



Infill Asymptotics and Bandwidth Selection for Kernel Estimators of Spatial Intensity Functions

M. N. M. van Lieshout^{1,2} 

Received: 6 May 2019 / Revised: 10 October 2019 / Accepted: 21 October 2019 /
Published online: 28 November 2019
© The Author(s) 2019

Abstract

We investigate the asymptotic mean squared error of kernel estimators of the intensity function of a spatial point process. We derive expansions for the bias and variance in the scenario that n independent copies of a point process in \mathbb{R}^d are superposed. When the same bandwidth is used in all d dimensions, we show that an optimal bandwidth exists and is of the order $n^{-1/(d+4)}$ under appropriate smoothness conditions on the true intensity function.

Keywords Bandwidth · Infill asymptotics · Intensity function · Kernel estimator · Mean squared error · Point process

Mathematics Subject Classification (2010) 60G55 · 62G07 · 60D05

1 Introduction

Often the first step in the analysis of a spatial point pattern is to estimate its intensity function. Various non-parametric estimators are available to do so. Some techniques are based on local neighbourhoods of a point, expressed for example in terms of its nearest neighbours (Granville 1998) or in terms of its cells in the Voronoi (Ord 1978) or Delaunay tessellation (Schaap 2007; Schaap and Van de Weygaert 2000) of the pattern. By far the most popular technique, however, is kernel smoothing (Diggle 1985). Specifically, let Φ be a simple point process that is observed in a bounded open subset $W \neq \emptyset$ of \mathbb{R}^d and assume that its first order moment measure exists as a σ -finite Borel measure and is absolutely continuous with respect to Lebesgue measure with a Radon–Nikodym derivative $\lambda : \mathbb{R}^d \rightarrow [0, \infty)$ known as its intensity function. Heuristically speaking $\lambda(x_0) dx_0$ can be interpreted as the

This research was partially supported by The Netherlands Organisation for Scientific Research NWO (project DEEP.NL.2018.033).

✉ M. N. M. van Lieshout
Marie-Colette.van.Lieshout@cwi.nl

¹ CWI, P.O. Box 94079, NL-1090 GB, Amsterdam, The Netherlands

² University of Twente, P.O. Box 217, NL-7500 AE, Enschede, The Netherlands

infinitesimal probability that Φ places a point in dx_0 . A kernel estimator of λ based on $\Phi \cap W$ then takes the form

$$\widehat{\lambda}(x_0; H) = \lambda(x_0; \widehat{H}, \Phi, W) = \frac{1}{\det(H)} \sum_{y \in \Phi \cap W} \kappa(H^{-1}(x_0 - y)) \quad (1)$$

at $x_0 \in W$. The function $\kappa : \mathbb{R}^d \rightarrow [0, \infty)$ is supposed to be kernel, that is, a d -dimensional symmetric probability density function (Silverman 1986, p. 13) and $H = \text{diag}(h_1, \dots, h_d)$ is a diagonal matrix with entries $h_i > 0$, $i = 1, \dots, d$. The choices of the bandwidths h_i determine the amount of smoothing per component. As an aside, note that the support of $\kappa(H^{-1}(x_0 - y))$ as a function of y could overlap the complement of W . For this reason various edge corrections have been proposed (Berman and Diggle 1989; Van Lieshout 2012). In the sequel, though, we shall be concerned with very small bandwidths so this aspect may be ignored.

The aim of this paper is to derive asymptotic expansions for the bias and variance of Eq. 1 in terms of the bandwidth matrix H . Such expansions are well known in the superficially similar case of estimating a probability density function based on a random sample. Indeed, there is a vast literature on this topic that is summarised in the textbooks by Bowman and Azzalini (1997), by Silverman (1986) or by Wand and Jones (1994) and the references therein. However, intensity function estimation is different in three respects: λ is not normalised, the number of points in $\Phi \cap W$ is random and their locations are not necessarily independent. In this spatial context, bandwidth selection is dominated by ad hoc (Berman and Diggle 1989) and non-parametric methods (Cronie and Van Lieshout 2018). Rigorous techniques, to the best of our knowledge, only exist for point processes that consist of independent and identically distributed points. For example, assuming a simple multiplicative model on the intensity function of a Poisson process on the real line, Brooks and Marron (1991) derived an asymptotically optimal least-squares cross-validation estimator. Lo (2017) studied the asymptotic (integrated) mean squared error for binomial point processes in any dimension without imposing specific model assumptions on λ . Note that the inherent conditioning on the number of points serves to reduce the problem of intensity estimation to that of multivariate density estimation. Bootstrap ideas can be used to construct confidence regions as proposed in one dimension by Cowling et al. (1996) and extended to \mathbb{R}^d by Fuentes–Santos et al. (2016). Our goal is to extend Lo's approach to point processes that do not consist of a given number of independent and identically distributed points.

The plan of this paper is as follows. In Section 2 we focus on the regime in which n independent copies Φ_i of the same point process are superposed and the bandwidths $h_{i,n}$ tend to zero at the same rate b_n as n tends to infinity. Note that the $h_{i,n}$ may depend on the location, x_0 , of interest but not on the points of the pattern Φ . Adaptive bandwidth selection (Abramson 1982) in which the bandwidths may depend on the Φ_i will be treated in a companion paper. We derive expansions for the bias and variance and deduce the asymptotically optimal bandwidth when the diagonal entries $h_{i,n} = h_n$ are identical. In the general case we show by counterexample that an asymptotically optimal bandwidth matrix may not exist. For the sake of readability, all proofs are deferred to Section 3.

