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

Summary

We propose a new bandwidth selection method for kernel estimators of spatial point process intensity functions. The method is based on an optimality criterion motivated by the Campbell formula applied to the reciprocal intensity function. The new method is fully nonparametric, does not require knowledge of higher-order moments, and is not restricted to a specific class of point process. Our approach is computationally straightforward and does not require numerical approximation of integrals.

Some key words: Bandwidth selection; Campbell formula; Intensity function; Kernel estimation; Point process.

1. Introduction

Spatial point patterns arise in many scientific domains. Typical examples include the map of trees in a forest stand, the addresses of individuals infected with some disease, and the locations of cells in a tissue (Diggle, 2014; Gelfand et al., 2010; Illian et al., 2008).

The analysis of such patterns usually involves estimating the intensity function, that is, the likelihood of finding a point as a function of location. Sometimes the scientific context suggests a parametric form for the intensity function, perhaps in terms of covariate information. More often, nonparametric estimation is called for. In both cases, it is important to estimate the intensity function in a reliable way. Indeed, this is more urgent than ever in light of the recent development of functional summary statistics that correct for spatial heterogeneity (Baddeley et al., 2000; van Lieshout, 2011); the current paper is motivated by our recent work (Cronie & van Lieshout, 2016) in this direction.

Here we focus on nonparametric estimation. A few approaches have been suggested in the literature. For example, one may divide the observation window into quadrats and count the number of points that fall in each. An obvious drawback of this method is its strong dependence on the size and shape of the quadrats. A partial solution is to use spline (Ogata, 1998) or kernel smoothing, where the shape of the kernel tends to have little impact; the size parameter, though, remains crucial. Therefore, data-dependent, adaptive methods have been proposed in which the quadrats are replaced by the cells in the pattern's Delaunay tessellation, as shown in a 2007 University of Groningen PhD thesis by W. E. Schaap and in Schaap & van de Weygaert (2000), or the cells in the pattern's Voronoi tessellation (Bernardeau & van de Weygaert, 1996; Ord, 1978). The relative merits of these approaches were investigated by Barr & Schoenberg (2010) and van Lieshout (2012). Further details can be found in Diggle (2014, §§ 3 and 5).

The goal of this paper is to discuss and compare various approaches to choosing the size parameter, the bandwidth, when performing kernel estimation. The bandwidth is the parameter determining to what degree the variations in the intensity are smoothed out. At first glance this task might appear to be trivial, but it turns out to be extremely challenging. Apart from various rules of thumb (Baddeley et al., 2015, § 6.5; Illian et al., 2008, § 3.3; Scott, 1992, § 6; or the first edition of Diggle, 2014), there are essentially two main approaches, which both rely on specific model assumptions: a Poisson process likelihood crossvalidation approach (Loader, 1999, § 5.3) and minimization of the mean squared error in state estimation for a stationary isotropic Cox process (Diggle, 1985). As an alternative, we introduce a new, computationally simple and natural approach that relies solely on the Campbell formula (Matthes et al., 1978), which relates pattern averages to intensity function-weighted spatial averages.

2. Preliminaries

2·1. *Intensity functions*

Throughout this paper, let Ψ be a point process in $\mathbb{R}^d$, $d \geqslant 1$, observed within some nonempty open and bounded observation window $W \subseteq \mathbb{R}^d$ (Daley & Vere-Jones, 2008; Diggle, 2014; Illian et al., 2008; van Lieshout, 2000).

The abundance of points as a function of location is captured by the intensity function $\lambda : \mathbb{R}^d \to \mathbb{R}^+$, the Radon–Nikodym derivative of the first-order moment measure, provided it exists (Daley & Vere-Jones, 2008, § 9.5). From now on, we will assume that Ψ admits a well-defined intensity function λ. The Campbell theorem (Daley & Vere-Jones, 2008, p. 65) then states that for any measurable and nonnegative function $f : \mathbb{R}^d \to \mathbb{R}^+$,

$$E \left\{ \sum_{x \in \Psi} f(x) \right\} = \int_{\mathbb{R}^d} f(x) \lambda(x) \, dx, \tag{1}$$

where the integration is with respect to Lebesgue measure $\ell(\cdot)$.

