson process likelihood cross-validation method, additionally we used a spatial discretization of $[0, 1]^2$ in a 128×128 grid for the numerical evaluation of the involved integral.

To assess the quality of the bandwidth selection approaches by means of mean integrated squared error, the average integrated squared error over the 100 samples was calculated for each method, where for each sample we used a Gaussian kernel with the selected bandwidths and applied local edge correction. To express the results on a comparable scale, we divided by the expected number of points in $[0, 1]^2$. The calculations were carried out in the R-package spatstat (Baddeley et al., 2015).

Below, we give the details of the models used in the study. The results are summarized in Table 1 and our overall conclusions can be found in the paper.

1·2. *Poisson processes*

We start by evaluating a set of Poisson processes, with different intensity functions. For a non-empty, open and bounded $W \subseteq \mathbb{R}^d$, the Poisson process with intensity function $\lambda : W \to \mathbb{R}^+$ is constructed as follows. First, generate a Poisson distributed random variable N with rate

$$\int_W \lambda(x)\,\mathrm{d}x.$$

Then, upon the outcome $N = n$, sample n independent and identically distributed points with common probability density function

$$\frac{\lambda(x)}{\int_W \lambda(y)\,\mathrm{d}y}, \quad x \in W.$$

The ensemble of points thus generated form a realization of the desired Poisson process. This model is particularly amenable to calculations due to the strong independence assumptions (Chiu et al., 2013, Chapter 2).

In the first experiment, we generated 100 independent realizations of a homogeneous Poisson process, $\lambda(\cdot) \equiv \lambda > 0$, in the unit square for low, $\lambda = 10$, medium, $\lambda = 50$, and high, $\lambda = 250$, intensity values. The expected number of points in $[0, 1]^2$ is λ.

In the second Poisson experiment, we generated 100 independent realizations of a Poisson process in the unit square with linear intensity function

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

for weak, $\alpha = 1$, medium, $\alpha = 80$, and strong, $\alpha = 480$, trend. The expected number of points in $[0, 1]^2$ is $10 + \alpha/2$.

Finally, we generated 100 independent realizations of a Poisson process in the unit square with modulated intensity function

$$\lambda(x, y) = \alpha + \beta \cos(10x), \quad (x, y) \in [0, 1]^2.$$

To generate patterns with a low, medium and large intensity, we considered the (α, β) combinations $(10, 2)$, $(50, 20)$ and $(250, 100)$. The expected number of points in $[0, 1]^2$ is $\alpha + \beta \sin(10)/10$.

The results, which are displayed in Table 1, correspond to one round of experiments and the conclusions are based on our overall observations; the same remark applies to the other models to be discussed below.

In the high level intensity setting, the state estimation approach performs best and the new approach has the highest average integrated squared error. In the homogeneous case, we see that for the low intensity level the likelihood-based approach is performing best and the state estimation approach is giving rise to the highest average integrated squared error. For the medium level intensity, the new and state estimation approaches yield comparable average integrated squared errors, both being outperformed by the likelihood-based approach. In the inhomogeneous cases, for small and medium intensities, it seems that the likelihood-based approach has the best performance, followed by the new approach.

1·3. *Cox processes*

We next turn to a class of clustered point processes, namely Cox processes (Chiu et al., 2013, Section 5.2). As indicated by Table 1, the new approach seems to strongly outperform the competing approaches.

A Cox process is the generalization of a Poisson process that allows for a random intensity function Λ. Such models are appropriate to describe clustering due to latent environmental heterogeneity. The log likelihood depends on the distribution of Λ and is usually not available in closed form. The intensity function equals $\lambda(x) = E\{\Lambda(x)\}$ for $x \in W$.

In the first experiment we generated 100 independent realizations of a homogeneous Matérn cluster process (Chiu et al., 2013, Section 5.3) in the unit square, for various degrees of clustering.

This is a Cox process with

$$\Lambda(y) = \sum_{x \in \Phi} 1\{y \in B(x, r)\}/(\pi r^2),$$

where Φ is a homogeneous Poisson process with intensity κ and $B(x, r)$ is the closed ball of radius r, centred at $x \in \mathbb{R}^2$. In words, each point of Φ acts as parent to a family that consists of a Poisson number of offspring, with mean size μ, born independently of each other at positions that are uniformly distributed in a ball of radius r around the parent. The Matérn cluster process Ψ consists of the ensemble of all offspring.

