Australian & New Zealand Journal of Statistics

Aust. N. Z. J. Stat. 2021 doi: 10.1111/anzs.12319

Infill asymptotics for adaptive kernel estimators of spatial intensity

M.N.M. van Lieshout[1,2]

*CWI and University of Twente**

Summary

We apply the Abramson principle to define adaptive kernel estimators for the intensity function of a spatial point process. We derive asymptotic expansions for the bias and variance under the regime that n independent copies of a simple point process in Euclidean space are superposed. The method is illustrated by means of a simple example and applied to tornado data.

Key words: adaptive kernel estimator; bandwidth; infill asymptotics; intensity function; mean squared error; point process.

1. Introduction

The field of kernel estimation for probability densities is well-developed. Indeed, there are many textbooks including Bowman & Azzalini (1997), Chacón & Duong (2018), Silverman (1986) or Wand & Jones (1995). Briefly, an independent sample $X_1, \ldots, X_n$ from some probability density function f on $\mathbb{R}$ is observed and for every $x_0 \in \mathbb{R}$, $f(x_0)$ is estimated by

$$\hat{f}(x_0; h) = \frac{1}{nh} \sum_{i=1}^{n} \kappa\left(\frac{x_0 - X_i}{h}\right),$$

using a kernel function κ that is a symmetric probability density function. The scaling factor $h > 0$ is usually referred to as the bandwidth and governs how far the mass of κ is spread around X_i. In many cases, it is useful to let h depend on the location. Indeed, asymptotic expansions of the mean squared error of $\hat{f}(x_0; h)$ result in an expression that depends on x_0. A more dramatic improvement in terms of mean squared error can be obtained by letting h depend on the X_i too. This idea, due to Abramson (1982), leads to a bias expansion in which the leading term is of fourth order rather than quadratic (Hall, Hu & Marron 1995). The intuition is that in regions with few points, their mass should be spread out over a larger area than in regions with many points. Doing so allows the preservation of local structure in well-populated regions without imposing spurious bumps on sparser ones.

[1]CWI, Science Park 123, P.O. Box 94079, Amsterdam, NL-1090 GB, The Netherlands.
[2]Department of Applied Mathematics, University of Twente, P.O. Box 217, Enschede, 7500 AE, The Netherlands. e-mail: m.n.m.van.lieshout@cwi.nl
Acknowledgements. This research was partially supported by The Netherlands Organisation for Scientific Research NWO (project DEEP.NL.2018.033). Thanks are due to the anonymous referees for constructive comments.
*Dedicated to Adrian J. Baddeley on the occasion of his 65-th birthday.

Kernel estimation ideas are also used in the superficially similar problem of reconstructing an unknown intensity function λ from an observation of a spatial point process Φ in some bounded set W (Diggle 1985). Heuristically speaking, the intensity $\lambda(\mathbf{x_0})$ at a point $\mathbf{x_0} \in \mathbb{R}^d$ governs how likely it is that the point process places a point at $\mathbf{x_0}$. However, in contrast to f, λ is not normalised. Moreover, the number of points in $\Phi \cap W$ is random and the point locations are not necessarily independent. For these reasons, bandwidth selection in this context has been dominated by rules of thumb (see, e.g. Section 6.5 in Baddeley, Rubak & Turner (2016), Section 3.3 in Illian *et al.* (2008), or Section 6 in Scott (1992)), or ad hoc approaches that rely on specific models. Examples of the latter include the Poisson process-based likelihood cross-validation of Loader (1999) Section 5.3) or minimisation of the mean squared error in state estimation for a stationary isotropic Cox process (Diggle 1985). A fully non-parametric technique was recently proposed by Cronie & van Lieshout (2018).

Rigorous techniques, to the best of our knowledge, only exist for point processes that consist of independent and identically distributed points. Working on the real line, Brooks & Marron (1991) assumed a simple multiplicative model to derive an asymptotically optimal least-squares cross-validation estimator when the number of points tends to infinity. Lo (2017) picked up the baton and studied the asymptotic (integrated) mean squared error in any dimension without imposing a specific intensity model, again in the regime that the number of points goes to infinity. The Abramson approach for the same model was studied by Davies, Flynn & Hazelton (2018), the numerical implementation of which was provided in a follow-up paper (Davies & Baddeley 2018). Assuming that the points are i.i.d. effectively reduces intensity estimation to probability density estimation. When this is not the case, asymptotics based on the number of points in the pattern cannot be used. Since we are interested in heterogeneous patterns, that is patterns in which the intensity function is spatially varying, a similar remark holds for increasing domain asymptotics. In a companion paper (van Lieshout 2020), we generalised Lo's work to point processes that may exhibit interaction between the points. We assumed that replicated patterns were available, so that infill asymptotics apply. The aim of this paper is to study adaptive kernel estimators and prove that a large reduction in bias is obtained by doing so.

The plan is as follows. In Section 2, we recall kernel estimators for spatial intensity functions. We then explain Abramson's idea and give explicit expressions for the mean and variance of the resulting estimators. Additionally, we review the important class of compactly supported Beta kernels. In the next section, we focus on the regime in which n independent copies of the same point process are superposed and the bandwidth tends to zero as n tends to infinity. We derive Taylor expansions for the mean squared error of the associated adaptive kernel estimators, compare them to their counterparts for a fixed bandwidth and deduce the asymptotically optimal bandwidth. The approach is illustrated on a heterogeneous log-Gaussian Cox model and on tornado data in Section 4. For the sake of readability, all proofs are deferred to Section 5.

2. Adaptive kernel estimation

2.1. Kernel estimators

Let Φ be a simple point process (Chiu *et al.* 2013; van Lieshout 2019) that is observed in a bounded open subset $\emptyset \neq W$ of $\mathbb{R}^d$. Suppose that its first-order moment measure Λ, which is defined by

$$\Lambda(A) = \mathrm{E}\left[\sum_{\mathbf{x}\in\Phi} \mathbf{1}(\mathbf{x}\in A)\right],$$

the expected number of points of Φ that fall in Borel subsets A of W – exists as a locally finite Borel measure and is absolutely continuous with respect to Lebesgue measure with Radon–Nikodym derivative $\lambda : \mathbb{R}^d \to [0, \infty)$. The function λ is called the intensity function of Φ.

The classic kernel estimator of $\lambda(\mathbf{x_0})$, $\mathbf{x_0} \in W$, based on $\Phi \cap W$ is given by Diggle (1985) as

$$\frac{1}{h^d} \sum_{\mathbf{y}\in\Phi\cap W} \kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}\right). \tag{1}$$

Here $\kappa : \mathbb{R}^d \to [0, \infty)$ is a kernel, that is a d-dimensional symmetric (in the sense that it is even in all its arguments) probability density function (Silverman 1986, p. 13). The choice of bandwidth $h > 0$ determines the amount of smoothing. In this paper, we will restrict ourselves to applying the same bandwidth in every coordinate direction. More generally, a bandwidth matrix could be allowed (Chacón & Duong 2018). Since we shall be concerned with asymptotics for small bandwidths, edge effects due to mass of κ leaking out of W may safely be ignored, but various correction techniques are available in the literature including the global one proposed by Berman & Diggle (1989) or the mass preserving local edge correction suggested by van Lieshout (2012).

The kernel estimator (1) applies the same amount of smoothing everywhere. Intuitively, however, one feels that in sparse regions more smoothing might be necessary than in regions that are rich in points. Indeed, in the context of estimating a probability density function, Abramson (1982) proposed to scale the bandwidth in proportion to the square root of the probability density function.

To make his idea precise, let $c : W \to (0, \infty)$ be a measurable positive-valued weight function. Then an adaptive kernel estimator takes the form

$$\hat{\lambda}(\mathbf{x_0}; h, \Phi, W) = \sum_{\mathbf{y}\in\Phi\cap W} \frac{c(\mathbf{y})^d}{h^d} \kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h} c(\mathbf{y})\right), \quad \mathbf{x_0} \in W. \tag{2}$$

In other words, the bandwidth at $\mathbf{y} \in \Phi$ is $h/c(\mathbf{y})$. Provided λ is positive, the proposal of Abramson (1982) was to set

$$c(\mathbf{y}) = \sqrt{\frac{\lambda(\mathbf{y})}{\lambda(\mathbf{x_0})}},$$

so points $\mathbf{y}$ located in regions with a low intensity are given a larger bandwidth than those in high-intensity regions. The scaling by $\lambda(\mathbf{x_0})$ is not necessary as it can be absorbed into h but has the advantage that it makes $c(\mathbf{y})$ dimensionless. This choice is important from an asymptotic point of view, as we shall see in the sequel. Indeed, the goal of this paper is to prove that Abramson's weight function leads to a smaller asymptotic bias than would be obtained for a fixed bandwidth (Lo 2017; van Lieshout 2020). In practice, other weight functions may perform well, for example the dth degree root rule proposed by Bowman & Foster (1993).

2.2. First- and second-order moments

The moments of (2) can be expressed in terms of the factorial moment measures of Φ. Suppose that the first- and second-order factorial moment measures of Φ exist as locally finite Borel measures (Chiu *et al.* 2013, Section 4.3.3) and are absolutely continuous with respect to Lebesgue measure with Radon–Nikodym derivative $\rho^{(1)} = \lambda$ and $\rho^{(2)}$, the second-order product density. Then

$$\mathrm{E}\left[\sum_{\substack{(\mathbf{y_1},\ldots,\mathbf{y_m}):\\ \mathbf{y_i}\in\Phi, i=1,\ldots,m}}^{\neq} f(\mathbf{y_1},\ldots,\mathbf{y_m})\right] = \int_{(\mathbb{R}^d)^m} f(\mathbf{y_1},\ldots,\mathbf{y_m})\rho^{(m)}(\mathbf{y_1},\ldots,\mathbf{y_m})d\mathbf{y_1}\cdots d\mathbf{y_m}$$

for all non-negative measurable functions $f:\mathbb{R}^d\times\cdots\times\mathbb{R}^d\to\mathbb{R}$, under the proviso that the left-hand side is infinite if and only if the right-hand side is. The superscript denotes that the sum is over m-tuples consisting of distinct points.