2·2. *Kernel estimation*

A kernel function $\kappa : \mathbb{R}^d \to \mathbb{R}^+$ is defined to be a d-dimensional symmetric probability density function (Silverman, 1986, p. 13). Given a bandwidth or scale parameter $h > 0$, the intensity function may be estimated by

$$\hat{\lambda}(x; h) = \hat{\lambda}(x; h, \Psi, W) = h^{-d} \sum_{y \in \Psi \cap W} \kappa \left(\frac{x - y}{h} \right) w_h(x, y)^{-1}, \quad x \in W, \tag{2}$$

where $w_h(x, y)$ is an edge correction factor; $w_h \equiv 1$ corresponds to no edge correction. The global edge correction factor proposed in Berman & Diggle (1989) and Diggle (1985) corresponds to

$$w_h(x, y) = h^{-d} \int_W \kappa \left(\frac{x - u}{h} \right) du,$$

whereas the choice of the local factor

$$w_h(x, y) = h^{-d} \int_W \kappa \left(\frac{u - y}{h} \right) du,$$

suggested in van Lieshout (2012), results in the mass preservation property that

$$\int_W \hat{\lambda}(x; h) \, dx = \Psi(W), \tag{3}$$

which is the number of points in W. At least for the beta kernels discussed below, which are strictly positive on an open ball of radius h around the centre point, the assumption that W is open implies that neither edge correction factor can be zero. We emphasize that upon taking expectations on both sides of (3) we obtain global unbiasedness.

For κ, one may for instance pick a member of the class of beta kernels, which are also known as multi-weight kernels (Hall et al., 2004). More specifically, for any $\gamma \geqslant 0$, set

$$\kappa(x) = \kappa^\gamma(x) = \frac{\Gamma(d/2 + \gamma + 1)}{\pi^{d/2}\Gamma(\gamma + 1)}(1 - x^T x)^\gamma \, 1\{x \in B(0, 1)\}, \quad x \in \mathbb{R}^d. \tag{4}$$

Here $B(0, 1)$ is the closed unit ball centred at the origin. Specific examples include the box kernel, $\gamma = 0$, and the Epanechnikov kernel, $\gamma = 1$. Another widely used choice is the Gaussian kernel

$$\kappa(x) = (2\pi)^{-d/2}\exp(-x^T x/2), \quad x \in \mathbb{R}^d. \tag{5}$$

More details and further examples can be found in Silverman (1986).

The quality of a kernel estimator may be measured by the mean integrated squared error

$$E\left[\int_W \left\{\hat{\lambda}(x; h) - \lambda(x)\right\}^2 \mathrm{d}x\right] = \int_W \left[\mathrm{var}\left\{\hat{\lambda}(x; h)\right\} + \mathrm{BIAS}^2\left\{\hat{\lambda}(x; h)\right\}\right]\mathrm{d}x, \tag{6}$$

where $\mathrm{BIAS}\{\hat{\lambda}(x; h)\} = E\{\hat{\lambda}(x; h)\} - \lambda(x)$. Since (6) depends on the unknown intensity and second-order moment measure, in practice it is only of limited value.

3. A new approach to bandwidth selection

To motivate the new approach, consider the function $f : \mathbb{R}^d \to \mathbb{R}^+$ defined by $f(x) = 1(x \in W)/\lambda(x)$, which is measurable if $\lambda(x) > 0$ for all $x \in W$. Applying the Campbell formula (1) to f we obtain

$$E\left\{\sum_{x \in \Psi \cap W} \frac{1}{\lambda(x)}\right\} = \int_W \frac{1}{\lambda(x)}\lambda(x)\,\mathrm{d}x = \ell(W). \tag{7}$$