2 Infill Asymptotics

Let Φ_1, Φ_2, \dots be independent and identically distributed simple point processes (Chiu et al. 2013) restricted to a non-empty bounded open set $W \subset \mathbb{R}^d$ for which the first order

moment measure exists, is locally finite and admits an intensity function $\lambda : W \rightarrow [0, \infty)$. For $n \in \mathbb{N}$, let

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

denote the union. Upon taking the limit for $n \rightarrow \infty$, one obtains an asymptotic regime known as ‘infill asymptotics’ (Ripley 1988). Since the Φ_i share the same intensity function, the intensity function of their union Y_n is $n\lambda$. Therefore $\lambda(x_0), x_0 \in W$, may be estimated by

$$\widehat{\lambda}_n(x_0) := \frac{\lambda(x_0; \widehat{H}, Y_n, W)}{n} = \frac{1}{n} \sum_{i=1}^n \lambda(x_0; \widehat{H}, \Phi_i, W) \tag{2}$$

where $\lambda(x_0; \widehat{H}, \Phi_i, W)$ is given by Eq. 1.

The next Lemma collects the first two moments of Eq. 1.

Lemma 1 *Let Φ be a simple point process observed in a bounded open subset $\emptyset \neq W \subset \mathbb{R}^d$ whose factorial moment measures exist up to second order as locally finite Borel measures and are absolutely continuous with intensity function λ and second order product density $\rho^{(2)}$. Let κ be a kernel and H a diagonal matrix with positive entries. Then the first two moments of Eq. 1 are*

$$\mathbb{E} \left[\lambda(x_0; \widehat{H}, \Phi, W) \right] = \frac{1}{\det(H)} \int_W \kappa \left(H^{-1}(x_0 - u) \right) \lambda(u) du$$

and

$$\begin{aligned} \mathbb{E} \left[\left(\lambda(x_0; \widehat{h}, \Phi, W) \right)^2 \right] &= \frac{1}{\det(H)^2} \int_W \int_W \kappa \left(H^{-1}(x_0 - u) \right) \kappa \left(H^{-1}(x_0 - v) \right) \\ &\quad \times \rho^{(2)}(u, v) du dv \\ &\quad + \frac{1}{\det(H)^2} \int_W \kappa \left(H^{-1}(x_0 - u) \right)^2 \lambda(u) du. \end{aligned}$$

The proof follows directly from the definition of product densities in, for example, Section 4.3.3 in Chiu et al. (2013). Provided that λ takes values in $(0, \infty)$ and the first moment is finite, the variance of $\lambda(x_0; \widehat{H}, \Phi, W)$ can be expressed in terms of the pair correlation function $g : W \times W \rightarrow [0, \infty)$, defined by

$$g(u, v) = \frac{\rho^{(2)}(u, v)}{\lambda(u)\lambda(v)},$$

as

$$\begin{aligned} \frac{1}{\det(H)^2} \int_{W \times W} \kappa \left(H^{-1}(x_0 - u) \right) \kappa \left(H^{-1}(x_0 - v) \right) (g(u, v) - 1) \lambda(u) \lambda(v) du dv \\ + \frac{1}{\det(H)^2} \int_W \kappa \left(H^{-1}(x_0 - u) \right)^2 \lambda(u) du. \end{aligned}$$

For Poisson processes, the first integral vanishes as $g \equiv 1$.

In this paper, we will restrict ourselves to kernels that belong to the Beta class:

$$\kappa^\gamma(x) = \frac{\Gamma(d/2 + \gamma + 1)}{\pi^{d/2} \Gamma(\gamma + 1)} (1 - x^T x)^\gamma 1\{x \in b(0, 1)\}, \quad x \in \mathbb{R}^d, \tag{3}$$

for $\gamma \geq 0$. Here $b(0, 1)$ is the closed unit ball in \mathbb{R}^d centred at the origin. The normalising constant will be abbreviated by

$$c(d, \gamma) = \int_{b(0,1)} (1 - x^T x)^\gamma dx = \frac{\pi^{d/2} \Gamma(\gamma + 1)}{\Gamma(d/2 + \gamma + 1)}, \quad d \in \mathbb{N}, \gamma \geq 0. \tag{4}$$

Note that Beta kernels are supported on the closed unit ball and that their smoothness is governed by the parameter γ . Indeed, the box kernel defined by $\gamma = 0$ is constant and therefore continuous on the interior of the unit ball; the Epanechnikov kernel corresponding to the choice $\gamma = 1$ is Lipschitz continuous and for $\gamma > k \in \mathbb{N}$ the function κ^γ is k times continuously differentiable on \mathbb{R}^d .

The following Lemma collects further basic properties of the Beta kernels. The proof can be found in Section 3.1.

Lemma 2 *For the Beta kernels κ^γ , $\gamma \geq 0$, defined in Eq. 3 the integrals*

$$\int_{\mathbb{R}} x_i \kappa^\gamma(x) dx_i = 0 = \int_{b(0,1)} x_i x_j \kappa^\gamma(x) dx_1 \cdots dx_d$$

vanish for all $i, j \in \{1, \dots, d\}$ such that $i \neq j$. Furthermore, with c defined in Eq. 4,

$$Q(d, \gamma) := \int_{\mathbb{R}^d} \kappa^\gamma(x)^2 dx = \frac{c(d, 2\gamma)}{c(d, \gamma)^2}$$

is finite and so is, for all $i = 1, \dots, d$,

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

For the important special case $d = 2$,

$$Q(2, \gamma) = \frac{(\gamma + 1)^2}{(2\gamma + 1)\pi}.$$

At this point it is important to stress that the restriction to Beta kernels is made for specificity only. Our results extend with minor modifications (different expressions for Q and V) to any non-negative squared integrable function κ that is compactly supported, integrates to one, has finite and positive second order moments and is symmetric in the sense that, for all $x \in \mathbb{R}^d$, $\kappa(x) = \kappa(-x)$ and the first displayed formula in Lemma 2 holds (Silverman 1986).