Upon considering the functions

$$f_1(\mathbf{y}) = \frac{c(\mathbf{y})^d}{h^d}\kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}c(\mathbf{y})\right)\mathbf{1}(\mathbf{y}\in W)$$

and $f_2(\mathbf{y},\mathbf{z}) = f_1(\mathbf{y})f_1(\mathbf{z})$, we deduce that the first two moments of (2) are given by

$$\mathrm{E}\hat{\lambda}(\mathbf{x_0}; h,\Phi,W) = \frac{1}{h^d}\int_W c(\mathbf{y})^d\kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}c(\mathbf{y})\right)\lambda(\mathbf{y})d\mathbf{y}$$

and

$$\mathrm{E}\left[\left(\hat{\lambda}(\mathbf{x_0}; h,\Phi,W)\right)^2\right] = \frac{1}{h^{2d}}\int_W c(\mathbf{y})^{2d}\kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}c(\mathbf{y})\right)^2\lambda(\mathbf{y})d\mathbf{y}$$
$$+\frac{1}{h^{2d}}\int_W\int_W c(\mathbf{y})^d c(\mathbf{z})^d\kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}c(\mathbf{y})\right)\kappa\left(\frac{\mathbf{x_0}-\mathbf{z}}{h}c(\mathbf{z})\right)\rho^{(2)}(\mathbf{y},\mathbf{z})d\mathbf{y}d\mathbf{z},$$

again under the proviso that the left-hand side is infinite if and only if the right-hand side is. If the second moment is finite and $\lambda(\cdot)>0$, the variance is finite and can be expressed in terms of the pair correlation function $g(\mathbf{y},\mathbf{z}) = \rho^{(2)}(\mathbf{y},\mathbf{z})/(\lambda(\mathbf{y})\lambda(\mathbf{z}))$ as

$$\begin{aligned}\operatorname{var}\hat{\lambda}(\mathbf{x_0}; h,\Phi,W) &= \frac{1}{h^{2d}}\int_W c(\mathbf{y})^{2d}\kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}c(\mathbf{y})\right)^2\lambda(\mathbf{y})d\mathbf{y}+\\ &\frac{1}{h^{2d}}\int_W\int_W c(\mathbf{y})^d c(\mathbf{z})^d\kappa\left(\frac{\mathbf{x_0}-\mathbf{y}}{h}c(\mathbf{y})\right)\kappa\left(\frac{\mathbf{x_0}-\mathbf{z}}{h}c(\mathbf{z})\right)\big(g(\mathbf{y},\mathbf{z})-1\big)\lambda(\mathbf{y})\lambda(\mathbf{z})d\mathbf{y}d\mathbf{z}.\end{aligned} \tag{3}$$

2.3. The Beta class

For specificity, we will restrict ourselves to kernels that belong to the Beta class

$$\kappa^\gamma(\mathbf{x}) = \frac{1}{c(d,\gamma)}(1-\mathbf{x}^\top\mathbf{x})^\gamma\mathbf{1}(\mathbf{x}\in b(\mathbf{0},1)),\quad \mathbf{x}\in\mathbb{R}^d, \tag{4}$$

for $\gamma\geqslant 0$. Here $b(\mathbf{0},1)$ is the closed unit ball in $\mathbb{R}^d$ centred at the origin and

$$c(d,\gamma) = \int_{b(\mathbf{0},1)}(1-\mathbf{x}^\top\mathbf{x})^\gamma d\mathbf{x} = \frac{\pi^{d/2}\Gamma(\gamma+1)}{\Gamma(d/2+\gamma+1)},\quad d\in\mathbb{N},\gamma\geqslant 0, \tag{5}$$

cf. Duong (2015). Beta kernels are supported on the compact unit ball. The parameter γ governs the smoothness. Indeed, the box kernel defined by $\gamma = 0$ is constant and therefore continuous on the interior of the unit ball; the Epanechnikov kernel ($\gamma = 1$) is Lipschitz continuous. For $\gamma > m$ the function κ^γ is m times continuously differentiable on $\mathbb{R}^d$. Furthermore, it can be shown that

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

is finite and positive. For $i = 1,\ldots,d$, the integrals

$$V(d,\gamma) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i^2 \kappa^\gamma(\mathbf{x}) dx_1 \cdots dx_d = \frac{1}{d+2\gamma+2}$$
$$V_4(d,\gamma) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i^4 \kappa^\gamma(\mathbf{x}) dx_1 \cdots dx_d = \frac{3}{(d+2\gamma+2)(d+2\gamma+4)}$$

are finite and positive and so is, for $d \geqslant 2$ and $i \neq j \in \{1,\ldots,d\}$,

$$V_2(d,\gamma) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i^2 x_j^2 \kappa^\gamma(\mathbf{x}) dx_1 \cdots dx_d = \frac{1}{(d+2\gamma+2)(d+2\gamma+4)}.$$

In particular, for $d = 2$,

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

Derivations for $Q(d,\gamma)$ and $V(d,\gamma)$ can be found in (Duong 2015; van Lieshout 2020), the expressions for V_2 and V_4 follow by similar arguments.

3. Infill asymptotics

The asymptotic regime considered in this paper is the following. Let $\Phi_1, \Phi_2, \ldots$ be independent and identically distributed simple point processes (Chiu *et al.* 2013; van Lieshout 2019) and assume that the first-order moment measure of the Φ_i exists, is locally finite and admits an intensity function $\lambda : W \to [0,\infty)$. For $n \in \mathbb{N}$, write

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

for the union. Taking the limit for $n \to \infty$, one obtains an asymptotic regime that Ripley (1988) calls 'infill asymptotics'. Clearly the intensity function of Y_n is $n\lambda(\cdot)$ so $\lambda(\mathbf{x_0})$, $\mathbf{x_0} \in W$, may be estimated by

$$\hat{\lambda}_n(\mathbf{x_0}) = \frac{\hat{\lambda}(\mathbf{x_0}; h, Y_n, W)}{n} = \frac{1}{n} \sum_{i=1}^{n} \hat{\lambda}(\mathbf{x_0}; h, \Phi_i, W), \tag{6}$$

the average of the $\hat{\lambda}(\mathbf{x_0}; h, \Phi_i, W)$.

Our aim in the remainder of this section is to derive an asymptotic expansion of the mean squared error for bandwidths h_n that depend on n in such a way that $h_n \to 0$ as $n \to \infty$. To fix notation, let E be an open subset of $\mathbb{R}^d$ and denote by $\mathcal{C}^k(E)$ the class of functions $f : E \to \mathbb{R}$ for which the kth-order partial derivatives $D_{j_1 \cdots j_k} f$ exist and are continuous on E. For $f \in \mathcal{C}^k(E)$, the order of taking partial derivatives may be interchanged. Moreover, a

Taylor expansion holds: if $\mathbf{x} \in E$ and $\mathbf{x} + t\mathbf{h} \in E$ for all $0 \leqslant t \leqslant 1$, then a $\theta \in (0,1)$ can be found such that

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \sum_{r=1}^{k-1} \frac{1}{r!} D^r f(\mathbf{x})(h^{(r)}) + \frac{1}{k!} D^k f(\mathbf{x}+\theta\mathbf{h})(h^{(k)}), \tag{7}$$

where $\mathbf{h} = (h_1, \ldots, h_d) \in \mathbb{R}^d$ and

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

We are now ready to state the first main result of this section in analogy to the theorem in Abramson (1982, p. 1218). The proof will be given in Section 5.2.

Theorem 1. *Let $\Phi_1, \Phi_2, \ldots$ be i.i.d. simple point processes observed in a bounded open subset $\emptyset \neq W \subseteq \mathbb{R}^d$ for which the first- and second-order factorial moment measures exist, are locally finite and admit well-defined intensity function λ and pair correlation function g. Suppose that $g : W \times W \to \mathbb{R}$ is bounded and that $\lambda : W \to [\underline{\lambda}, \bar{\lambda}]$ is bounded (from below by $\underline{\lambda} > 0$, from above by $\bar{\lambda} > \underline{\lambda}$), bounded away from zero and twice continuously differentiable on W with first- and second-order partial derivatives $\lambda_i = D_i \lambda$, $\lambda_{ij} = D_{ij} \lambda$, $i, j = 1, \ldots, d$.*

Consider $\hat{\lambda}_n$ defined in (2) and (6) with $c(\mathbf{x}) = \left(\lambda(\mathbf{x})/\lambda(\mathbf{x_0})\right)^{1/2}$ for a Beta kernel κ^γ with $\gamma > 2$ and bandwidth h_n chosen in such a way that, as $n \to \infty$, $h_n \to 0$ and $nh_n^d \to \infty$. Then, for $\mathbf{x_0} \in W$, as $n \to \infty$,

(i) bias $\hat{\lambda}_n(\mathbf{x_0}) = o(h_n^2)$.
(ii) var $\hat{\lambda}_n(\mathbf{x_0}) = \lambda(\mathbf{x_0}) Q(d, \gamma)(nh_n^d)^{-1} + O\left(n^{-1} h_n^{1-d}\right)$.

A few remarks are in order. The restriction to intensity functions that are bounded away from zero may seem restrictive. In practice, however, a homogeneous Poisson process with fixed and known intensity $\epsilon > 0$ may be superposed to each of the Φ_i. Their common intensity function $\lambda + \epsilon$ is then bounded from below by ϵ and may be estimated as described before and, finally, subtraction of ϵ yields an estimator for the intensity function of interest.

Second, it is striking to observe that the variance is the same as that for a non-adaptive bandwidth (van Lieshout 2020, Theorem 1). The bias term on the other hand is of a smaller order, $o(h_n^2)$ compared to the leading term

$$h_n^2 \sum_{i=1}^{d} \frac{\lambda_{ii}(\mathbf{x_0})}{2(d + 2\gamma + 2)}$$

for a fixed bandwidth.

A disadvantage of Theorem 1 is that, as the leading bias term is not specified, it cannot be used to calculate an asymptotically optimal bandwidth. To remedy this, stronger smoothness assumptions seem needed in analogy to those of Hall *et al.* (1995) in the context of probability density estimation.

Theorem 2 *Let $\Phi_1, \Phi_2, \ldots$ be i.i.d. simple point processes observed in a bounded open subset $\emptyset \neq W \subseteq \mathbb{R}^d$ for which the first- and second-order factorial moment measures exist, are locally finite and admit well-defined intensity function λ and pair correlation function g.*

Suppose that $g : W \times W \to \mathbb{R}$ is bounded and that $\lambda : W \to [\underline{\lambda}, \bar{\lambda}]$ is bounded, bounded away from zero and four times continuously differentiable on W.