In other words, $\sum_{x \in \Psi \cap W}\lambda(x)^{-1}$, known as the Stoyan & Grabarnik (1991) statistic, is an unbiased estimator of the window size $\ell(W)$. If λ is replaced by its estimated counterpart, $\hat{\lambda}$, the left-hand side of (7) is a function of the bandwidth while the right-hand side is not. We may therefore minimize the discrepancy between $\ell(W)$ and $\sum_{x \in \Psi \cap W}\hat{\lambda}(x; h, \Psi, W)^{-1}$ to select an appropriate bandwidth h. Formally, we define

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

and choose the bandwidth $h > 0$ by minimizing

$$F_\kappa\{h; \Psi, W, \ell(W)\} = \{T_\kappa(h; \Psi, W) - \ell(W)\}^2. \tag{8}$$

Since W is assumed to be open and $\kappa(0) > 0$ for the kernels considered in this paper, (8) is well-defined. We obtain the final estimate $\hat{\lambda}$ by inserting the chosen h into (2). In essence, we have transformed the problem of selecting the unknown optimal bandwidth to that of estimating the known size of the study region. Therefore, our new approach is fully non-model-based and computationally straightforward. We

emphasize that $\Psi \cap W$ almost surely contains finitely many points, since W is bounded. The convention that $T_\kappa(h; \Psi, W) \equiv \ell(W)$ when $\Psi \cap W = \emptyset$ is simply a way of saying that if nothing is observed, there is nothing to estimate, so we always estimate correctly. As an aside, the Campbell–Mecke formula has also been used for parametric inference, which is sometimes referred to as Takacs–Fiksel estimation. For an overview and references to the literature, see Møller & Waagepetersen (2017).

Next, we consider the continuity properties of $T_\kappa(\cdot; \Psi, W)$ and its limits as the bandwidth approaches zero and infinity.

THEOREM 1. *Let ψ be a locally finite point pattern of distinct points in $\mathbb{R}^d$, observed in some nonempty open and bounded window W, and exclude the trivial case that $\psi \cap W = \emptyset$. Let $\kappa(\cdot)$ be a Gaussian or beta kernel with $\gamma > 0$. Then $T_\kappa(h; \psi, W)$ is a continuous function of h on $(0, \infty)$. For the box kernel, $T_\kappa(h; \psi, W)$ is piecewise continuous in h. In all cases, $\lim_{h \to 0} T_\kappa(h; \psi, W) = 0$. Also,*

$$\lim_{h \to \infty} T_\kappa(h; \psi, W) = \begin{cases} \infty, & w_h \equiv 1, \\ \ell(W), & \text{for the global and local edge corrections.} \end{cases}$$

Proof. The functions κ^γ defined in (4) are continuous for $\gamma > 0$, as is the Gaussian kernel (5). The function $h \mapsto 1/h$ is continuous on the open interval $(0, \infty)$, so both the function $h \mapsto 1/h^d$ and the composition $h \mapsto \kappa^\gamma(x/h)$ are also continuous in h on $(0, \infty)$. Hence by the dominated convergence theorem, using the fact that W and κ^γ are bounded, $w_h(x, y)$ is continuous in h on $(0, \infty)$ for local, global and no edge correction and $x, y \in W$; moreover, $w_h(x, y) > 0$ since W is open. Therefore, for fixed $x \in W$, the function $\hat{\lambda}(x; h)$ is also continuous in h on $(0, \infty)$. Finally, since for $x \in \psi \cap W$, the kernel estimator $\hat{\lambda}(x; h) \geqslant h^{-d} \kappa^\gamma(0)/w_h(x, x) > 0$, we conclude that $T_{\kappa^\gamma}(h; \psi, W)$ is continuous in $h \in (0, \infty)$. Box kernels are discontinuous at 1, making T_{κ^0} piecewise continuous.