Lemma 1 can be used to derive the mean squared error of Eq. 2. For the proof we refer to Section 3.2.

Assumption A Let Φ_1, Φ_2, \dots be independent and identically distributed simple point processes observed in a bounded open subset $\emptyset \neq W \subset \mathbb{R}^d$. Assume that their factorial moment measures exist up to second order as locally finite Borel measures and are absolutely continuous with strictly positive intensity function $\lambda : W \rightarrow (0, \infty)$ and second order product densities $\rho^{(2)}$. Write $Y_n = \bigcup_{i=1}^n \Phi_i$ for the union, $n \in \mathbb{N}$.

Proposition 1 *Let κ^γ be a Beta kernel (3) with $\gamma \geq 0$. Then under Assumption A the mean squared error of Eq. 2 is given by*

$$\begin{aligned} \text{mse } \widehat{\lambda}_n(x_0) &= \left(\frac{1}{\det(H)} \int_W \kappa^\gamma \left(H^{-1}(x_0 - u) \right) \lambda(u) \, du - \lambda(x_0) \right)^2 \\ &+ \frac{1}{n \det(H)^2} \int_W \int_W \kappa^\gamma \left(H^{-1}(x_0 - u) \right) \kappa^\gamma \left(H^{-1}(x_0 - v) \right) \\ &\times (g(u, v) - 1) \lambda(u) \lambda(v) \, du \, dv \\ &+ \frac{1}{n \det(H)^2} \int_W \kappa^\gamma \left(H^{-1}(x_0 - u) \right)^2 \lambda(u) \, du. \end{aligned} \tag{5}$$

If $h_1 = \dots = h_d = h$ then Eq. 5 reduces to

$$\begin{aligned} \text{mse } \widehat{\lambda}_n(x_0) &= \left(\frac{1}{h^d} \int_{b(x_0, h) \cap W} \kappa^\gamma \left(\frac{x_0 - u}{h} \right) \lambda(u) \, du - \lambda(x_0) \right)^2 \\ &+ \frac{1}{nh^{2d}} \int_{(b(x_0, h) \cap W)^2} \kappa^\gamma \left(\frac{x_0 - u}{h} \right) \kappa^\gamma \left(\frac{x_0 - v}{h} \right) \\ &\times (g(u, v) - 1) \lambda(u) \lambda(v) \, du \, dv \\ &+ \frac{1}{nh^{2d}} \int_{b(x_0, h) \cap W} \kappa^\gamma \left(\frac{x_0 - u}{h} \right)^2 \lambda(u) \, du. \end{aligned}$$

The first term in Eq. 5 is the squared bias. It depends on λ and on the bandwidth matrix H but not on n . The remaining terms come from the variance and depend on λ , on g , on H and on n . Note that λ and g are unknown. Therefore, if one were to use Eq. 5 for selecting a bandwidth, these quantities would have to be estimated. Moreover, Eq. 5 involves integrals that would have to be approximated numerically. Therefore, our aim in the remainder of this section is to derive an asymptotic expansion for the mean squared error for bandwidths $h_{i,n}$ that depend on n in such a way that for all components $i = 1, \dots, d$, $h_{i,n} \rightarrow 0$ as $n \rightarrow \infty$. It will turn out that the leading terms no longer depend on the pair correlation function and do not involve integrals that cannot be evaluated explicitly.

First recall some basic facts from analysis. Let E be an open subset of \mathbb{R}^d and denote by $C^k(E)$ the class of functions $f : E \rightarrow \mathbb{R}^m$ for which all k^{th} order partial derivatives $D_{j_1 \dots j_k} f$ exist and are continuous on E . For such functions the order of taking partial derivatives may be interchanged and the Taylor theorem states that if $x \in E$ and $x + th \in E$ for all $0 \leq t \leq 1$ then a $\theta \in (0, 1)$ can be found such that

$$f(x + h) - f(x) = \sum_{r=1}^{k-1} \frac{1}{r!} D^r f(x)(h^{(r)}) + \frac{1}{k!} D^k f(x + \theta h)(h^{(k)}). \tag{6}$$

Here $h^{(r)}$ is the r -tuple (h, \dots, h) and

$$D^r f(x)(h^{(r)}) := \sum_{j_1, \dots, j_r=1}^d h_{j_1} \cdots h_{j_r} D_{j_1 \dots j_r} f(x)$$

for $h = (h_1, \dots, h_d) \in \mathbb{R}^d$.

We are now ready to state the main result of this section, generalising Theorem 2 in Lo (2017) for the union of independent random points. The proof can be found in

Section 3.2. We will make the following assumptions on the asymptotic regime and the factorial moment measures of the point processes.

Assumption B Let $H_n, n \in \mathbb{N}$, be a sequence of diagonal matrices with entries $h_{i,n} > 0$ for $i = 1, \dots, d$. Suppose that, as $n \rightarrow \infty, h_{i,n}/b_n \rightarrow \beta_i$ for some constants $0 < \beta_i < \infty$ and a sequence of positive numbers $b_n > 0$ such that $b_n \rightarrow 0$ and $nb_n^d \rightarrow \infty$.

