

NUMERICAL EXPERIMENTS FOR THE ESTIMATION OF MEAN DENSITIES OF RANDOM SETS

FEDERICO CAMERLENGHI¹, VINCENZO CAPASSO^{2,3} AND ELENA VILLA^{✉,2}

¹Dept. of Mathematics, Università degli Studi di Pavia, via Ferrata 1, 27100 Pavia, Italy; ²Dept. of Mathematics, Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy ; ³Escuela Politecnica Superior, Universidad Carlos III de Madrid, Av. de la Universidad 30 28911 Leganes, Spain
e-mail: federico.camerlenghi01@ateneopv.it, vincenzo.capasso@unimi.it, elena.villa@unimi.it
(Received November 13, 2013; revised February 25, 2014; accepted April 24, 2014)

ABSTRACT

Many real phenomena may be modelled as random closed sets in \mathbb{R}^d , of different Hausdorff dimensions. The problem of the estimation of pointwise mean densities of absolutely continuous, and spatially inhomogeneous, random sets with Hausdorff dimension $n < d$, has been the subject of extended mathematical analysis by the authors. In particular, two different kinds of estimators have been recently proposed, the first one is based on the notion of Minkowski content, the second one is a kernel-type estimator generalizing the well-known kernel density estimator for random variables. The specific aim of the present paper is to validate the theoretical results on statistical properties of those estimators by numerical experiments. We provide a set of simulations which illustrates their valuable properties via typical examples of lower dimensional random sets.

Keywords: density estimator, Hausdorff dimension, Hausdorff measure, kernel estimate, Minkowski content, random closed set, stochastic geometry.

INTRODUCTION

Given an Euclidean space \mathbb{R}^d , the problem of the evaluation and estimation of the mean density of lower dimensional random closed sets (*i.e.*, with Hausdorff dimension less than d), such as fiber processes, boundaries of germ-grain models, n -facets of random tessellations, and surfaces of full dimensional random sets, has been of great interest in many different scientific and technological fields over the last decades (see Camerlenghi *et al.*, 2014 and references therein). Recently, in Villa (2014), and Camerlenghi *et al.* (2014), two different kinds of estimators have been proposed by the authors, the first one is based on the notion of Minkowski content, the second one is a kernel-type estimator generalizing the well-known kernel density estimator for random variables.

The specific aim of the present paper is to validate the theoretical results on statistical properties of those estimators by numerical experiments. We provide a set of simulations which illustrates their valuable properties via typical examples of lower dimensional random sets. To complete the picture, we have included an additional estimator that naturally derives from the Besicovitch derivation theorem (Ambrosio *et al.*, 2000).

The required background regarding the global and local approximation of mean densities of random closed sets has been presented in a series of papers by Capasso and Villa (Capasso and Villa, 2006; 2007;

2008; Ambrosio *et al.*, 2009; Villa, 2014); we will report here the basic definitions, while for a detailed mathematical analysis of the proposed estimators we refer to the paper by Camerlenghi *et al.* (2014).

In Section *Basics and notation* we introduce basic notations and definitions; in Section *Estimation of $\lambda_{\Theta_n}(x)$* and Section *Optimal bandwidth* we recall our proposed estimators as mentioned above, together with their statistical properties; in Section *Particular cases* we discuss the problem of the optimal bandwidth for the three cases and provide a set of simulations in several cases of interest; finally, in the concluding remarks we offer some comments based on the compared results of the numerical simulations.

BASICS AND NOTATION

We remind that, given a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, a *random closed set* Θ in \mathbb{R}^d is a measurable map $\Theta : (\Omega, \mathfrak{F}) \rightarrow (\mathbb{F}, \sigma_{\mathbb{F}})$, where \mathbb{F} denotes the class of the closed subsets in \mathbb{R}^d , and $\sigma_{\mathbb{F}}$ is the σ -algebra generated by the so called *Fell topology*, or *hit-or-miss topology* (Matheron, 1975). We say that a random closed set $\Theta : (\Omega, \mathfrak{F}) \rightarrow (\mathbb{F}, \sigma_{\mathbb{F}})$ satisfies a certain property (*e.g.*, Θ has Hausdorff dimension n) if Θ satisfies that property \mathbb{P} -a.s.

Here \mathcal{H}^n is the n -dimensional Hausdorff measure, dx stands for $\mathcal{H}^d(dx)$, and $\mathcal{B}_{\mathcal{X}}$ is the Borel σ -algebra of any topological space \mathcal{X} . $B_r(x)$ and b_n will denote

the closed ball with centre x and radius $r > 0$ and the volume of the unit ball in \mathbb{R}^n , respectively. For any function f , $\text{disc}f$ will denote the set of its discontinuity points. We remind that a compact set $A \subset \mathbb{R}^d$ is called *countably \mathcal{H}^n -rectifiable* if there exist countably many n -dimensional Lipschitz graphs $\Gamma_i \subset \mathbb{R}^d$ such that $A \setminus \cup_i \Gamma_i$ is \mathcal{H}^n -negligible. Throughout the paper we shall deal with countably \mathcal{H}^n -rectifiable random closed sets Θ_n . For definitions and basic properties of Hausdorff measure and rectifiable sets (Federer, 1969; Falconer, 1985; Ambrosio *et al.*, 2000).