Now let h tend to zero. For the beta kernels with $\gamma \geqslant 0$, the support of $h^{-d}\kappa^\gamma\{(\cdot - y)/h\}$, $y \in W$, is the ball centred at y with radius h. Since the volume of such a ball tends to zero as h tends to zero and W is open, using the fact that κ^γ is a symmetric probability density, we obtain that $\lim_{h \to 0} w_h(x, y) = 1$ for all $x, y \in W$ for global and local edge correction. This holds trivially if no edge correction is applied. For the Gaussian kernel, $h^{-d}\kappa\{(\cdot - y)/h\}$, $y \in W$, corresponds to the probability density of d independent normally distributed random variables with standard deviation h centred at the components of y. Take a sequence $h_n \to 0$ and write X_n for the corresponding Gaussian random vector. Then X_n converges in distribution to the Dirac mass at y by Lévy's continuity theorem. Since W is open, it is a continuity set, and the weak convergence of X_n implies that w_{h_n} tends to 1, the probability that y belongs to W, for local edge correction. The global case follows from the symmetry of κ, and the case $w_{h_n} \equiv 1$ is trivial. For $x = y$ and all $h > 0$, $\kappa\{(x - y)/h\} = \kappa(0)$. Therefore, having excluded the case that $\psi \cap W$ is empty, for all $x \in \psi \cap W$, the sum $\sum_{y \in \psi \cap W} \kappa\{(x - y)/h\} w_h^{-1}(x, y) \geqslant \kappa(0) w_h^{-1}(x, x)$, which tends to $\kappa(0) > 0$ as $h \to 0$. Observing that h^d tends to zero with h, we obtain the desired limit $\lim_{h \to 0} T_\kappa(h; \psi, W) = 0$.

Next, let h tend to infinity. Since $\kappa(\cdot) \leqslant \kappa(0)$, $\lim_{h \to \infty} h^{-d}\kappa\{(x - y)/h\} = 0$ for all $x, y \in W$. Therefore, $T_\kappa(h; \psi, W)$ tends to infinity when $w_h \equiv 1$ and $\psi \cap W \neq \emptyset$. If one does correct for edge effects, by the continuity of κ in 0 for all considered kernels, $\kappa\{(x - y)/h\}$ tends to $\kappa(0)$ as $h \to \infty$. Furthermore, since W is bounded, by the dominated convergence theorem, $\int_W \kappa\{(x - u)/h\} \, du \to \kappa(0)\ell(W)$. Upon using the symmetry of κ, we obtain that for both local and global edge correction, $h^d w_h(x, y) \to \kappa(0)\ell(W)$. Therefore $\hat{\lambda}(x; h) \to \psi(W)/\ell(W)$ and $\lim_{h \to \infty} T_\kappa(h; \psi, W) = \ell(W)$. This completes the proof. □

Theorem 1 may not hold if a leave-one-out estimator is used for the intensity function; indeed $T_\kappa(h; \Psi, W)$ may not be well-defined. Furthermore, Theorem 1 demonstrates that when $w_h(\cdot) \equiv 1$, with probability 1, the minimum of (8) is zero and this minimum is attained. It need not be unique, however, since $T_\kappa(h; \Psi, W)$ may not be a monotone function of h, but we conjecture that monotonicity holds when κ is the Gaussian kernel.

Expression (8) may be exploited as a general intensity estimation criterion: given some collection Γ of functions $\gamma : W \to (0, \infty)$, possibly depending on Ψ and W, which are estimating the underlying intensity function λ, one may use as estimator $\hat{\lambda}$ a minimizer of the objective functional $F : \Gamma \to [0, \infty)$

defined by

$$\gamma \mapsto F(\gamma) = \{T(\gamma; \Psi, W) - \ell(W)\}^2 = \left\{ \sum_{x \in \Psi \cap W} \frac{1}{\gamma(x)} - \ell(W) \right\}^2, \quad \gamma \in \Gamma. \tag{9}$$

The estimating equation (9) suggests a general way of estimating marginal Radon–Nikodym derivatives, without explicit knowledge of the multivariate structures, the topic of our ongoing research. Depending on the choice of function class Γ, (9) may generate existing estimators. For example, in the simplest case where each $\gamma(\cdot) \equiv \gamma > 0$ is assumed constant, we obtain $\hat{\lambda}(\cdot) \equiv \Psi(W)/\ell(W)$. This is the classical intensity estimator for a homogeneous point process and, in particular, the maximum likelihood estimator for a homogeneous Poisson process (Chiu et al., 2013, § 4.7.3).