Assumption C Additional to Assumption A suppose that the pair correlation function $g : W \times W \rightarrow [0, \infty)$ is bounded and that the intensity function $\lambda : W \rightarrow (0, \infty)$ is twice continuously differentiable with second order partial derivatives $\lambda_{ij} = D_{ij}\lambda, i, j = 1, \dots, d$, that are Hölder continuous with exponent $\alpha > 0$ on W , that is, there exists some $C > 0$ such that, for all $i, j \in \{1, \dots, d\}$,

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

Theorem 1 Under Assumptions A and C and in the asymptotic regime of Assumption B the bias and variance of the estimator (2) with Beta kernel $\kappa^\gamma, \gamma \geq 0$, satisfy

1. $\text{bias } \widehat{\lambda}_n(x_0) = \frac{\sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0)}{2(d+2\gamma+2)} + O(b_n^{2+\alpha})$
2. $\text{Var } \widehat{\lambda}_n(x_0) = \frac{\lambda(x_0) Q(d,\gamma)}{n \det(H_n)} + O\left(\frac{1}{nb_n^{d-1}}\right)$

as $n \rightarrow \infty$.

The bias expansion depends on the second order partial derivatives of the unknown intensity function and on the smoothness parameter α . The smoothness of the kernel, measured by γ , also plays a role. The leading term of the variance depends on $\lambda(x_0)$ and on the smoothness of the kernel.

Theorem 1 implies an expansion for the mean squared error, cf. Section 3.2 for a proof.

Corollary 1 Consider the setting of Theorem 1. Then, as $n \rightarrow \infty$,

$$\begin{aligned} \text{mse } \widehat{\lambda}_n(x_0) &= \frac{V(d, \gamma)^2}{4} \left(\sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0) \right)^2 + \frac{\lambda(x_0) Q(d, \gamma)}{n \det(H_n)} \\ &+ O(b_n^{4+\alpha}) + O\left(\frac{1}{nb_n^{d-1}}\right). \end{aligned}$$

If $h_{i,n} = h_n = b_n$ for $i = 1, \dots, d$ then, provided $\sum_{i=1}^d \lambda_{ii}(x_0) \neq 0$, the asymptotic mean squared error is minimal for

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

In words, $h_n^*(x_0)$ is of the order $n^{-1/(d+4)}$. Clearly $h_n^*(x_0)$ tends to zero as $n \rightarrow \infty$. Moreover, $nh_n^*(x_0)^d$ is of the order n to the $1 - d/(d+4)$ and therefore tends to infinity with

n . The expression is similar to the Parzen formula (1962) for classic density estimation. For the special case $d = 2$,

$$h_n^*(x_0) = \frac{1}{n^{1/6}} \left(\frac{8\lambda(x_0)(\gamma + 1)^2(\gamma + 2)^2}{(2\gamma + 1)\pi(\lambda_{11}(x_0) + \lambda_{22}(x_0))^2} \right)^{1/6}.$$

The optimal bandwidth $h_n^*(x_0)$ depends on the unknown intensity function and its second order partial derivatives. A simple approach would be to take a fully non-parametric pilot estimator (for example the one proposed by Cronie and Van Lieshout (2018) that is sufficiently smooth to allow taking second order partial derivatives and to plug these into the expression for $h_n^*(x_0)$. More sophisticated approaches involve iterations of kernel estimators for $\lambda(x_0)$. Again provided that the kernel is sufficiently smooth, the second order partial derivatives of these kernel estimators can be used to estimate $\lambda_{ii}(x_0)$, $i = 1, \dots, d$ (Engel et al. 1994; Lo 2017), possibly using a somewhat larger bandwidth.

Note that when the second order partial derivatives have different signs, the leading term in the expansion for the bias in Theorem 1 may vanish. If the bandwidth components may differ, a unique asymptotically optimal bandwidth matrix may not exist even when the leading bias term is non-zero. These points are illustrated in the following example.

Example 1 Take $d = 2$, $W = (0, 1)^2$, the box kernel κ^0 and suppose that $\lambda(x, y) = 2 - x^2 + y^2$ for $(x, y) \in W$. Then the leading terms in the mean squared error expansion add up to

$$\left(h_{1,n}^2 \lambda_{11}(x_0) + h_{2,n}^2 \lambda_{22}(x_0) \right)^2 \frac{V(2, \gamma)^2}{4} + \frac{\lambda(x_0) Q(2, \gamma)}{nh_{1,n}h_{2,n}},$$

which at $x_0 = (1/2, 1/2)$ reduce to

$$\frac{1}{16} \left(h_{2,n}^2 - h_{1,n}^2 \right)^2 + \frac{2}{\pi nh_{1,n}h_{2,n}}$$

by Lemma 2. For this example the score equations do not have a zero in $(0, \infty)^2$. Specialising to the case that $h_{1,n} = h_{2,n} = h_n$, note that for $h_n < 1/2$, the bias is zero (cf. Lemma 1 and Proposition 1). Hence, under the asymptotic regime of Assumption B, there can be no trade-off between bias and variance.

In two dimensions a sufficient condition for the existence of an asymptotically optimal bandwidth matrix is that $\lambda_{11}(x_0)\lambda_{22}(x_0) > 0$, cf. Section 2.2 in Lo (2017), in which case the optimal components satisfy

$$h_{2,n}^*(x_0) = h_{1,n}^*(x_0) \sqrt{\frac{\lambda_{11}(x_0)}{\lambda_{22}(x_0)}}.$$

We conclude this section with an expansion for the random variable $\widehat{\lambda}_n(x_0)$ itself. Its proof can be found in Section 3.2.

Proposition 2 *Under Assumptions A and C and in the asymptotic regime of Assumption B the estimator (2) with Beta kernel κ^γ , $\gamma \geq 0$, satisfies*

$$\widehat{\lambda}_n(x_0) = \lambda(x_0) + O(b_n^2) + O_P \left(n^{-1/2} b_n^{-d/2} \right)$$

as $n \rightarrow \infty$.