Consider $\hat{\lambda}_n$ defined in (2) and (6) with $c(\mathbf{x}) = \left(\lambda(\mathbf{x})/\lambda(\mathbf{x_0})\right)^{1/2}$ for a Beta kernel κ^γ with $\gamma > 5$ and bandwidth h_n chosen in such a way that, as $n \to \infty$, $h_n \to 0$ and $nh_n^d \to \infty$. Suppose that the fourth-order partial derivatives $c_{ijkl} = D_{ijkl}c$, $i, j, k, l = 1, \ldots, d$, are Hölder continuous with index $\alpha \in (0, 1]$ on W, that is there exists some $C > 0$ such that for all $i, j, k, l = 1, \ldots, d$:

$$|c_{ijkl}(\mathbf{x}) - c_{ijkl}(\mathbf{y})| \leqslant C \|\mathbf{x} - \mathbf{y}\|^\alpha, \quad \mathbf{x}, \mathbf{y} \in W.$$

Then, for $\mathbf{x_0} \in W$, as $n \to \infty$,

(i) $\operatorname{bias} \hat{\lambda}_n(\mathbf{x_0}) = \lambda(\mathbf{x_0}) h_n^4 \int_{\mathbb{R}^d} A(\mathbf{u}; \mathbf{x_0}) d\mathbf{u} + O(h_n^{4+\alpha})$, *for an integrable function*

$$\begin{aligned} A(\mathbf{u}; \mathbf{x_0}) = {} & \frac{g'_{\mathbf{u}}(1)}{24} D^4 c(\mathbf{x_0})(u^{(4)}) + \frac{g^{(iv)}_{\mathbf{u}}(1)}{24} (Dc(\mathbf{x_0})(u^{(1)}))^4 \\ & + \frac{g''_{\mathbf{u}}(1)}{2} \left\{ \frac{1}{3} Dc(\mathbf{x_0})(u^{(1)}) D^3 c(\mathbf{x_0})(u^{(3)}) + \frac{1}{4} (D^2 c(\mathbf{x_0})(u^{(2)}))^2 \right\} \\ & + \frac{g'''_{\mathbf{u}}(1)}{4} (Dc(\mathbf{x_0})(u^{(1)})^2 D^2 c(\mathbf{x_0})(u^{(2)}) \end{aligned}$$

with $g_{\mathbf{u}}(v) = v^{d+2} \kappa^\gamma(v\mathbf{u})$, $v \in \mathbb{R}$, having derivatives $g'_{\mathbf{u}}$, $g''_{\mathbf{u}}$, $g'''_{\mathbf{u}}$ and $g^{(iv)}_{\mathbf{u}}$ up to order four.
(ii) $\operatorname{var} \hat{\lambda}_n(\mathbf{x_0}) = \lambda(\mathbf{x_0}) Q(d, \gamma)(nh_n^d)^{-1} + O\left(n^{-1} h_n^{1-d}\right)$.

The proof will be given in Section 5.2. For the important special cases $d = 1$ and $d = 2$, the expression for $A(\mathbf{u}; \mathbf{x_0})$ may be simplified according to the following propositions. Similar results in the context of probability density estimation were obtained by Hall *et al.* (1995) in dimension one and by Davies *et al.* (2018) in dimension two. Note though that these authors did not assume Hölder continuity of the fourth-order partial derivatives and therefore obtained a remainder term of $o(h_n^4)$.

Proposition 1 *Consider the framework of Theorem 2 in one dimension ($d = 1$). Then the coefficient of h_n^4 in the expansion of bias $\hat{\lambda}_n(x_0)$ is*

$$\lambda(x_0) V_4(1, \gamma) \left[\frac{-1}{12} c^{(iv)}(x_0) + c'''(x_0) c'(x_0) + \frac{3}{4} (c''(x_0))^2 - 6 c''(x_0)(c'(x_0))^2 + 5(c'(x_0))^4 \right],$$

where $V_4(1, \gamma) = 3/((3 + 2\gamma)(5 + 2\gamma))$ and the superscript (iv) indicates the fourth-order derivative.

Proposition 2 Consider the framework of Theorem 2 in two dimensions ($d = 2$). Then the coefficient of h_n^4 in the expansion of bias $\hat{\lambda}_n(\mathbf{x_0})$ is

$$\lambda(\mathbf{x_0}) \{ C_4(\mathbf{x_0}) V_4(2, \gamma) + C_2(\mathbf{x_0}) V_2(2, \gamma) \},$$

with $V_4(2, \gamma) = 3/((4 + 2\gamma)(6 + 2\gamma))$, $V_2(2, \gamma) = 1/((4 + 2\gamma)(6 + 2\gamma))$ and constants $C_4(\mathbf{x_0})$ equal to

$$\sum_{i=1}^{2} \left[\frac{-D_{iiii} c(\mathbf{x_0})}{12} + D_i c(\mathbf{x_0}) D_{iii} c(\mathbf{x_0}) + \frac{3}{4} (D_{ii} c(\mathbf{x_0}))^2 - 6 (D_i c(\mathbf{x_0}))^2 D_{ii} c(\mathbf{x_0}) + 5 (D_i c(\mathbf{x_0}))^4 \right]$$

and

$$\begin{aligned} C_2(\mathbf{x_0}) = & \, 30(D_1 c(\mathbf{x_0}))^2 (D_2 c(\mathbf{x_0}))^2 - 6(D_1 c(\mathbf{x_0}))^2 D_{22} c(\mathbf{x_0}) - 6(D_2 c(\mathbf{x_0}))^2 D_{11} c(\mathbf{x_0}) \\ & - 24 D_1 c(\mathbf{x_0}) D_2 c(\mathbf{x_0}) D_{12} c(\mathbf{x_0}) + 3 D_1 c(\mathbf{x_0}) D_{122} c(\mathbf{x_0}) + 3 D_2 c(\mathbf{x_0}) D_{112} c(\mathbf{x_0}) \\ & + \frac{3}{2} D_{11} c(\mathbf{x_0}) D_{22} c(\mathbf{x_0}) + 3(D_{12} c(\mathbf{x_0}))^2 - \frac{1}{2} D_{1122} c(\mathbf{x_0}). \end{aligned}$$

Theorem 2 immediately yields the asymptotically optimal bandwidth.

Corollary 1 *Consider the setting of Theorem 2. Then*

$$\text{mse } \hat{\lambda}_n(\mathbf{x_0}) = \lambda(\mathbf{x_0})^2 \left(\int_{\mathbb{R}^d} A(\mathbf{u}; \mathbf{x_0}) d\mathbf{u} \right)^2 h_n^8 + \frac{\lambda(\mathbf{x_0}) Q(d, \gamma)}{n h_n^d} + O(h_n^{8+\alpha}) + O\left(\frac{1}{n h_n^{d-1}} \right).$$

Provided that $\int_{\mathbb{R}^d} A(\mathbf{u}; \mathbf{x_0}) d\mathbf{u} \neq 0$, *the asymptotic mean squared error is optimised at*

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

Note that the optimal bandwidth $h_n^*(\mathbf{x_0})$ as well as the weight function $(\lambda(\mathbf{x})/\lambda(\mathbf{x_0}))^{1/2}$ depend on the unknown intensity function. Clearly $h_n^*(\mathbf{x_0})$ tends to zero as $n \to \infty$ at a rate $n^{-1/(d+8)}$ which should be compared to $n^{-1/(d+4)}$ for estimators based on (1) with a fixed bandwidth (van Lieshout 2020), bearing in mind that we assume that the intensity function is four times continuously differentiable rather than two.

To conclude this section, we study asymptotics for the estimator $\hat{\lambda}_n$ itself. The proof can be found in Section 5.2.

Proposition 3 In the framework of Theorem 2, for $\mathbf{x_0} \in W$,

$$\hat{\lambda}_n(\mathbf{x_0}) = \lambda(\mathbf{x_0}) + O(h_n^4) + O_P\left(n^{-1/2} h_n^{-d/2} \right)$$

as $n \to \infty$.

4. Discussion and examples

To illustrate our approach, we first present an example in which the ground truth is known, then consider a real life example, and close with some discussion.

4.1. Theoretical example

As an illustration, let us consider a log-Gaussian Cox process Φ on the plane observed in the unit rectangle $(0,1)^2$ (Coles & Jones 1991). Such a process is defined as follows. Conditional on a realisation Z of a Gaussian field, points are generated according to a Poisson process with intensity function e^Z. The underlying Gaussian field is defined by its mean function m and covariance function ρ. Under the assumption that $\rho(\mathbf{y}, \mathbf{y}) \equiv \sigma^2$ is not location dependent, the intensity function of Φ is given by

$$\lambda(\mathbf{y}) = \exp\left[m(\mathbf{y}) + \sigma^2/2 \right].$$

The pair correlation function exceeds one (Møller, Syversveen & Waagepetersen 1998), which is indicative of attraction between the points in Φ.

For convenience, set

$$m(\mathbf{y}) = -\beta(y_1 + y_2), \quad \mathbf{y} = (y_1, y_2)^\top \in \mathbb{R}^2,$$

for some $\beta > 0$, and use a Beta kernel with $\gamma = 6$ to estimate the intensity function based on n samples from the model. Then

$$h_n^*(\mathbf{x_0}) = 2.75(n\lambda(\mathbf{x_0})\{3C_4(\mathbf{x_0}) + C_2(\mathbf{x_0})\}^2)^{-1/10} = 2.75\left(\frac{4}{n\lambda(\mathbf{x_0})\beta^8}\right)^{1/10}.$$

Thus, h_n^* decreases with the intensity. The same is true for the optimal asymptotic bandwidth

$$\tilde{h}_n^*(\mathbf{x_0}) = 2.92\left(\frac{\lambda(\mathbf{x_0})}{n\left(\lambda_{11}(\mathbf{x_0}) + \lambda_{22}(\mathbf{x_0})\right)^2}\right)^{1/6} = 2.92\left(\frac{1}{4n\lambda(\mathbf{x_0})\beta^4}\right)^{1/6}$$

in the classical regime (van Lieshout 2020). To illustrate the effect of the Abramson square root rule, let us focus on the point $\mathbf{x_0} = (0.1, 0.1)^\top$. The weighted bandwidths $h_n^*(\mathbf{x_0})(\lambda(\mathbf{x_0})/\lambda(\mathbf{y}))^{1/2}$ that govern the amount of smoothing for the observed points $\mathbf{y}$ decrease with $\lambda(\mathbf{y})$ and are shown in Figure 1 for $n = 1$, $\beta = 5$ and $\sigma = 4$. Note that the weighted bandwidths are much larger in low-intensity regions than in those with a high intensity. For this particular choice of parameters, the asymptotic mean squared error at $\mathbf{x_0} = (0.1, 0.1)^\top$ is $8,767n^{-8/10}$. For comparison, the asymptotically optimal fixed bandwidth is $\tilde{h}_n^*(\mathbf{x_0}) = 0.25n^{-1/6}$ with corresponding asymptotic mean squared error $32,493n^{-4/6}$.

4.2. Tornadoes example

Let us consider data submitted by US National Weather Service field offices and published by the Storm Prediction Center at http://www.spc.noaa.gov/wcm/#data.

The distribution of tornadoes over the year is bi-modal, with peaks in spring and autumn. Figure 2 plots maps of tornadoes (in starting longitude/latitude coordinates) in the apring season (February–July) in Kansas during the years 2008–2018. Focussing on only one of the seasons suggests that the time elapsed between two such seasons is so large that patterns in different years can be treated as independent.

We are interested in the tornado intensity at the state capital Wichita, $\mathbf{x_0} = (-97.33, 37.68)^\top$, indicated by a cross in the panels in Figure 2.