4. NUMERICAL EVALUATION

To compare the performance of the bandwidth selection approach in § 3 with current methods, we performed a simulation study. First, we describe the two most common methods.

In state estimation (Diggle, 1985), one treats $\hat{\lambda}$ as a state estimator for the random intensity function Λ of a stationary isotropic Cox process Ψ and minimizes the mean squared error $E[\{\hat{\lambda}(0; h, \Psi) - \Lambda(0)\}^2]$ for an arbitrary origin 0. For the box kernel in two dimensions and ignoring edge correction, this expression reduces to

$$\rho^{(2)}(0) + \frac{\lambda^2}{\pi^2 h^4} \int_0^{2h} \left\{ 2h^2 \arccos\left(\frac{t}{2h}\right) - \frac{t}{2}(4h^2 - t^2)^{1/2} \right\} dK(t) + \lambda \frac{1 - 2\lambda K(h)}{\pi h^2}.$$

To implement this bandwidth selection method, one needs an estimator of the constant intensity $\lambda > 0$ of Ψ, an estimator $\hat{K}$ of Ripley's K-function (Chiu et al., 2013, § 4.7), a Riemann integral over the bandwidth range and an optimization algorithm. The estimator $\hat{K}$ requires a pattern with at least two points as well as some edge correction which limits the range of h-values one can consider.

Likelihood crossvalidation (Baddeley et al., 2015, § 6.5; Loader, 1999, § 5.3) ignores all interaction and assumes that Ψ is an inhomogeneous Poisson process. Then the leave-one-out crossvalidation loglikelihood equals, up to an h-independent normalization constant,

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

Conditions are needed to ensure that $\hat{\lambda}$ is strictly positive, and one must specify which, if any, edge correction is used. Simulations suggest that the clearest optimum is found when ignoring edge effects. From a numerical point of view, to implement this bandwidth selection method, one needs to discretize the observation window into a lattice and at each lattice point calculate a kernel estimator for every h considered by an optimization algorithm. The data pattern must contain at least two points.

The results of our simulation study are described in full in the Supplementary Material. Here we only present results when Ψ is a log-Gaussian Cox process (Møller et al., 1998). Thus, let Z be a Gaussian random field on W with mean zero and covariance function $\sigma^2 \exp(-\beta \|x - y\|)$, $x, y \in W$, where $\| \cdot \|$ denotes the Euclidean norm, and consider the random measure defined by its random intensity function $\eta(x) \exp\{Z(x)\}$. Then the intensity function of the resulting Cox process is $\eta(x) \exp(\sigma^2/2)$.

Given parameters β and σ^2 and a function η, we generate 100 simulations in the window $W = [0, 1]^2$. For each of the three bandwidth selection approaches considered, we select the bandwidth using no edge correction, with a discretization of 128 values in the range $[0 \cdot 01, 1 \cdot 5]$. For the crossvalidation method, we use a spatial discretization of $[0, 1]^2$ in a 128×128 grid for the numerical evaluation of the integral.

To assess the quality of the selection approaches by means of (6), the average integrated squared error over the 100 samples is calculated for each method, where for each sample we use a Gaussian kernel