3 Proofs and Technicalities

3.1 Properties of the Beta Kernel

Proof of Lemma 2 The first two claims follow from the symmetry of the Beta kernel. Furthermore

$$Q(d, \gamma) = \int_{\mathbb{R}^d} \kappa^\gamma(x)^2 dx = \frac{1}{c(d, \gamma)^2} \int_{b(0,1)} (1 - x^T x)^{2\gamma} dx = \frac{c(d, 2\gamma)}{c(d, \gamma)^2}.$$

From the symmetry of the Beta kernel it is clear that the definition of $V(d, \gamma)$ does not depend on the choice of i . First consider the case $d = 1$. By the symmetry of κ^γ and a change of variables $v = x^2, dx = dv/(2\sqrt{v})$,

$$\begin{aligned} V(1, \gamma) &= \int_{-\infty}^{\infty} x^2 \kappa^\gamma(x) dx = \frac{2}{c(1, \gamma)} \int_0^1 v(1-v)^\gamma \frac{1}{2v^{1/2}} dv = \frac{B(\frac{3}{2}, \gamma + 1)}{c(1, \gamma)} \\ &= \frac{1}{2\gamma + 3}. \end{aligned}$$

For dimensions $d > 1$, use Fubini's theorem to write $V(d, \gamma)$ as a repeated integral and note that the innermost integral takes the form

$$\int_{\left\{ \frac{s^2}{1 - \|x\|_{d-1}^2} \leq 1 \right\}} s^2 (1 - \|x\|_{d-1}^2 - s^2)^\gamma ds.$$

By the symmetry and a change of variables $t = s^2/(1 - \|x\|_{d-1}^2)$,

$$V(d, \gamma) = \frac{B\left(\frac{3}{2}, \gamma + 1\right)}{c(d, \gamma)} c\left(d - 1, \gamma + \frac{3}{2}\right)$$

in accordance with the claim. □

3.2 Proofs

Proof of Proposition 1 Since $\widehat{\lambda}_n(x_0)$ is the average of n independent and identically distributed random variables $\lambda(x_0; H, \Phi_i, W), i = 1, \dots, n$,

$$\mathbb{E} \widehat{\lambda}_n(x_0) = \mathbb{E} \lambda(x_0; H, \Phi_1, W)$$

and

$$\text{Var} \widehat{\lambda}_n(x_0) = \frac{1}{n} \text{Var} \lambda(x_0; H, \Phi_1, W).$$

As $\text{mse} \widehat{\lambda}_n(x_0)$ is the sum of the squared bias and the variance the claim follows from Lemma 1. If the diagonal of H is constant, the fact that κ^γ is supported on $b(0, 1)$ implies that

$$\mathbb{E} \widehat{\lambda}_n(x_0) = \frac{1}{h^d} \int_{b(x_0, h) \cap W} \kappa^\gamma\left(\frac{x_0 - u}{h}\right) \lambda(u) du$$

and

$$\begin{aligned} \text{Var } \widehat{\lambda}_n(x_0) &= \frac{1}{nh^{2d}} \int_{b(x_0,h) \cap W} \int_{b(x_0,h) \cap W} \kappa^\gamma \left(\frac{x_0 - u}{h} \right) \kappa^\gamma \left(\frac{x_0 - v}{h} \right) \\ &\quad \times (g(u, v) - 1) \lambda(u) \lambda(v) du dv \\ &\quad + \frac{1}{nh^{2d}} \int_{b(x_0,h) \cap W} \kappa^\gamma \left(\frac{x_0 - u}{h} \right)^2 \lambda(u) du. \end{aligned}$$

□

Proof of Theorem 1 To prove 1. note that since each $h_{i,n}$ goes to zero, $x_0 \in W$ and W is open, for n large enough,

$$\{x \in \mathbb{R}^d : \|H_n^{-1}(x_0 - x)\| \leq 1\} \subset \prod_{i=1}^d [x_{0,i} - h_{i,n}, x_{0,i} + h_{i,n}] \subset W. \tag{7}$$

For such n , by a change of variables, the symmetry and support of the Beta kernels and the proof of Proposition 1, the bias is

$$\begin{aligned} &\frac{1}{\det(H)} \int_{\mathbb{R}^d} \kappa^\gamma \left(\frac{x_{0,1} - u_1}{h_{1,n}}, \dots, \frac{x_{0,d} - u_d}{h_{d,n}} \right) \lambda(u) du - \lambda(x_0) \\ &= \int_{b(0,1)} \kappa^\gamma(u) \{\lambda(x_0 + H_n u) - \lambda(x_0)\} du. \end{aligned} \tag{8}$$

The term $\lambda(x_0)$ can be brought under the integral since κ^γ is a probability density function.

Fix $u \in b(0, 1)$. As $x_0 + tH_n u \in W$ for all $0 \leq t \leq 1$ and λ is twice continuously differentiable on W the term between curly brackets in the integrand in Eq. 8 may be expanded as a Taylor series (6) with $k = 2$:

$$\lambda(x_0 + H_n u) - \lambda(x_0) = D^1 \lambda(x_0)(H_n u) + \frac{1}{2} D^2 \lambda(x_0 + \theta H_n u)(H_n u, H_n u)$$

for some $0 < \theta = \theta(u) < 1$ that may depend on u . Write

$$\begin{aligned} D^2 \lambda(x_0 + \theta H_n u)(H_n u, H_n u) &= D^2 \lambda(x_0 + \theta H_n u)(H_n u, H_n u) \\ &\quad - D^2 \lambda(x_0)(H_n u, H_n u) + D^2 \lambda(x_0)(H_n u, H_n u). \end{aligned}$$