We considered two levels, $\kappa = 10, 20$, for the parent intensity, two ranges of clustering, $r = 0.05, 0.1$, and two mean offspring sizes, $\mu = 3, 10$. To avoid edge effects, the parent process was generated on a dilated window. The resulting point process is homogeneous with intensity $\kappa\mu$ so the expected number of points in $[0, 1]^2$ is $\kappa\mu$.

Table 1 indicates that, throughout, the new approach outperforms its competitors. For a low degree of clustering, that is, small values of μ, the likelihood-based method outperforms the state estimation approach; for larger degrees the state estimation method works better. The state estimation and likelihood-based methods tend to have decreased integrated squared errors when clusters are more diffuse.

Our next set of experiments concerned the log-Gaussian Cox process (Møller et al., 1998) described in the paper. In addition to the two experiments presented in the paper, we also sampled from the homogeneous model with $\eta(x, y) = \lambda > 0$ and intensity $\lambda \exp(\sigma^2/2)$, which is also equal to the expected number of points in the unit square.

We considered two levels $\lambda = 10, 50$, two degrees of variability, $\sigma^2 = 2 \log 5$ and $\sigma^2 = 2 \log 2$, as well as two degrees of clustering $\beta = 10, 50$. The results are summarized in Table 1 and, as already mentioned, the new approach works best. It further seems that the likelihood-based approach has the second best performance. Increasing β, that is, decreasing the range of interaction, tends to lead to a smaller average integrated squared error. Increasing the variability, and hence the intensity, yields a higher average integrated squared error.

1·4. *Determinantal point processes*

We finally turn to planar determinantal point processes (Lavancier et al., 2015). They exhibit inhibition and allow explicit expressions

$$\rho^{(n)}(x_1, \ldots, x_n) = \det\{C(x_i, x_j)\}_{i,j}, \quad x_i \in \mathbb{R}^d,$$

for the product densities in terms of the determinant, det, of a kernel $C : \mathbb{R}^d \to \mathbb{R}$. Hence, the intensity function is $\lambda(x) = \rho^{(1)}(x) = C(x, x)$. We used the kernel

$$\sigma^2 \exp\left(-\beta \|x_1 - x_2\|\right), \quad x_1, x_2 \in \mathbb{R}^d,$$

which results in a homogeneous determinantal point process with intensity $\lambda(x, y) = C\{(x, y), (x, y)\} = \sigma^2$. Hence, the expected number of points in $[0, 1]^2$ is σ^2.

We generated 100 independent realizations of this homogeneous model, on the unit square, for low, $\sigma^2 = 10$, medium, $\sigma^2 = 50$, and high, $\sigma^2 = 250$, intensity values. We further considered two values for β, namely 20 and 50. For $\sigma^2 = 250$, the model is not valid for the larger of the interaction ranges, that is, for $\beta = 20$. To alleviate edge effects, we generated realizations on a dilated window of size $2/\beta$ and clipped them in the unit square. Table 1 indicates that for small intensities the likelihood-based method performs best and the state estimation approach generates the highest average integrated squared error. For medium intensities the likelihood-based approach performs best, with the other two approaches giving rise to similar average integrated squared errors. In the case of high intensities, the state estimation method performs best and the new method gives rise to the highest average integrated squared error.

Next, we applied independent thinning (Chiu et al., 2013, Section 5.1) to the realizations of the homogeneous determinantal point process described above. We employed the linear retention probability function $p(x, y) = (10 + 80x)/90$, $(x, y) \in [0, 1]^2$, which results in the intensity function

$$\lambda(x, y) = \sigma^2 p(x, y) = \frac{\sigma^2}{90}(10 + 80x).$$

We considered the cases $\lambda = 50, 250$ and $\beta = 20, 50$. Here, the expected number of points in $[0, 1]^2$ is $5\sigma^2/9$. The results are summarized in Table 1. There are indications that in the low and medium intensity cases the likelihood-based approach performs best, followed by the new approach. On the other hand, in the high intensity setting the likelihood-based approach and the state estimation approach seem to perform almost equally well.

In the last experiment, we applied independent thinning to realizations of the homogeneous determinantal point process described above, using the modulated retention probability function

$$p(x, y) = \frac{10 + 2\cos(10x)}{12}, \quad (x, y) \in [0, 1]^2,$$

for $\lambda = 10, 50$ and $\beta = 20, 50$. The expected number of points in $[0, 1]^2$ is $\sigma^2\{10 + \sin(10)/5\}/12$. The results are summarized in Table 1. For the smallest intensity, the new approach and the likelihood-based approach seem to have similar, best, performance. As we increase the intensity, the new approach tends to generate a higher average integrated squared error than the likelihood-based approach. The state estimation method tends to perform poorly for small intensities, perhaps because there are not enough points to reliably estimate the involved K-function.