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

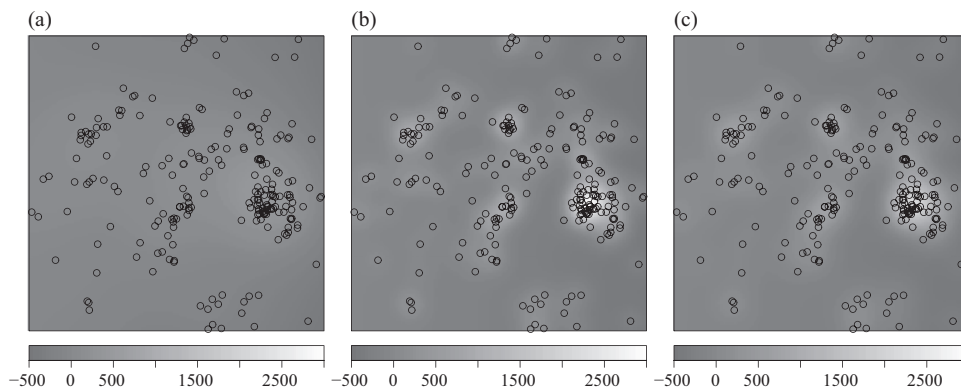| Model/Method	New	State estimation	Crossvalidation
Log-Gaussian Cox process with linear trend			
$(\sigma^2, \beta) = (2\log 5, 50)$	89·6	1477·2	536·0
$(\sigma^2, \beta) = (2\log 2, 10)$	57·5	136·9	112·6
$(\sigma^2, \beta) = (2\log 5, 10)$	335·3	2960·6	2251·2
Modulated log-Gaussian Cox process			
$(\sigma^2, \beta) = (2\log 5, 50)$	16·3	93·6	21·2
$(\sigma^2, \beta) = (2\log 2, 10)$	9·1	29·4	11·3
$(\sigma^2, \beta) = (2\log 5, 10)$	78·4	730·7	392·9

Fig. 1. Estimation error plots, $\hat{\lambda}(x, y; h) - \lambda(x, y)$, $(x, y) \in [0, 1]^2$, for a realization of the linear trend model with $(\sigma^2, \beta) = (2\log 5, 10)$. The spatial locations are superimposed. (a) New method, $h = 0·127$. (b) State estimation, $h = 0·035$. (c) Crossvalidation, $h = 0·045$.

with the selected bandwidths and then apply local edge correction. To express the results on a comparable scale, we divide them by the expected number of points in $[0, 1]^2$. Calculations were carried out using the R package spatstat (Baddeley et al., 2015). The state estimation approach is the fastest of the methods considered, the new approach is slightly slower, and the Poisson likelihood approach is the slowest.

In our first experiment, we took the linear trend function

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

and considered 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 expected number of points in $[0, 1]^2$ is then $50\exp(\sigma^2/2)$. In a second experiment we replaced the linear trend function by the modulated function

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

The expected number of points in $[0, 1]^2$ in this case is $(10 + \sin(10)/5)\exp(\sigma^2/2)$.

Table 1 shows that the new method performs best and the state estimation approach worst. Increasing β, that is, decreasing the range of interaction, tends to lead to a smaller integrated squared error. Increasing the variability σ^2, and hence the intensity, yields a larger integrated squared error. For illustration, in Fig. 1, for each of the methods we provide estimation error plots, $\hat{\lambda}(x, y; h) - \lambda(x, y)$, $(x, y) \in [0, 1]^2$, based on one of the realizations of the linear trend model with $(\sigma^2, \beta) = (2\log 5, 10)$. The new method yields a significantly smaller estimation error than its competitors.

In our simulation study we considered a range of models exhibiting clustering as well as regularity. The following tentative conclusions can be drawn. For clustered patterns, everything points to the new method

performing best. For Poisson processes, as one would expect, likelihood-based crossvalidation seems to perform better than its competitors, at least for moderately sized patterns. The picture is more varied for regular patterns. Broadly speaking, both the new and the likelihood-based methods give good results for low and moderate intensity values; state estimation seems suited for very dense patterns. We generally suggest sticking to the new approach, unless (i) the pattern exhibits clear inhibition/regularity, or (ii) the number of points is very large and the pattern clearly exhibits no clustering. When exception (i) holds we suggest the likelihood-based approach and when exception (ii) holds the state estimation approach.