Now

$$\begin{aligned} &\left| D^2 \lambda(x_0 + \theta H_n u)(H_n u, H_n u) - D^2 \lambda(x_0)(H_n u, H_n u) \right| \\ &= \left| \sum_{i=1}^d \sum_{j=1}^d h_{i,n} h_{j,n} u_i u_j (\lambda_{ij}(x_0 + \theta H_n u) - \lambda_{ij}(x_0)) \right| \end{aligned}$$

is dominated by

$$\sum_{i=1}^d \sum_{j=1}^d h_{i,n} h_{j,n} |\lambda_{ij}(x_0 + \theta H_n u) - \lambda_{ij}(x_0)|$$

because $|u_i| \leq 1$. Since n was chosen large enough for $x_0 + \theta H_n u$ to lie in W we may use the Hölder assumption to obtain the inequality

$$\begin{aligned} &\left| D^2 \lambda(x_0 + \theta H_n u)(H_n u, H_n u) - D^2 \lambda(x_0)(H_n u, H_n u) \right| \\ &\leq C \|\theta H_n u\|^\alpha \sum_{i=1}^d \sum_{j=1}^d h_{i,n} h_{j,n}. \end{aligned}$$

Write

$$\|H_n u\| = \left(\sum_{i=1}^d h_{i,n}^2 u_i^2 \right)^{1/2} = b_n \left(\sum_{i=1}^d \frac{h_{i,n}^2}{b_n^2} u_i^2 \right)^{1/2}.$$

The sum in the expression in the right hand side is uniformly bounded over $u \in b(0, 1)$ since each $h_{i,n}/b_n$ converges. Similarly

$$\sum_{i=1}^d \sum_{j=1}^d h_{i,n} h_{j,n} = b_n^2 \sum_{i=1}^d \sum_{j=1}^d \frac{h_{i,n} h_{j,n}}{b_n^2}$$

and the sum in the right hand side is bounded. In summary,

$$\left| D^2 \lambda(x_0 + \theta H_n u)(H_n u, H_n u) - D^2 \lambda(x_0)(H_n u, H_n u) \right| \leq \tilde{C} b_n^{2+\alpha}$$

for a constant \tilde{C} that does not depend on the particular choice of $u \in b(0, 1)$ nor on $\theta \in (0, 1)$. We conclude that, for n large enough,

$$\lambda(x_0 + H_n u) - \lambda(x_0) = D^1 \lambda(x_0)(H_n u) + \frac{1}{2} D^2 \lambda(x_0)(H_n u, H_n u) + R(H_n u)$$

for a remainder term $R(H_n u)$ that satisfies $|R(H_n u)| \leq b_n^{2+\alpha} \tilde{C}/2$ uniformly over $u \in b(0, 1)$.

Returning to the bias (8),

$$\begin{aligned} \text{bias } \widehat{\lambda_n(x_0)} &= \int_{b(0,1)} \kappa^\gamma(u) D^1 \lambda(x_0)(H_n u) du \\ &+ \frac{1}{2} \int_{b(0,1)} \kappa^\gamma(u) D^2 \lambda(x_0)(H_n u, H_n u) du \\ &+ \int_{b(0,1)} \kappa^\gamma(u) R(H_n u) du. \end{aligned}$$

By Lemma 2,

$$\int_{b(0,1)} \kappa^\gamma(u) D^1 \lambda(x_0)(H_n u) du = \sum_{i=1}^d h_{i,n} D_i \lambda(x_0) \int_{b(0,1)} u_i \kappa^\gamma(u) du = 0.$$

Furthermore,

$$\frac{1}{2} \int_{b(0,1)} \kappa^\gamma(u) D^2 \lambda(x_0)(H_n u, H_n u) du$$

is equal to

$$\frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d h_{i,n} h_{j,n} \lambda_{ij}(x_0) \int_{b(0,1)} u_i u_j \kappa^\gamma(u) du = \frac{1}{2} \sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0) V(d, \gamma)$$

because, again by Lemma 2, the cross terms with $i \neq j$ are zero. Finally, since κ^γ is a probability density function, for n large enough,

$$\left| \int_{b(0,1)} \kappa^\gamma(u) R(H_n u) du \right| \leq \frac{\tilde{C}}{2} b_n^{2+\alpha}$$

and 1. is proved.

To prove 2. note that, as for the bias, n may be chosen large enough for the inclusions in Eq. 7 to hold. For such n , by a change of variables $u = H_n^{-1}(x - x_0)$ and the symmetry and support of the Beta kernels,

$$\begin{aligned} & \frac{1}{n \det(H_n)^2} \int_{\mathbb{R}^d} \kappa^\gamma \left(H_n^{-1}(x_0 - x) \right)^2 \lambda(x) dx \\ &= \frac{1}{n \det(H_n)} \int_{b(0,1)} \kappa^\gamma(u)^2 \lambda(x_0 + H_n u) du. \end{aligned}$$

Fix $u \in b(0, 1)$. As $x_0 + tH_n u \in W$ for all $0 \leq t \leq 1$ and λ is continuously differentiable on W we may use the Taylor expansion (6) with $k = 1$ to write

$$\begin{aligned} \lambda(x_0 + H_n u) &= \lambda(x_0) + D^1 \lambda(x_0 + \theta H_n u)(H_n u) \\ &= \lambda(x_0) + \sum_{i=1}^d h_{i,n} D_i \lambda(x_0 + \theta H_n u) u_i \end{aligned}$$

for some $0 < \theta = \theta(u) < 1$ that may depend on u . Thus we may write

$$\lambda(x_0 + H_n u) = \lambda(x_0) + R(H_n u) \tag{9}$$

for a remainder term

$$R(H_n u) = \sum_{i=1}^d h_{i,n} D_i \lambda(x_0 + \theta H_n u) u_i = b_n \sum_{i=1}^d \frac{h_{i,n}}{b_n} D_i \lambda(x_0 + \theta H_n u) u_i.$$