To evaluate $h_n^*(\mathbf{x_0})$, a pilot estimator is needed. We use a non-parametric non-adaptive bandwidth selector (Cronie & van Lieshout 2018). Since the latter uses a Gaussian kernel, according to the principle of equal variance we scale the result by four to get an appropriate bandwidth for the Beta kernel. For $\gamma = 6$, κ^γ is five times continuously differentiable and can also be used to estimate $C_2(\mathbf{x_0})$ and $C_4(\mathbf{x_0})$. An appeal to Proposition 2 and Corollary 1 results in $h_n^*(\mathbf{x_0}) \approx 0.8$ and $\hat{\lambda}_n(\mathbf{x_0}) \approx 2.6$. For fixed bandwidth 0.8, a classic kernel estimator would give an intensity value of 2.9 at Wichita.

It is important to stress that we used a non-parametric pilot estimator for simplicity only. Clearly, there is ample room for improvement, for instance along the lines of the iterative techniques proposed by Engel, Herrmann & Gasser (1994) and Lo (2017). However, such techniques involve tuning parameters that require additional research.

Figure 1. Asymptotically optimal adaptive bandwidth $n^{1/10}h_n^*(\mathbf{x_0})(\lambda(\mathbf{x_0})/\lambda(\mathbf{y}))^{1/2}$ plotted for $\mathbf{x_0} = (0.1, 0.1)^T$ as a function of $\mathbf{y} \in (0,1)^2$ for the log-Gaussian Cox model discussed in the text with $\beta = 5$, $\sigma = 4$ using a Beta kernel with $\gamma = 6$.

4.3. Discussion

Note that our asymptotic framework relies on the availability of replicates. In situations in which only a single pattern is available, no asymptotic results can be hoped for unless one is willing to assume that the points are independently distributed. Indeed, a pattern's apparent inhomogeneity could also arise from attraction between the points and kernel estimators may be misleading.

With these reservations, for data consisting of a single pattern, a two-step implementation of the bandwidth selector proposed by Cronie & van Lieshout (2018) could be helpful. Briefly, provided $\Phi \neq \emptyset$, choose a non-adaptive pilot bandwidth h_g as before and then optimise

$$\left| \sum_{\mathbf{x} \in \Phi \cap W} \frac{1}{\hat{\lambda}(\mathbf{x}; h, \Phi, W)} - \ell(W) \right|$$

over $h > 0$ where $\ell(W)$ is the Lebesgue measure of W and $\hat{\lambda}$ the adaptive kernel estimator defined in (2) with $c(\mathbf{y})$ equal to the scaled square root of

$$\frac{1}{h_g^d} \sum_{\mathbf{z} \in \Phi \cap W} \kappa \left(\frac{\mathbf{y} - \mathbf{z}}{h_g} \right).$$

Instead of scaling by the estimated intensity at some target point $\mathbf{x_0}$, we may scale by its geometric mean over all points in $\Phi \cap W$.

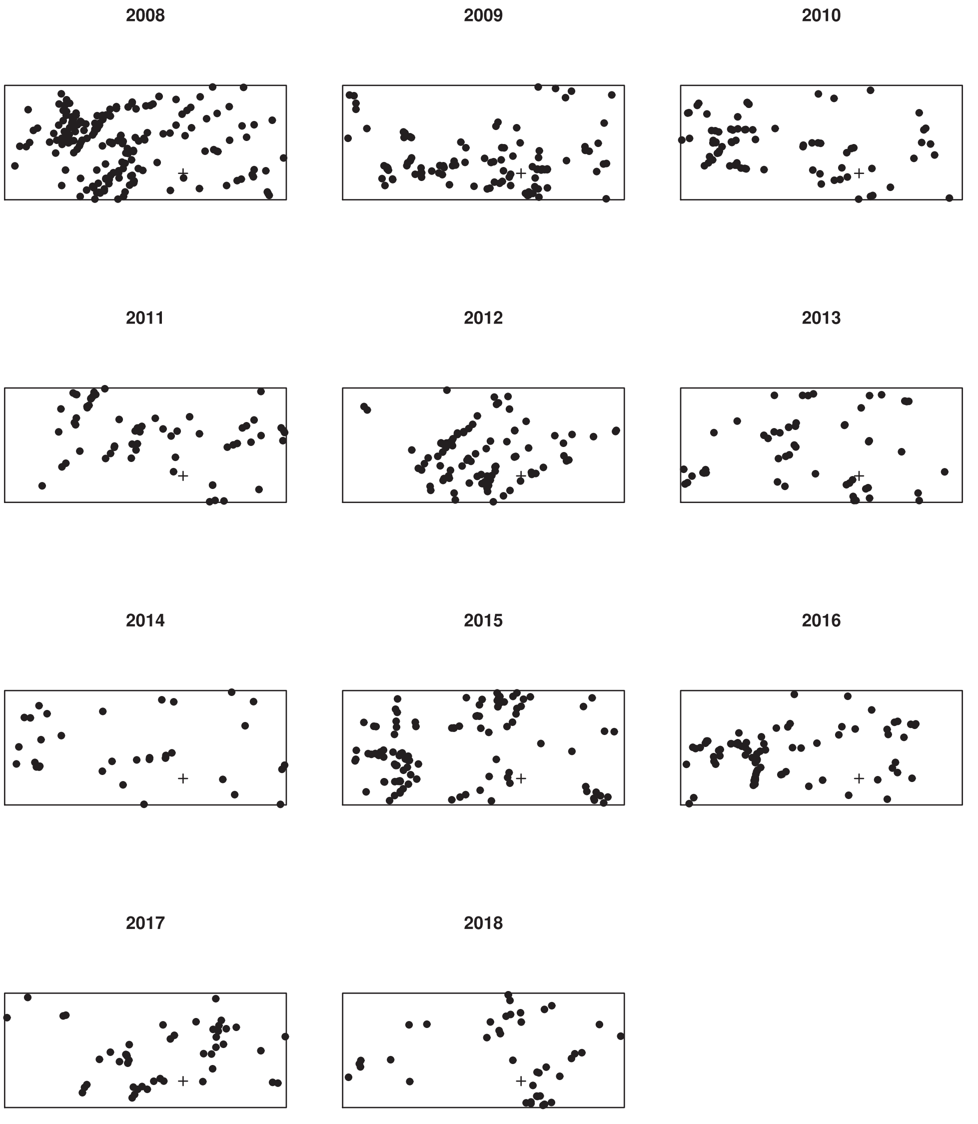

Figure 2. Tornadoes in Kansas during the Spring seasons of 2008–2018.

5. Proofs

5.1. Auxiliary lemmas

In the sequel, the following additional properties of the Beta kernels will be needed.

Lemma 1. *Consider the Beta kernels κ^γ defined in (4) with $\gamma > 1$. Then, for all $i \in \{1, \ldots, d\}$,*

$$\int_{\mathbb{R}^d} u_i D_i \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = -1,$$

the integrals of second- and fourth-order products in $\mathbf{u} \in \mathbb{R}^d$ *that include* u_i *with respect to* $D_i\kappa^\gamma$ *vanish and for distinct* $i, j \in \{1, \ldots, d\}$,

$$\int_{\mathbb{R}^d} u_i u_j^2 D_i \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = -V(d,\gamma); \quad \int_{\mathbb{R}^d} u_i^3 D_i \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = -3V(d,\gamma).$$

The integrals of other third-order products in $\mathbf{u} \in \mathbb{R}^d$ *that include* u_i *with respect to* $D_i\kappa^\gamma$ *vanish. Finally the following identities hold for all* $i \neq j \in \{1, \ldots, d\}$:

$$\int_{\mathbb{R}^d} u_i u_j \sum_{k=1}^{d} u_k D_k \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = 0$$

and

$$\int_{\mathbb{R}^d} u_i^2 \sum_{k=1}^{d} u_k D_k \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = -(d+2)V(d,\gamma).$$

Lemma 2 *Consider the Beta kernels* κ^γ *defined in (4) with* $\gamma > 2$. *Then, for all* $i \neq j \in \{1, \ldots, d\}$,

$$\int_{\mathbb{R}^d} u_i u_j \sum_{k=1}^{d} \sum_{l=1}^{d} u_k u_l D_{kl} \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = 0$$

and

$$\int_{\mathbb{R}^d} u_i^2 \sum_{k=1}^{d} \sum_{l=1}^{d} u_k u_l D_{kl} \kappa^\gamma(\mathbf{u}) du_1 \cdots du_d = (d+2)(d+3)V(d,\gamma).$$

The integrals of fifth-order products in $\mathbf{u} \in \mathbb{R}^d$ *that include* $u_k u_l$ *with respect to* $D_{kl}\kappa^\gamma$ *vanish.*

Lemma 3 *Consider the Beta kernels* κ^γ *defined in (4). For* γ *large enough for the partial derivatives to exist and be continuous, and for all* $i \in \{1, \ldots, d\}$,

$$\int_{\mathbb{R}^d} u_i^5 D_i \kappa^\gamma(\mathbf{u}) d\mathbf{u} = -5V_4(d,\gamma); \quad \int_{\mathbb{R}^d} u_i^6 D_{ii} \kappa^\gamma(\mathbf{u}) d\mathbf{u} = 30V_4(d,\gamma);$$

$$\int_{\mathbb{R}^d} u_i^7 D_{iii} \kappa^\gamma(\mathbf{u}) d\mathbf{u} = -210V_4(d,\gamma); \quad \int_{\mathbb{R}^d} u_i^8 D_{iiii} \kappa^\gamma(\mathbf{u}) d\mathbf{u} = 1680V_4(d,\gamma).$$

The proofs of Lemmas 1–3 rely on partial integrations and are left to the reader.