5. DISCUSSION

To deal with zero or near-zero intensities, one may consider a further point process $\Xi \subseteq W$, independent of Ψ, with known intensity function $\lambda_\Xi(x) > 0$. Since the superposition $\Psi \cup \Xi$ has intensity function $\lambda_{\Psi \cup \Xi}(x) = \lambda(x) + \lambda_\Xi(x) > 0$, $x \in W$ (Chiu et al., 2013, p. 165), by employing (8) to obtain an estimate $\hat{\lambda}_{\Psi \cup \Xi}(\,\cdot\,; h)$ based on $(\Psi \cup \Xi) \cap W$, we may subtract $\lambda_\Xi(\cdot)$ to obtain the final estimate $\hat{\lambda}(\,\cdot\,; h)$. Some care must be taken when choosing Ξ. If λ_Ξ is too small the superposition will have no effect. If λ_Ξ is too large the features of Ψ may not be detectable through $\Psi \cup \Xi$.

One may also exploit the Campbell formula to select the bandwidth as a minimizer $h > 0$ of the squared difference between $\sum_{x \in \Psi \cap W} f\{\hat{\lambda}(x; h, \Psi, W)\}$ and $\int_W f\{\hat{\lambda}(x; h, \Psi, W)\}\hat{\lambda}(x; h, \Psi, W)\,\mathrm{d}x$, for suitable functions f other than the current choice $f(x) = 1/x$ motivated by the ratio estimators in Cronie & van Lieshout (2016).

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes an extended simulation study.

REFERENCES

BADDELEY, A. J., MØLLER, J. & WAAGEPETERSEN, R. P. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neer.* **54**, 329–50.

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

BARR, C. D. & SCHOENBERG, F. P. (2010). On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. *Biometrika* **97**, 977–84.

BERMAN, M. & DIGGLE, P. J. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *J. R. Statist. Soc. B* **51**, 81–92.

BERNARDEAU, F. & VAN DE WEYGAERT, R. (1996). A new method for accurate estimation of velocity field statistics. *Mon. Not. R. Astron. Soc.* **279**, 693–711.

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

CRONIE, O. & VAN LIESHOUT, M. N. M. (2016). Summary statistics for inhomogeneous marked point processes. *Ann. Inst. Statist. Math.* **68**, 905–28.

DALEY, D. J. & VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. New York: Springer, 2nd ed.

DIGGLE, P. J. (1985). A kernel method for smoothing point process data. *Appl. Statist.* **34**, 138–47.

DIGGLE, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Boca Raton, Florida: Taylor & Francis/CRC Press, 3rd ed.

GELFAND, A. E., DIGGLE, P. J., FUENTES, M. & GUTTORP, P. E. (2010). *Handbook of Spatial Statistics*. Boca Raton, Florida: Taylor & Francis/CRC Press.

HALL, P., MINNOTTE, M. C. & ZHANG, C. (2004). Bump hunting with non-Gaussian kernels. *Ann. Statist.* **32**, 2124–41.

Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley.

Loader, C. (1999). *Local Regression and Likelihood*. New York: Springer.

Matthes, K., Kerstan, J. & Mecke, J. (1978). *Infinitely Divisible Point Processes*. Chichester: Wiley.

Møller, J., Syversveen, A. & Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scand. J. Statist.* **25**, 451–82.

Møller, J. & Waagepetersen, R. (2017). Some recent developments in statistics for spatial point patterns. *Ann. Rev. Statist. Appl.* **4**, 317–42.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402.

Ord, J. K. (1978). How many trees in a forest? *Math. Sci.* **3**, 23–33.

Schaap, W. E. & van de Weygaert, R. (2000). Letter to the editor. Continuous fields and discrete samples: Reconstruction through Delaunay tessellations. *Astron. Astrophys.* **363**, L29–L32.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. New York: Wiley.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Stoyan, D. & Grabarnik, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Math. Nachr.* **151**, 95–100.

van Lieshout, M. N. M. (2000). *Markov Point Processes and Their Applications*. London/Singapore: Imperial College Press/World Scientific.

van Lieshout, M. N. M. (2011). A J-function for inhomogeneous point processes. *Statist. Neer.* **65**, 183–201.

van Lieshout, M. N. M. (2012). On estimation of the intensity function of a point process. *Methodol. Comp. Appl. Prob.* **14**, 567–78.

[*Received on* 22 *March* 2017. *Editorial decision on* 12 *December* 2017]