Since the partial derivatives are continuous and hence bounded on compact sets contained in W , $h_{i,n}/b_n \rightarrow \beta_i > 0$ and $|u_i| \leq 1$ on $b(0, 1)$ we see that $|R(H_n u)| \leq \tilde{D} b_n$ for some constant \tilde{D} and consequently

$$\begin{aligned} & \frac{1}{n \det(H_n)} \int_{b(0,1)} \kappa^\gamma(u)^2 \lambda(x_0 + H_n u) du \\ &= \frac{1}{n \det(H_n)} \lambda(x_0) Q(d, \gamma) + \frac{1}{n \det(H_n)} \int_{b(0,1)} \kappa^\gamma(u)^2 R(H_n u) du \end{aligned}$$

by Lemma 2. The bound on the remainder term $R(H_n u)$ implies that

$$\begin{aligned} \left| \frac{1}{n \det(H_n)} \int_{b(0,1)} \kappa^\gamma(u)^2 R(H_n u) du \right| &\leq \frac{1}{n \det(H_n)} \int_{b(0,1)} \kappa^\gamma(u)^2 |R(H_n u)| du \\ &\leq \frac{b_n \tilde{D}}{n \det(H_n)} Q(d, \gamma). \end{aligned}$$

Now

$$\frac{b_n}{n \det(H_n)} = \frac{b_n}{n b_n^d \prod_{i=1}^d (h_{i,n}/b_n)} = O\left(\frac{1}{n b_n^{d-1}}\right)$$

because $h_{i,n}/b_n \rightarrow \beta_i$ for some $\beta_i > 0$ as $n \rightarrow \infty$. We arrive at

$$\begin{aligned} & \frac{1}{n \det(H_n)^2} \int_W \kappa^\gamma \left(H_n^{-1}(x_0 - x) \right)^2 \lambda(x) dx \\ &= \frac{\lambda(x_0) Q(d, \gamma)}{n \det(H_n)} + O\left(\frac{1}{n b_n^{d-1}}\right) \end{aligned} \tag{10}$$

as $n \rightarrow \infty$.

We will now show that the contribution of the interaction structure (via the pair correlation function) to the variance is negligible. Again choose n so large that the inclusions

in Eq. 7 hold. Then, by a change of variables and the symmetry and support of the Beta kernels, the double integral in Eq. 5 in Proposition 1 reduces to

$$\frac{1}{n} \int_{b(0,1)^2} \kappa^\gamma(u) \kappa^\gamma(v) \times (g(x_0 + H_n u, x_0 + H_n v) - 1) \lambda(x_0 + H_n u) \lambda(x_0 + H_n v) du dv.$$

Since the pair correlation function is assumed to be bounded on W , say $g(\cdot, \cdot) \leq \bar{g}$, and $x_0 + H_n u \in W$ for all $u \in b(0, 1)$ the double integral can be bounded in absolute value by

$$\begin{aligned} & \frac{1 + \bar{g}}{n} \left(\int_{b(0,1)} \kappa^\gamma(u) \lambda(x_0 + H_n u) du \right)^2 \\ &= \frac{1 + \bar{g}}{n} \left(\int_{b(0,1)} \kappa^\gamma(u) \{ \lambda(x_0) + R(H_n u) \} du \right)^2, \end{aligned}$$

cf. Eq. 9. The integrand in the right hand side is bounded in absolute value by $\kappa^\gamma(u) \{ \lambda(x_0) + \bar{D}b_n \}$ and therefore the interaction structure contributes $O(1/n)$ to the mean squared error. Upon adding Eq. 10, we conclude that

$$\text{Var } \widehat{\lambda}_n(x_0) = \frac{\lambda(x_0) Q(d, \gamma)}{n \det(H_n)} + O\left(\frac{1}{nb_n^{d-1}}\right) + O\left(\frac{1}{n}\right)$$

as $n \rightarrow \infty$. To complete the proof note that the last term in the right hand side is negligible or of the same order compared to the middle one. \square

Proof of Corollary 1 By Theorem 1 the squared bias reads

$$\left(\frac{\sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0)}{2(d + 2\gamma + 2)} \right)^2 + 2R(b_n) \frac{\sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0)}{2(d + 2\gamma + 2)} + R(b_n)^2 \tag{11}$$

for a remainder term $R(b_n)$ for which there exists a scalar M such that $|R(b_n)| \leq Mb_n^{2+\alpha}$ for large n . Because

$$R(b_n) \sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0) = b_n^2 R(b_n) \sum_{i=1}^d \frac{h_{i,n}^2}{b_n^2} \lambda_{ii}(x_0)$$

and $h_{i,n}^2/b_n^2 \rightarrow \beta_i^2 > 0$, $n \rightarrow \infty$, the second term in Eq. 11 is $O(b_n^2 b_n^{2+\alpha}) = O(b_n^{4+\alpha})$. The third term $R(b_n)^2$ is $O(b_n^{4+2\alpha})$ and therefore negligible. Hence

$$(\text{bias } \widehat{\lambda}(x_0))^2 = \left(\frac{\sum_{i=1}^d h_{i,n}^2 \lambda_{ii}(x_0)}{2(d + 2\gamma + 2)} \right)^2 + O(b_n^{4+\alpha})$$

as $n \rightarrow \infty$ and the claimed expression for the mean squared error follows from Theorem 1.