Lemma 4 *For fixed* $\mathbf{u} \in \mathbb{R}^d$, *the function* $g_{\mathbf{u}} : \mathbb{R} \to \mathbb{R}$ *defined by* $g_{\mathbf{u}}(v) = v^{d+2}\kappa^\gamma(v\mathbf{u})$ *is, for a Beta kernel* κ^γ *with* $\gamma > 4$, *four times continuously differentiable. The first three derivatives are given by*

$$
\begin{aligned}
g_{\mathbf{u}}'(v) &= (d+2)v^{d+1}\kappa^\gamma(v\mathbf{u}) + v^{d+2}D\kappa^\gamma(v\mathbf{u})(u^{(1)}),\\
g_{\mathbf{u}}''(v) &= (d+1)(d+2)v^{d}\kappa^\gamma(v\mathbf{u}) + 2(d+2)v^{d+1}D\kappa^\gamma(v\mathbf{u})(u^{(1)})\\
&\quad + v^{d+2}D^2\kappa^\gamma(v\mathbf{u})(u^{(2)}),\\
g_{\mathbf{u}}'''(v) &= d(d+1)(d+2)v^{d-1}\kappa^\gamma(v\mathbf{u}) + 3(d+1)(d+2)v^{d}D\kappa^\gamma(v\mathbf{u})(u^{(1)})\\
&\quad + 3(d+2)v^{d+1}D^2\kappa^\gamma(v\mathbf{u})(u^{(2)}) + v^{d+2}D^3\kappa^\gamma(v\mathbf{u})(u^{(3)})
\end{aligned}
$$

and the fourth-order derivative is

$$
\begin{aligned}
g_{\mathbf{u}}^{(iv)}(v) &= (d-1)d(d+1)(d+2)v^{d-2}\kappa^\gamma(v\mathbf{u})\\
&\quad + 4d(d+1)(d+2)v^{d-1}D\kappa^\gamma(v\mathbf{u})(u^{(1)})\\
&\quad + 6(d+1)(d+2)v^{d}D^2\kappa^\gamma(v\mathbf{u})(u^{(2)})\\
&\quad + 4(d+2)v^{d+1}D^3\kappa^\gamma(v\mathbf{u})(u^{(3)}) + v^{d+2}D^4\kappa^\gamma(v\mathbf{u})(u^{(4)}).
\end{aligned}
$$

Proof of Lemma 4 For $\gamma > 4$, the function κ^γ is four times continuously differentiable. The expressions for the derivatives follow by straightforward calculation.

5.2. Proofs of propositions and theorems

Proof of Theorem 1 To prove Part 1 note that since h_n goes to zero, $\mathbf{x_0} \in W$, W is open and λ is bounded away from zero, if n is large enough,

$$
\left\{ \mathbf{y} \in \mathbb{R}^d : \frac{c(\mathbf{y})}{h_n} \|\mathbf{x_0} - \mathbf{y}\| \leqslant 1 \right\} \subseteq b\left(\mathbf{x_0}, \frac{\lambda(\mathbf{x_0})^{1/2} h_n}{\underline{\lambda}^{1/2}} \right) \subseteq W. \tag{8}
$$

For such n, by a change of variables, the properties of the Beta kernels and the definition of the intensity function, the bias is equal to

$$
\begin{aligned}
\operatorname{bias} \hat{\lambda}_n(\mathbf{x_0}) &= \int_{\mathbb{R}^d} \left[\lambda(\mathbf{x_0}) c(\mathbf{x_0} + h_n\mathbf{u})^{d+2} \kappa^\gamma \left(c(\mathbf{x_0} + h_n\mathbf{u})\mathbf{u} \right) - \lambda(\mathbf{x_0}) \kappa^\gamma(\mathbf{u}) \right] d\mathbf{u}\\
&= \lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} \left[g_{\mathbf{u}}(c(\mathbf{x_0} + h_n\mathbf{u})) - g_{\mathbf{u}}(1) \right] d\mathbf{u}
\end{aligned} \tag{9}
$$

for the functions $g_{\mathbf{u}} : \mathbb{R} \to \mathbb{R}$, $\mathbf{u} \in \mathbb{R}^d$, defined by

$$
g_{\mathbf{u}}(v) = v^{d+2} \kappa^\gamma(v\mathbf{u}).
$$

Note that the integral in (9) is compactly supported, say on $K \subseteq b(\mathbf{0}, (\lambda(\mathbf{x_0})/\underline{\lambda})^{1/2}) \subseteq \mathbb{R}^d$, a property it inherits from the Beta kernel since λ is bounded away from zero. Moreover, the assumption that λ is bounded and the fact that the Beta kernels are bounded imply that the integrand is bounded in absolute value, so $\operatorname{bias} \hat{\lambda}_n(\mathbf{x_0})$ is finite.

Since we are after the coefficient of h_n^2 and, for $\gamma > 2$, κ^γ is twice continuously differentiable, we use a Taylor expansion (7) with $k=2$. Fix $\mathbf{u} \in K$. Then

$$\begin{aligned} g_{\mathbf{u}}(1+v) - g_{\mathbf{u}}(1) &= g'_{\mathbf{u}}(1)v + R(\mathbf{u}, v) \\ &= \left\{ (d+2)\kappa^\gamma(\mathbf{u}) + \sum_{i=1}^{d} u_i D_i \kappa^\gamma(\mathbf{u}) \right\} v + R(\mathbf{u}, v), \end{aligned} \tag{10}$$

where the remainder term is

$$R(\mathbf{u}, v) = \frac{v^2}{2} g''_{\mathbf{u}}(1+\theta v)$$

for some $0 < \theta = \theta(v) < 1$ that may depend on $v \in \mathbb{R}$. Moreover, by Lemma 4, $g''_{\mathbf{u}}(v)$ can be written as

$$(d+1)(d+2)v^d \kappa^\gamma(v\mathbf{u}) + 2(d+2)v^{d+1} \sum_{i=1}^{d} u_i D_i \kappa^\gamma(v\mathbf{u}) + v^{d+2} \sum_{i=1}^{d} \sum_{j=1}^{d} u_i u_j D_{ij} \kappa^\gamma(v\mathbf{u}).$$

Recall that in (9), the function $g_{\mathbf{u}}$ is evaluated at v of the form $c(\mathbf{x_0} + h_n \mathbf{u})$. Since the function c is bounded, we may restrict ourselves to a compact interval I for v and on this interval $g''_{\mathbf{u}}(v)$ is bounded as κ^γ and its partial derivatives are bounded too. Moreover, the bound can be chosen uniformly in $\mathbf{u}$ over the compact set K. In summary, there exists a constant $C > 0$ such that $|R(\mathbf{u}, v)| \leqslant Cv^2$ for all $\mathbf{u} \in K$ and $v \in I$.

We also need a Taylor expansion (7) with $k=2$ for the function c around $\mathbf{x_0} \in W$. Note that for $\mathbf{u} \in K$,

$$h_n \mathbf{u} \in b\left(\mathbf{0}, \frac{\lambda(\mathbf{x_0})^{1/2} h_n}{\underline{\lambda}^{1/2}}\right)$$

so the same is true for any $th_n\mathbf{u}$ with $t \in [0,1]$. Hence, by (8), for large n, $\mathbf{x_0} + th_n\mathbf{u} \in W$ and

$$c(\mathbf{x_0} + h_n\mathbf{u}) - 1 = h_n \sum_{i=1}^{d} u_i D_i c(\mathbf{x_0}) + \tilde{R}_n(\mathbf{u}) = h_n \sum_{i=1}^{d} u_i \frac{D_i \lambda(\mathbf{x_0})}{2\lambda(\mathbf{x_0})} + \tilde{R}_n(\mathbf{u}) \tag{11}$$

where the remainder term is

$$\tilde{R}_n(\mathbf{u}) = \frac{h_n^2}{2} D^2 c(\mathbf{x_0} + \theta h_n \mathbf{u})(u^{(2)}) = \frac{h_n^2}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} u_i u_j D_{ij} c(\mathbf{x_0} + \theta h_n \mathbf{u})$$

for some $0 < \theta = \theta(\mathbf{u}) < 1$ that may depend on $\mathbf{u} \in K$. The function c is twice continuously differentiable and, for $\mathbf{x} \in W$, the second-order partial derivatives are, for $i, j \in \{1, \ldots, d\}$,

$$D_{ij} c(\mathbf{x}) = \frac{1}{2\lambda(\mathbf{x_0})^{1/2}} \frac{\lambda_{ij}(\mathbf{x})}{\lambda(\mathbf{x})^{1/2}} - \frac{1}{4\lambda(\mathbf{x_0})^{1/2}} \frac{\lambda_i(\mathbf{x})\lambda_j(\mathbf{x})}{\lambda(\mathbf{x})^{3/2}},$$

where we use the notation $\lambda_i = D_i \lambda$. By assumption, λ is bounded and bounded away from zero and its partial derivatives are continuous and therefore bounded on compact sets. Returning to (11), for $\mathbf{u} \in K$ and $t \in [0,1]$, $\mathbf{x_0} + th_n\mathbf{u}$ is contained in the closed ball around $\mathbf{x_0}$ with radius $(\lambda(\mathbf{x_0})/\underline{\lambda})^{1/2} h_n$ on which $D_{ij}c$ is bounded. Also the $|u_i|$ are bounded on K. We conclude that there exists a constant $\tilde{C} > 0$ such that $|\tilde{R}_n(\mathbf{u})| \leqslant \tilde{C} h_n^2$, for all $\mathbf{u} \in K$.

Our next step is to combine the Taylor series (10) and (11). Write $E_n(\mathbf{u}) = c(\mathbf{x_0} + h_n\mathbf{u}) - 1$. For large n, (10) implies that the bias (9) can be written as

$$\begin{aligned} \text{bias}\,\hat{\lambda}_n(\mathbf{x_0}) &= \lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} \left[g_{\mathbf{u}}(1 + E_n(\mathbf{u})) - g_{\mathbf{u}}(1) \right] d\mathbf{u} \\ &= \lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} \left[E_n(\mathbf{u}) g'_{\mathbf{u}}(1) + \frac{E_n(\mathbf{u})^2}{2} g''_{\mathbf{u}}(1 + \eta(\mathbf{u}) E_n(\mathbf{u})) \right] d\mathbf{u} \end{aligned} \tag{12}$$

for some $\eta(\mathbf{u}) \in (0,1)$. We will show that the first- and second-order terms in (12) vanish. By (11) and Lemma 4, the first-order term is equal to $h_n \lambda(\mathbf{x_0})$ multiplied by

$$\int_{\mathbb{R}^d} Dc(\mathbf{x_0})(u^{(1)}) g'_{\mathbf{u}}(1) d\mathbf{u} = \sum_{i=1}^d D_i c(\mathbf{x_0}) \int_{\mathbb{R}^d} u_i \left[(d+2)\kappa^\gamma(\mathbf{u}) + \sum_{j=1}^d u_j D_j \kappa^\gamma(\mathbf{u}) \right] d\mathbf{u}$$

and vanishes because of Lemma 1 and the symmetry properties of κ^γ. Also by (11), the second-order term reads $h_n^2 \lambda(\mathbf{x_0}) I_n / 2$ where

$$I_n = \int_{\mathbb{R}^d} \left[g'_{\mathbf{u}}(1) D^2 c(\mathbf{x_0} + \theta(\mathbf{u}) h_n \mathbf{u})(u^{(2)}) + g''_{\mathbf{u}}(1 + \eta(\mathbf{u}) E_n(\mathbf{u})) \left\{ Dc(\mathbf{x_0})(u^{(1)}) \right\}^2 \right] d\mathbf{u}$$

for some $\theta(\mathbf{u})$ and $\eta(\mathbf{u})$ in $(0,1)$. Recall that (12) is compactly supported. This property is inherited by I_n. Moreover, its integrand is bounded on compact sets. By the dominated convergence theorem and Lemma 4,

$$\begin{aligned} \lim_{n\to\infty} I_n &= \sum_{i=1}^d \sum_{j=1}^d D_{ij} c(\mathbf{x_0}) \int_{\mathbb{R}^d} u_i u_j \left[(d+2)\kappa^\gamma(\mathbf{u}) + \sum_{k=1}^d u_k D_k \kappa^\gamma(\mathbf{u}) \right] d\mathbf{u} \\ &+ \sum_{i=1}^d \sum_{j=1}^d D_i c(\mathbf{x_0}) D_j c(\mathbf{x_0}) \int_{\mathbb{R}^d} u_i u_j \times \\ &\times \left[(d+1)(d+2)\kappa^\gamma(\mathbf{u}) + 2(d+2) \sum_{k=1}^d u_k D_k \kappa^\gamma(\mathbf{u}) + \sum_{k=1}^d \sum_{l=1}^d u_k u_l D_{kl} \kappa^\gamma(\mathbf{u}) \right] d\mathbf{u}. \end{aligned}$$

The first double sum is zero because of the symmetry of the Beta kernels and Lemma 1, the second one because of the symmetry of the Beta kernels, Lemmas 1 and 2. By the bounds on the remainder terms R and $\tilde{R}$, all other terms in (12) are of the order $o(h_n^2)$ and the proof is complete.