REFERENCES

BADDELEY, A., RUBAK, E. & TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton: Taylor & Francis/CRC Press.

CHIU, S. N., STOYAN, D., KENDALL, W. S. & MECKE, J. (2013). *Stochastic Geometry and its Applications*. Chichester: Wiley, 3rd ed.

LAVANCIER, F., MØLLER, J. & RUBAK, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**, 853–877.

MØLLER, J., SYVERSVEEN, A. & WAAGEPETERSEN, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

[*Received January .... Revised January ...*]

Table 1. *Average integrated squared error, divided by the expected number of points, over* 100 *simulations of different models on the unit square.*

Model/Method	New	State estimation	Cross-validation
Homogeneous Poisson process			
$\lambda = 10$	3.4	11.0	2.8
$\lambda = 50$	10.0	10.6	5.9
$\lambda = 250$	30.7	12.2	16.4
Poisson process with linear trend			
$\alpha = 1$	4.2	13.0	3.8
$\alpha = 80$	8.9	11.4	7.4
$\alpha = 480$	23.0	14.8	17.9
Modulated Poisson process			
$(\alpha, \beta) = (10, 2)$	4.0	11.8	3.7
$(\alpha, \beta) = (50, 20)$	10.8	12.7	9.0
$(\alpha, \beta) = (250, 100)$	29.6	17.9	20.0
Homogeneous Matérn cluster process			
$(\kappa, r, \mu) = (10, 0.05, 3)$	19.5	233.5	158.3
$(\kappa, r, \mu) = (10, 0.1, 3)$	15.0	64.5	47.7
$(\kappa, r, \mu) = (10, 0.05, 10)$	44.5	627.4	783.9
$(\kappa, r, \mu) = (10, 0.1, 10)$	44.2	175.2	214.6
$(\kappa, r, \mu) = (20, 0.05, 3)$	22.8	203.8	158.4
$(\kappa, r, \mu) = (20, 0.1, 3)$	21.7	58.5	50.6
$(\kappa, r, \mu) = (20, 0.05, 10)$	62.9	552.5	765.5
$(\kappa, r, \mu) = (20, 0.1, 10)$	71.9	168.9	236.2
Homogeneous log-Gaussian Cox process			
$(\lambda, \sigma^2, \beta) = (10, 2\log 5, 50)$	14.1	70.8	17.8
$(\lambda, \sigma^2, \beta) = (10, 2\log 2, 10)$	9.7	28.1	12.4
$(\lambda, \sigma^2, \beta) = (10, 2\log 5, 10)$	55.0	376.5	208.6
$(\lambda, \sigma^2, \beta) = (50, 2\log 5, 50)$	73.7	641.7	336.6
$(\lambda, \sigma^2, \beta) = (50, 2\log 2, 10)$	61.1	130.9	102.2
$(\lambda, \sigma^2, \beta) = (50, 2\log 5, 10)$	383.4	2,875.1	2,229.9
Homogeneous determinantal point process			
$(\sigma^2, \beta) = (10, 50)$	4.0	11.0	3.2
$(\sigma^2, \beta) = (10, 20)$	4.1	13.1	3.2
$(\sigma^2, \beta) = (50, 50)$	10.9	10.6	6.3
$(\sigma^2, \beta) = (50, 20)$	9.6	8.4	4.8
$(\sigma^2, \beta) = (250, 50)$	27.6	10.4	11.7
Determinantal point process with linear trend			
$(\sigma^2, \beta) = (50, 50)$	7.0	10.6	5.3
$(\sigma^2, \beta) = (50, 20)$	6.2	9.7	5.1
$(\sigma^2, \beta) = (250, 50)$	16.7	10.8	10.9
Modulated determinantal point process			
$(\sigma^2, \beta) = (10, 50)$	4.0	11.7	4.9
$(\sigma^2, \beta) = (10, 20)$	3.7	13.2	3.3
$(\sigma^2, \beta) = (50, 50)$	9.8	10.6	6.4
$(\sigma^2, \beta) = (50, 20)$	8.5	8.4	4.8