If we restrict ourselves to the case that the entries on the diagonal of H_n are equal, the asymptotic mean squared error takes the form

$$\alpha h_n^4 + \frac{\beta}{n h_n^d}$$

for some scalars α and β that are strictly positive under the assumption that $\sum_i \lambda_{ii}(x_0) \neq 0$. Equating the derivative with respect to h_n to zero yields the score equation

$$(h_n)^{3+d+1} = \frac{d\beta}{4n\alpha}.$$

The second derivative with respect to h_n , $12\alpha h_n^2 + d(d + 1)\beta n^{-1}h_n^{-d-2}$, is strictly positive and therefore the solution to the score equation corresponds to the unique minimum. Plugging in the expressions for α and β completes the proof. \square

Proof of Proposition 2 Since each $h_{i,n}$ goes to zero, $x_0 \in W$ and W is open, for n large enough, the inclusions in Eq. 7 hold and, using Lemma 1,

$$\widehat{\lambda}_n(x_0) - \mathbb{E}\widehat{\lambda}_n(x_0) = \widehat{\lambda}_n(x_0) - \frac{1}{\det(H_n)} \int_{\mathbb{R}^d} \kappa^\gamma \left(H_n^{-1}(x_0 - x) \right) \lambda(x) dx$$

can be written as an average of n independent random variables

$$Z_i := \lambda(x_0; \widehat{H}_n, \widehat{\Phi}_i, W) - \frac{1}{\det(H_n)} \int_{\mathbb{R}^d} \kappa^\gamma \left(H_n^{-1}(x_0 - x) \right) \lambda(x) dx$$

with $\mathbb{E}Z_i = 0$. Furthermore, by Theorem 1,

$$\text{Var} \left(\frac{1}{n} \sum_{i=1}^n Z_i \right) = \frac{\lambda(x_0) Q(d, \gamma)}{n \det(H_n)} + R(b_n)$$

for a remainder term $R(b_n)$ satisfying $nb_n^{d-1}|R(b_n)| \leq M$ for some $M > 0$ and large n . By Chebychev’s inequality, for all $\epsilon > 0$,

$$\begin{aligned} \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n Z_i \right| \geq \epsilon^{-1/2} \sqrt{\frac{\lambda(x_0) Q(d, \gamma)}{nb_n^d} \prod_{i=1}^d \beta_i^{-1}} \right) \\ \leq \epsilon \prod_{i=1}^d \beta_i \frac{nb_n^d}{\lambda(x_0) Q(d, \gamma)} \left(\frac{\lambda(x_0) Q(d, \gamma)}{n \det(H_n)} + R(b_n) \right). \end{aligned}$$

Since $b_n/h_{i,n} \rightarrow 1/\beta_i$ and $\beta_i > 0$, the upper bound tends to ϵ as $n \rightarrow \infty$ and therefore

$$\frac{1}{n} \sum_{i=1}^n Z_i = O_P \left(n^{-1/2} b_n^{-d/2} \right).$$

To finish the proof note that the bias expansion 1. in Theorem 1 implies that the bias is $O(b_n^2)$. \square

Acknowledgements We are grateful to the referee and associate editor for their careful reading of the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Abramson IA (1982) On bandwidth variation in kernel estimates – a square root law. *Ann Statist* 10:1217–1223
 Berman M, Diggle PJ (1989) Estimating weighted integrals of the second-order intensity of a spatial point process. *J R Stat Soc Ser B* 51:81–92
 Bowman AW, Azzalini A (1997) Applied smoothing techniques for data analysis. The kernel approach with S-Plus illustrations. University Press, Oxford

- Brooks MM, Marron JS (1991) Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Process Appl* 38:157–165
- Chiu SN, Stoyan D, Kendall WS, Mecke J (2013) *Stochastic geometry and its applications*, 3rd edn. Wiley, Chichester
- Cowling A, Hall P, Phillips MJ (1996) Bootstrap confidence regions for the intensity of a Poisson point process. *J Amer Statist Assoc* 91:1516–1524
- Cronie O, Van Lieshout MNM (2018) A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* 105:455–462
- Diggle PJ (1985) A kernel method for smoothing point process data. *J Appl Stat* 34:138–147
- Engel J, Herrmann E, Gasser T (1994) An iterative bandwidth selector for kernel estimation of densities and their derivatives. *J Nonparametr Statist* 4:21–34
- Fuentes-Santos I, González-Manteiga W, Mateu J (2016) Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial Poisson point processes. *Scand J Stat* 43:416–435
- Granville V (1998) Estimation of the intensity of a Poisson point process by means of nearest neighbour distances. *Stat Neerl* 52:112–124
- Lo PH (2017) An iterative plug-in algorithm for optimal bandwidth selection in kernel intensity estimation for spatial data. PhD thesis, Technical University of Kaiserslautern
- Ord JK (1978) How many trees in a forest? *Math Sci* 3:23–33
- Parzen E (1962) On estimation of a probability density function and mode. *Ann Math Statist* 33:1065–1076
- Ripley BD (1988) *Statistical inference for spatial processes*. University Press, Cambridge
- Schaap WE (2007) DTFE. The Delaunay tessellation field estimator. PhD Thesis, University of Groningen
- Schaap WE, Van de Weygaert R (2000) Letter to the editor. Continuous fields and discrete samples: reconstruction through Delaunay tessellations. *Astronom Astrophys* 363:L29–L32
- Silverman BW (1986) *Density estimation for statistics and data analysis*. Chapman & Hall, Boca Raton
- Van Lieshout MNM (2012) On estimation of the intensity function of a point process. *Methodol Comput Appl Probab* 14:567–578
- Wand MP, Jones MC (1994) *Kernel smoothing*. Chapman & Hall, Boca Raton

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.