To prove Part 2 note that, as for the bias, n may be chosen so large that (8) holds. For such n, by a change of variables $\mathbf{u} = (\mathbf{y} - \mathbf{x_0})/h_n$ and the symmetry of the Beta kernels,

$$\frac{1}{n h_n^{2d}} \int_W c(\mathbf{y})^{2d} \kappa^\gamma \left(\frac{\mathbf{x_0} - \mathbf{y}}{h_n} c(\mathbf{y}) \right)^2 \lambda(\mathbf{y}) d\mathbf{y} = \tag{13}$$

$$\int_{\mathbb{R}^d} \frac{c(\mathbf{x_0} + h_n\mathbf{u})^{2d}}{n h_n^d} \kappa^\gamma \left(c(\mathbf{x_0} + h_n\mathbf{u})\mathbf{u} \right)^2 \lambda(\mathbf{x_0} + h_n\mathbf{u}) d\mathbf{u} = \frac{\lambda(\mathbf{x_0})}{n h_n^d} \int_{\mathbb{R}^d} h_{\mathbf{u}} \left(c(\mathbf{x_0} + h_n\mathbf{u}) \right) d\mathbf{u}$$

for the function $h_{\mathbf{u}} : \mathbb{R} \to \mathbb{R}$, $\mathbf{u} \in \mathbb{R}^d$, defined by

$$h_{\mathbf{u}}(v) = v^{2d+2} \kappa^\gamma(v\mathbf{u})^2.$$

Note that the integrals in the equation below (13) are compactly supported on $K \subseteq \mathbb{R}^d$ as in the proof of Part 1. Moreover, the assumption that λ is bounded and the fact that the Beta kernels are bounded imply that the integrand is bounded in absolute value, so (13) is finite.

Fix $\mathbf{u} \in K$. Then, by a Taylor expansion (7) with $k=1$,

$$h_{\mathbf{u}}(1+v) = h_{\mathbf{u}}(1) + h'_{\mathbf{u}}(1+\theta v)v$$

for some $0 < \theta(v) < 1$, $v \in \mathbb{R}$, with

$$h'_{\mathbf{u}}(v) = 2(d+1)v^{2d+1}\kappa^{\gamma}(v\mathbf{u})^2 + 2v^{2d+2}\kappa^{\gamma}(v\mathbf{u})\sum_{i=1}^{d} u_i D_i \kappa^{\gamma}(v\mathbf{u}).$$

Recall that in (13), the function $h_{\mathbf{u}}$ is evaluated at v of the form $c(\mathbf{x_0} + h_n\mathbf{u})$. Since the function c is bounded we may restrict ourselves to a compact interval I for v and on this interval $h'_{\mathbf{u}}(v)$ is bounded as κ^{γ} and its partial derivatives are bounded too. Moreover, the bound can be chosen uniformly in $\mathbf{u}$ over the compact set K. In summary, there exists a constant H such that $|h'_{\mathbf{u}}(v)| \leqslant H$ for all $\mathbf{u} \in K$ and $v \in I$. Hence, with $E_n(\mathbf{u}) = c(\mathbf{x_0} + h_n\mathbf{u}) - 1$ as before, (13) can be written as

$$\frac{\lambda(\mathbf{x_0})}{nh_n^d}\int_{\mathbb{R}^d}\left[h_{\mathbf{u}}(1) + E_n(\mathbf{u})h'_{\mathbf{u}}(1+\theta(\mathbf{u})E_n(\mathbf{u}))\right]d\mathbf{u} = \frac{\lambda(\mathbf{x_0})}{nh_n^d}\int_{\mathbb{R}^d}\kappa^{\gamma}(\mathbf{u})^2 d\mathbf{u} + R_n$$

for a remainder term

$$R_n = \frac{\lambda(\mathbf{x_0})}{nh_n^d}\int_{\mathbb{R}^d} E_n(\mathbf{u})h'_{\mathbf{u}}(1+\theta(\mathbf{u})E_n(\mathbf{u}))d\mathbf{u}$$

with $0 < \theta(\mathbf{u}) < 1$. By (11),

$$|E_n(\mathbf{u})h'_{\mathbf{u}}(1+\theta(\mathbf{u})E_n(\mathbf{u}))| \leqslant H|E_n(\mathbf{u})| \leqslant h_n H\left|\sum_{i=1}^{d}\frac{\lambda_i(\mathbf{x_0})}{2\lambda(\mathbf{x_0})}u_i\right| + H\left|\tilde{R}_n(\mathbf{u})\right|.$$

As $\mathbf{u} \in K$ and, for such $\mathbf{u}$, $|\tilde{R}_n(\mathbf{u})| \leqslant \tilde{C}h_n^2$, we conclude that

$$\frac{\lambda(\mathbf{x_0})}{nh_n^d}\int_{\mathbb{R}^d} h_{\mathbf{u}}\left(c(\mathbf{x_0} + h_n\mathbf{u})\right)d\mathbf{u} = \frac{\lambda(\mathbf{x_0})}{nh_n^d}Q(d,\gamma) + O\left(\frac{1}{nh_n^{d-1}}\right).$$

We will finally show that the contribution of the interaction structure (through the pair correlation function) to the variance of $\hat{\lambda}_n(\mathbf{x_0})$ vanishes. Again, choose n so large that (8) holds. For such n, by a change of variables and the symmetry of the Beta kernels, and writing $\bar{g}$ for an upper bound to the pair correlation function, the integral in the last line in (3) divided by n can be bounded in absolute value by

$$\frac{(1+\bar{g})\lambda(\mathbf{x_0})^2}{n}\left(\int_{\mathbb{R}^d} c(\mathbf{x_0} + h_n\mathbf{u})^{d+2}\kappa^{\gamma}(c(\mathbf{x_0} + h_n\mathbf{u})\mathbf{u})d\mathbf{u}\right)^2 = O\left(\frac{1}{n}\right),$$

since the integral is compactly supported and both c and κ^{γ} are bounded.

Proof of Theorem 2 As in the proof of Theorem 1, for n large enough, the bias is given by (9) and the integral involved is finite and supported on a compact set $K \subseteq b(\mathbf{0}, (\lambda(\mathbf{x_0})/\underline{\lambda})^{1/2}) \subseteq \mathbb{R}^d$. Finiteness is ensured by the assumptions that λ and κ^{γ} are bounded. Moreover, $\mathbf{x_0} + th_n\mathbf{u}$ lies in W for all $\mathbf{u} \in K$ and $t \in [0,1]$. To calculate the coefficient of h_n^4, we use Taylor expansions

(7) with $k=4$ for both c and $g_{\mathbf{u}}$. Note that this can be done because of the smoothness assumptions on λ and κ^γ. Now,

$$\begin{aligned} E_n(\mathbf{u}) &= c(\mathbf{x_0}+h_n\mathbf{u})-1 \\ &= h_n Dc(\mathbf{x_0})(u^{(1)}) + \frac{1}{2}h_n^2 D^2 c(\mathbf{x_0})(u^{(2)}) + \frac{1}{6}h_n^3 D^3 c(\mathbf{x_0})(u^{(3)}) \\ &\quad + \frac{1}{24}h_n^4 D^4 c(\mathbf{x_0})(u^{(4)}) + \tilde{R}_n(\mathbf{u}) \end{aligned} \tag{14}$$

for a remainder term

$$\tilde{R}_n(\mathbf{u}) = \frac{1}{24}h_n^4 D^4 c(\mathbf{x_0}+\theta h_n\mathbf{u})(u^{(4)}) - \frac{1}{24}h_n^4 D^4 c(\mathbf{x_0})(u^{(4)})$$

for some $0<\theta=\theta(\mathbf{u})<1$ that may depend on $\mathbf{u}\in K$. Now,

$$\left| D^4 c(\mathbf{x_0}+\theta h_n\mathbf{u})(u^{(4)}) - D^4 c(\mathbf{x_0})(u^{(4)}) \right| =$$

$$\left| \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d u_i u_j u_k u_l \left(D_{ijkl} c(\mathbf{x_0}+\theta h_n\mathbf{u}) - D_{ijkl} c(\mathbf{x_0}) \right) \right|$$

is dominated by

$$\sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d \left| u_i u_j u_k u_l \right| \times \left| D_{ijkl} c(\mathbf{x_0}+\theta h_n\mathbf{u}) - D_{ijkl} c(\mathbf{x_0}) \right| .$$

The first term in the summand is bounded on the compact set K. For the second term we may use the Hölder assumption to obtain

$$\left| D_{ijkl} c(\mathbf{x_0}+\theta h_n\mathbf{u}) - D_{ijkl} c(\mathbf{x_0}) \right| \leqslant C\|\theta h_n\mathbf{u}\|^\alpha \leqslant C h_n^\alpha \|\mathbf{u}\|^\alpha.$$

The norm of $\mathbf{u}\in K$ is bounded on the compact set K, therefore, in summary, for all $\mathbf{u}\in K$,

$$|\tilde{R}_n(\mathbf{u})| \leqslant \tilde{C} h_n^{4+\alpha}.$$

Next, for fixed $\mathbf{u}\in K$,

$$g_{\mathbf{u}}(1+v) - g_{\mathbf{u}}(1) = v g_{\mathbf{u}}'(1) + \frac{1}{2}v^2 g_{\mathbf{u}}''(1) + \frac{1}{6}v^3 g_{\mathbf{u}}'''(1) + \frac{1}{24}v^4 g_{\mathbf{u}}^{(iv)}(1) + R(\mathbf{u},v),$$

where $R(\mathbf{u},v) = D^5 g_{\mathbf{u}}(1+\theta v)(v^{(5)})/120$ for some $\theta=\theta(v)$ in $(0,1)$ that may depend on $v\in\mathbb{R}$. Recall that $g_{\mathbf{u}}$ is evaluated at v of the form $c(\mathbf{x_0}+h_n\mathbf{u})$, $\mathbf{u}\in K$. Since the function c is bounded we may restrict ourselves to a compact interval I for v that contains 1. On this interval, the fifth-order derivative of $g_{\mathbf{u}}$ is bounded as, for $\gamma>5$, the function κ^γ and its partial derivatives up to fifth order are bounded. Moreover, the bound can be chosen uniformly in $\mathbf{u}$ over the compact set K. In summary, $|R(\mathbf{u},v)|\leqslant Cv^5$ for $\mathbf{u}\in K$ and $v\in I$.

Next, plug the Taylor expansions into (9). Then

$$\text{bias}\,\hat{\lambda}_n(\mathbf{x_0}) = \lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} \left[g_{\mathbf{u}}(1+E_n(\mathbf{u})) - g_{\mathbf{u}}(1) \right] d\mathbf{u} = \lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} R(\mathbf{u}, E_n(\mathbf{u})) d\mathbf{u} +$$

$$\lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} \left[E_n(\mathbf{u}) g_{\mathbf{u}}'(1) + \frac{1}{2}E_n(\mathbf{u})^2 g_{\mathbf{u}}''(1) + \frac{1}{6}E_n(\mathbf{u})^3 g_{\mathbf{u}}'''(1) + \frac{1}{24}E_n(\mathbf{u})^4 g_{\mathbf{u}}^{(iv)}(1) \right] d\mathbf{u}. \tag{15}$$

To obtain an expansion in terms of powers of h_n, plug in (14). For example, the first term in the integrand in (15) reads

$$\lambda(\mathbf{x_0})\int_{\mathbb{R}^d} E_n(\mathbf{u})g'_{\mathbf{u}}(1)d\mathbf{u} = \lambda(\mathbf{x_0})\int_{\mathbb{R}^d} g'_{\mathbf{u}}(1)\times$$

$$\times\{h_n Dc(\mathbf{x_0})(u^{(1)}) + \frac{1}{2}h_n^2 D^2 c(\mathbf{x_0})(u^{(2)}) + \frac{1}{6}h_n^3 D^3 c(\mathbf{x_0})(u^{(3)})$$

$$+\frac{1}{24}h_n^4 D^4 c(\mathbf{x_0})(u^{(4)}) + \tilde{R}_n(\mathbf{u})\}d\mathbf{u},$$

where $\tilde{R}_n(\mathbf{u})$ is bounded on K in absolute value by $\tilde{C}h_n^{4+\alpha}$.

By Theorem 1, the first- and second-order terms are zero. We will show that the third-order term $h_n^3\lambda(\mathbf{x_0})I_{n,3}$ vanishes too. By (14) and (15), $I_{n,3}$ is given by

$$\int_{\mathbb{R}^d}\left[\frac{g'_{\mathbf{u}}(1)}{6}D^3 c(\mathbf{x_0})(u^{(3)}) + \frac{g''_{\mathbf{u}}(1)}{2}Dc(\mathbf{x_0})(u^{(1)})D^2 c(\mathbf{x_0})(u^{(2)}) + \frac{g'''_{\mathbf{u}}(1)}{6}(Dc(\mathbf{x_0})(u^{(1)}))^3\right]d\mathbf{u}.$$

Lemma 4 implies that the first term of $I_{n,3}$ is

$$\frac{1}{6}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} D_{ijk}c(\mathbf{x_0})\int_{\mathbb{R}^d} u_i u_j u_k\left[(d+2)\kappa^\gamma(\mathbf{u}) + \sum_l u_l D_l \kappa^\gamma(\mathbf{u})\right]d\mathbf{u},$$

which vanishes by the symmetry properties of κ^γ and Lemma 1. By Lemma 4, the second term is a linear combination of integrals of the form

$$\int_{\mathbb{R}^d} u_i u_j u_k\left[(d+1)(d+2)\kappa^\gamma(\mathbf{u}) + 2(d+2)\sum_{l=1}^{d} u_l D_l \kappa^\gamma(\mathbf{u}) + \sum_{l=1}^{d}\sum_{m=1}^{d} u_l u_m D_{lm}\kappa^\gamma(\mathbf{u})\right]d\mathbf{u},$$

which vanish because of the symmetry properties of the Beta kernel, Lemma 1 and Lemma 2. Similar arguments, the details of which are left to the reader, apply to the third and last term of $I_{n,3}$, which, by Lemma 4, is a linear combination of integrals of the form

$$\int_{\mathbb{R}^d} u_i u_j u_k\{d(d+1)(d+2)\kappa^\gamma(\mathbf{u}) + 3(d+1)(d+2)\sum_{l=1}^{d} u_l D_l \kappa^\gamma(\mathbf{u}) +$$

$$3(d+2)\sum_{l=1}^{d}\sum_{m=1}^{d} u_l u_m D_{lm}\kappa^\gamma(\mathbf{u}) + \sum_{l=1}^{d}\sum_{m=1}^{d}\sum_{n=1}^{d} u_l u_m u_n D_{lmn}\kappa^\gamma(\mathbf{u})\}d\mathbf{u}.$$

The coefficient of h_n^4 in (15) reads $\lambda(\mathbf{x_0})\int A(\mathbf{u};\mathbf{x_0})d\mathbf{u}$ with A as claimed and does not vanish in general. The integral is finite because it is compactly supported and the partial derivatives of κ^γ are bounded up to fourth order for $\gamma > 5$.

Contributions of the order $h_n^{4+\alpha}$ come from

$$\lambda(\mathbf{x_0})\int_{\mathbb{R}^d} g'_{\mathbf{u}}(1)\tilde{R}_n(\mathbf{u})d\mathbf{u} = \lambda(\mathbf{x_0})\int_{\mathbb{R}^d}\left[(d+2)\kappa^\gamma(\mathbf{u}) + \sum_{i=1}^{d} u_i D_i \kappa^\gamma(\mathbf{u})\right]\tilde{R}_n(\mathbf{u})d\mathbf{u}$$

since $|\tilde{R}_n(\mathbf{u})| \leqslant \tilde{C} h_n^{4+\alpha}$ on K. Because

$$\left| \int_{\mathbb{R}^d} R(\mathbf{u}, E_n(\mathbf{u})) d\mathbf{u} \right| \leqslant \int_{\mathbb{R}^d} C |E_n(\mathbf{u})|^5 d\mathbf{u},$$

and because by (14) the leading term in $E_n(\mathbf{u})^5$ is $h_n^5 (Dc(\mathbf{x_0})(u^{(1)}))^5$ the integral contributes a term $O\left(h_n^5\right)$. All other terms in (15) of the order $O(h_n^5)$.

Proof of Proposition 1 By Theorem 2 the coefficient of h_n^4 is $\lambda(x_0) \int_{\mathbb{R}} A(u; x_0) du$, where

$$\frac{A(u;x_0)}{u^4} = \frac{1}{24} g_u'(1) c^{(iv)}(x_0) + \frac{1}{4} g_u'''(1) (c'(x_0))^2 c''(x_0) + \frac{1}{24} g_u^{(iv)}(1) (c'(x_0))^4 + \frac{1}{2} g_u''(1) \left[\frac{1}{3} c'(x_0) c'''(x_0) + \frac{1}{4} (c''(x_0))^2 \right].$$

Lemma 4 and Lemma 3 can be used to derive the following equations:

$$\begin{aligned}
\int_{\mathbb{R}} u^4 g_u'(1) du &= \int_{\mathbb{R}} u^4 \left[3\kappa^\gamma(u) + u D_1 \kappa^\gamma(u) \right] du = -2V_4(1,\gamma), \\
\int_{\mathbb{R}} u^4 g_u''(1) du &= \int_{\mathbb{R}} u^4 \left[6\kappa^\gamma(u) + 6u D_1 \kappa^\gamma(u) + u^2 D_{11} \kappa^\gamma(u) \right] du = 6V_4(1,\gamma), \\
\int_{\mathbb{R}} u^4 g_u'''(1) du &= \int_{\mathbb{R}} u^4 \left[6\kappa^\gamma(u) + 18u D_1 \kappa^\gamma(u) + 9u^2 D_{11} \kappa^\gamma(u) + u^3 D_{111} \kappa^\gamma(u) \right] du \\
&= -24 V_4(1,\gamma)
\end{aligned}$$

and $\int_{\mathbb{R}} u^4 g_u^{(iv)}(1)\, du$ is equal to

$$\int_{\mathbb{R}} u^4 \left[24u D_1 \kappa^\gamma(u) + 36u^2 D_{11} \kappa^\gamma(u) + 12u^3 D_{111} \kappa^\gamma(u) + u^4 D_{1111} \kappa^\gamma(u) \right] du$$

which in turn reduces to $120 V_4(1,\gamma)$. Hence, upon a rearrangement of terms, $\int_{\mathbb{R}} A(u;x_0) du$ is equal to

$$V_4(1,\gamma) \left[\frac{-1}{12} c^{(iv)}(x_0) + c'''(x_0) c'(x_0) + \frac{3}{4} (c''(x_0))^2 - 6c''(x_0)(c'(x_0))^2 + 5(c'(x_0))^4 \right].$$

Proof of Proposition 2 Theorem 2 states that the coefficient of h_n^4 is $\lambda(\mathbf{x_0}) \int A(\mathbf{u}; \mathbf{x_0}) d\mathbf{u}$ with an explicit expression for $A(\mathbf{u}; \mathbf{x_0})$.

The non-zero terms in this expression can be reduced by repeated partial integration to a scalar multiple of either $V_4(2,\gamma)$ or $V_2(2,\gamma)$ as the integrals of other fourth-order products in $\mathbf{u} \in \mathbb{R}^2$ with respect to κ^γ vanish by the symmetry properties of the Beta kernel.

The scalar multipliers can be calculated as in Lemma 3: for $i \neq j \in \{1,2\}$, integrals with respect to first-order partial derivatives reduce to

$$\int_{\mathbb{R}^2} u_i^4 u_j D_j \kappa^\gamma(\mathbf{u}) d\mathbf{u} = -V_4(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^3 u_j^2 D_i \kappa^\gamma(\mathbf{u}) d\mathbf{u} = -3V_2(2,\gamma),$$

integrals with respect to second-order partial derivatives reduce to

$$\int_{\mathbb{R}^2} u_i^5 u_j D_{ij} \kappa^\gamma(\mathbf{u}) d\mathbf{u} = 5V_4(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^4 u_j^2 D_{jj} \kappa^\gamma(\mathbf{u}) d\mathbf{u} = 2V_4(2,\gamma);$$

$$\int_{\mathbb{R}^2} u_i^4 u_j^2 D_{ii}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 12V_2(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^3 u_j^3 D_{ij}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 9V_2(2,\gamma),$$

and integrals with respect to third-order partial derivatives reduce to

$$\int_{\mathbb{R}^2} u_i^6 u_j D_{iij}\kappa^\gamma(\mathbf{u})d\mathbf{u} = -30V_4(2,\gamma)$$

and

$$\int_{\mathbb{R}^2} u_i^5 u_j^2 D_{ijj}\kappa^\gamma(\mathbf{u})d\mathbf{u} = -10V_4(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^4 u_j^3 D_{jjj}\kappa^\gamma(\mathbf{u})d\mathbf{u} = -6V_4(2,\gamma),$$

$$\int_{\mathbb{R}^2} u_i^5 u_j^2 D_{iii}\kappa^\gamma(\mathbf{u})d\mathbf{u} = -60V_2(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^4 u_j^3 D_{iij}\kappa^\gamma(\mathbf{u})d\mathbf{u} = -36V_2(2,\gamma).$$

Finally,

$$\int_{\mathbb{R}^d} u_i^7 u_j D_{iiij}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 210V_4(2,\gamma); \quad \int_{\mathbb{R}^d} u_i^6 u_j^2 D_{iijj}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 60V_4(2,\gamma);$$

$$\int_{\mathbb{R}^2} u_i^5 u_j^3 D_{ijjj}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 30V_4(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^4 u_j^4 D_{jjjj}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 24V_4(2,\gamma),$$

and

$$\int_{\mathbb{R}^2} u_i^6 u_j^2 D_{iiii}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 360V_2(2,\gamma); \quad \int_{\mathbb{R}^2} u_i^5 u_j^3 D_{iiij}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 180V_2(2,\gamma);$$

$$\int_{\mathbb{R}^2} u_i^4 u_j^4 D_{iijj}\kappa^\gamma(\mathbf{u})d\mathbf{u} = 144V_2(2,\gamma).$$

Evaluation of the expression for $A(\mathbf{u};\mathbf{x_0})$ implies the claim by elementary but tedious calculation. For example, the coefficient of $D_{11}c(\mathbf{x_0})D_{22}c(\mathbf{x_0})$ arises from terms with these coefficients in

$$\frac{1}{8}\int_{\mathbb{R}^2}(D^2c(\mathbf{x_0})(u^{(2)}))^2 g''_{\mathbf{u}}(1)du_1du_2,$$

which, by Lemma 4, is equal to

$$\frac{1}{8}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} D_{ij}c(\mathbf{x_0})D_{kl}c(\mathbf{x_0})\int_{\mathbb{R}^2} u_iu_ju_ku_l \times$$

$$\times\left[12\kappa^\gamma(\mathbf{u}) + 8D\kappa^\gamma(\mathbf{u})(u^{(1)}) + D^2\kappa^\gamma(\mathbf{u})(u^{(2)})\right]du_1du_2.$$

The desired coefficients occur when $i=j=1$ and $k=l=2$ or when $i=j=2$ and $k=l=2$. Their sum is

$$\frac{D_{11}c(\mathbf{x_0})D_{22}c(\mathbf{x_0})}{4}\int_{\mathbb{R}^2} u_1^2u_2^2\left[12\kappa^\gamma(\mathbf{u}) + 8D\kappa^\gamma(\mathbf{u})(u^{(1)}) + D^2\kappa^\gamma(\mathbf{u})(u^{(2)})\right]du_1du_2,$$

so the coefficient of $D_{11}c(\mathbf{x_0})D_{22}c(\mathbf{x_0})$ is equal to

$$\frac{1}{4}\left[12V_2(2,\gamma) + 8(-3-3)V_2(2,\gamma) + (12+9+9+12)V_2(2,\gamma)\right] = \frac{6V_2(2,\gamma)}{4}.$$

Proof of Corollary 1 By Theorem 2,

$$\left(\text{bias}\, \hat{\lambda}_n(\mathbf{x_0})\right)^2 = \lambda(\mathbf{x_0})^2 \left(\int_{\mathbb{R}^d} A(\mathbf{u};\mathbf{x_0})d\mathbf{u}\right)^2 h_n^8 + 2h_n^4 R(h_n)\lambda(\mathbf{x_0}) \int_{\mathbb{R}^d} A(\mathbf{u};\mathbf{x_0})d\mathbf{u} + R(h_n)^2$$

for a remainder term $R(h_n)$ for which there exists a constant $M>0$ for which $|h_n^{-4-\alpha}R(h_n)| \leqslant M$ for large n. Hence

$$\left(\text{bias}\, \hat{\lambda}_n(\mathbf{x_0})\right)^2 = \lambda(\mathbf{x_0})^2 \left(\int_{\mathbb{R}^d} A(\mathbf{u};\mathbf{x_0})d\mathbf{u}\right)^2 h_n^8 + O(h_n^{8+\alpha})$$

from which the claimed expression for the mean squared error follows upon adding the variance. Consequently, since the integral of $A(\mathbf{u};\mathbf{x_0})$ and λ are finite and non-zero by assumption, the asymptotic mean squared error takes the form

$$\alpha h_n^8 + \frac{\beta}{n h_n^d},$$

for some scalars $\alpha, \beta > 0$. Equating the derivative with respect to h_n to zero yields

$$(h_n^*)^{7+d+1} = \frac{d\beta}{8n\alpha}.$$

The second derivative with respect to h_n, $56\alpha h_n^6 + d(d+1)\beta n^{-1} h_n^{-d-2}$, is strictly positive, so h_n^* is the unique minimum. Plugging in the expressions for α and β completes the proof.

Proof of Proposition 3 Since $h_n \to 0$, $\mathbf{x_0} \in W$, W is open and λ is bounded away from zero, if n is large enough (8) holds. For such n,

$$\hat{\lambda}_n(\mathbf{x_0}) - \mathrm{E}\hat{\lambda}_n(\mathbf{x_0}) = \hat{\lambda}_n(\mathbf{x_0}) - \frac{1}{h_n^d}\int_{\mathbb{R}^d} c(\mathbf{x})^d \kappa^\gamma \left(\frac{\mathbf{x_0}-\mathbf{x}}{h_n} c(\mathbf{x})\right) \lambda(\mathbf{x})d\mathbf{x} = \frac{1}{n}\sum_{i=1}^n Z_i$$

can be written as an average of independent random variables

$$Z_i = \hat{\lambda}(\mathbf{x_0}; h_n, \Phi_i, W) - \frac{1}{h_n^d}\int_{\mathbb{R}^d} c(\mathbf{x})^d \kappa^\gamma \left(\frac{\mathbf{x_0}-\mathbf{x}}{h_n} c(\mathbf{x})\right) \lambda(\mathbf{x})d\mathbf{x}$$

with $\mathrm{E}Z_i = 0$. By Theorem 2

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \frac{\lambda(\mathbf{x_0})Q(d,\gamma)}{n h_n^d} + R(h_n)$$

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

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| \geqslant \epsilon^{-1/2}\sqrt{\frac{\lambda(\mathbf{x_0})Q(d,\gamma)}{n h_n^d}}\right) \leqslant \frac{\epsilon n h_n^d}{\lambda(\mathbf{x_0})Q(d,\gamma)}\left(\frac{\lambda(\mathbf{x_0})Q(d,\gamma)}{n h_n^d} + R(h_n)\right).$$

The right-hand side tends to ϵ as $n \to \infty$ so

$$\frac{1}{n}\sum_{i=1}^{n} Z_i = O_P\left(\sqrt{\frac{\lambda(\mathbf{x_0})Q(d,\gamma)}{nh_n^d}}\right).$$

The proof is completed on noticing that the bias is of the order $O(h_n^4)$ according to Theorem 2.

References

ABRAMSON, I.A. (1982). On bandwidth variation in kernel estimates – A square root law. *Annals of Statistics* **10**, 1217–1223.

BADDELEY, A., RUBAK, E. & TURNER, R. (2016). *Spatial Point Patterns: Methodology and Applications with R.* Boca Raton: CRC Press.

BERMAN, M. & DIGGLE, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society, series B* **51**, 81–92.

BOWMAN, A.W. & AZZALINI, A. (1997). *Applied Smoothing Techniques for Data Analysis. The Kernel Approach with S-Plus Illustrations.* Oxford: University Press.

BOWMAN, A.W. & FOSTER, P.J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *Journal of the American Statistical Association* **88**, 529–537.

BROOKS, M.M. & MARRON, J.S. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications* **38**, 157–165.

CHACÓN, J.E. & DUONG, T. (2018). *Multivariate Kernel Smoothing and its Applications.* Boca Raton: CRC Press.

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

COLES, P. & JONES, B. (1991). A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society* **248**, 1–13.

CRONIE, O. & VAN LIESHOUT, M.N.M. (2018). A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* **105**, 455–462.

DAVIES, T.M. & BADDELEY, A. (2018). Fast computation of spatially adaptive kernel estimates. *Statistics and Computing* **28**, 937–956.

DAVIES, T.M., FLYNN, C.R. & HAZELTON, M.L. (2018). On the utility of asymptotic bandwidth selectors for spatially adaptive kernel density estimation. *Statistics and Probability Letters* **138**, 75–81.

DIGGLE, P.J. (1985). A kernel method for smoothing point process data. *Journal of Applied Statistics* **34**, 138–147.

DUONG, T. (2015). Spherically symmetric multivariate Beta family kernels. *Statistics and Probability Letters* **104**, 141–145.

ENGEL J., HERRMANN, E. & GASSER, T. (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives. *Journal of Nonparametric Statistics* **4**, 21–34.

HALL, P., HU, T.C. & MARRON, J.S. (1995). Improved variable window kernel estimates of probability densities. *Annals of Statistics* **23**, 1–10.

ILLIAN, J., PENTTINEN, A., STOYAN, H. & STOYAN, D.. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns.* Chichester: Wiley.

VAN LIESHOUT, M.N.M. (2012). On estimation of the intensity function of a point process. *Methodology and Computing in Applied Probability* **14**, 567–578.

VAN LIESHOUT, M.N.M. (2019). *Theory of Spatial Statistics: A Concise Introduction.* Boca Raton: CRC Press.

VAN LIESHOUT, M.N.M. (2020). Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. *Methodology and Computing in Applied Probability* **22**, 995–1008.

LO, P.H. (2017). An Iterative Plug-in Algorithm for Optimal Bandwidth Selection in Kernel Intensity Estimation for Spatial Data. Technical University of Kaiserslautern: PhD Thesis.

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

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

RIPLEY, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge: University Press.
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*. Boca Raton: Chapman & Hall.
WAND, M.P. & JONES, M.C. (1995). *Kernel Smoothing*. Boca Raton: Chapman & Hall.