We briefly recall here that, by means of marked point processes in \mathbb{R}^d with marks in the class of compact subset of \mathbb{R}^d , every random closed set in \mathbb{R}^d can be represented as a *germ-grain model* (Hug *et al.*, 2002; Baddeley *et al.*, 2007, p. 192 and references therein). Therefore, throughout the paper we shall consider random sets Θ described by marked point processes $\Phi = \{(\xi_i, S_i)\}_{i \in \mathbb{N}}$ in \mathbb{R}^d with marks in a suitable mark space \mathbf{K} so that $Z_i = Z(S_i)$, $i \in \mathbb{N}$ is a random set containing the origin (*i.e.*, $Z : \mathbf{K} \rightarrow \mathbb{F}$):

$$\Theta(\omega) = \bigcup_{(x_i, s_i) \in \Phi(\omega)} x_i + Z(s_i), \quad \omega \in \Omega. \quad (1)$$

We assume that Φ has intensity measure $\Lambda(d(x, s)) = f(x, s) dx Q(ds)$ and second factorial moment measure $\nu_{[2]}(d(x, s, y, t)) = g(x, s, y, t) dx dy Q_{[2]}(d(s, t))$ (Karr, 1986; Daley *et al.*, 1988; Stoyan *et al.*, 1995, for general theory of point processes).

For the reader's convenience we have put in the Appendix basic notation and assumptions on Φ , which will appear throughout the paper (see also Villa, 2010; 2014; Camerlenghi *et al.*, 2014 for further details).

Given a RACS Θ_n of integer Hausdorff dimension $n \leq d$, whenever $\mathbb{E}[\mathcal{H}^n(\Theta_n \cap \cdot)]$ is absolutely continuous with respect to the measure \mathcal{H}^d on \mathbb{R}^d , its density (*i.e.*, its Radon-Nikodym derivative) with respect to \mathcal{H}^d has been called *mean density* of Θ_n , and it is denoted by λ_{Θ_n} (for an exhaustive discussion about the existence of λ_{Θ_n} , see Capasso and Villa, 2007; 2008). In the representation via point processes, as in Section *Basics and notation*, we may write (see Villa, 2014)

$$\lambda_{\Theta_n}(x) = \int_{\mathbf{K}} \int_{x-Z(s)} f(y, s) \mathcal{H}^n(dy) Q(ds), \quad (2)$$

where $-Z(s)$ is the reflection of $Z(s)$ at the origin. It is easy to see that if $n = 0$ and $\Theta_0 = X$ is a random vector with pdf f_X , then $\lambda_X(x) = f_X(x)$.

ESTIMATION OF $\lambda_{\Theta_n}(x)$

In the sequel we will assume that an i.i.d. random sample $\Theta_n^1, \dots, \Theta_n^N$ is available for the random closed set Θ_n , with mean density λ_{Θ_n} .

We list here three different kinds of estimators for $\lambda_{\Theta_n}(x)$. (See also Camerlenghi *et al.*, 2014).

A natural estimator

By the Besicovitch derivation theorem (Ambrosio *et al.*, 2000, Theorem 2.22), we know that

$$\lambda_{\Theta_n}(x) = \lim_{r \downarrow 0} \frac{\mathbb{E}[\mathcal{H}^n(\Theta_n \cap B_r(x))]}{b_d r^d} \quad \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d;$$

such approximation suggests the following *natural estimator* for the mean density $\lambda_{\Theta_n}(x)$ of Θ_n , at a point $x \in \mathbb{R}^d$,

$$\widehat{\lambda}_{\Theta_n}^{v, N}(x) := \frac{1}{N b_d r_N^d} \sum_{i=1}^N \mathcal{H}^n(\Theta_n^i \cap B_{r_N}(x)). \quad (3)$$

Here and later the scaling parameter r_N will be called the *bandwidth* associated with the sample size N , as usual in literature.

Kernel estimator

Kernel density estimation was firstly introduced by Parzen (1962) and Rosenblatt (1956) (see also Wertz, 1978, Deheuvels and Mason, 2004 for a survey of additional foundational papers).

We recall here that a measurable function $k : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to be a *multivariate kernel* if it satisfies the following conditions:

1. $0 \leq k(z) \leq M$ for all $z \in \mathbb{R}^d$, for some $M > 0$;
2. k is radially symmetric;
3. $\int_{\mathbb{R}^d} k(z) dz = 1$.

As a natural extension of the kernel density estimation theory for random vectors, the following kernel estimator for the mean density of Θ_n has been introduced in Camerlenghi *et al.* (2014)

$$\begin{aligned} \widehat{\lambda}_{\Theta_n}^{k, N}(x) &:= \frac{1}{N} \sum_{i=1}^N k_{r_N} * \mathcal{H}_{\Theta_n^i}^n(x) \\ &= \frac{1}{N r_N^d} \sum_{i=1}^N \int_{\Theta_n^i} k\left(\frac{x-y}{r_N}\right) \mathcal{H}^n(dy), \end{aligned} \quad (4)$$

where $*$ stands for the usual convolution product.

Remark 1 By choosing the kernel function

$$k(z) = \frac{1}{b_d} \mathbf{1}_{B_1(0)}(z),$$

it is easy to see that we may reobtain the natural density estimator as a particular case of the kernel estimator, i.e., $\widehat{\lambda}_{\Theta_n}^{k,N}(x) = \widehat{\lambda}_{\Theta_n}^{v,N}(x)$.

“Minkowski content”-based estimator

Within the mathematical framework provided in Ambrosio *et al.* (2009) and in Villa (2014, Theorem 7), based on a stochastic version of the Minkowski content notion, it is proved that if Θ_n satisfies (A1), (A2) and (A3), given in the Appendix, then

$$\lambda_{\Theta_n}(x) = \lim_{r \downarrow 0} \frac{\mathbb{P}(x \in \Theta_{n \oplus r})}{b_{d-n} r^{d-n}}, \quad \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d. \quad (5)$$

As a natural byproduct, the following “Minkowski content”-based estimator of $\lambda_{\Theta_n}(x)$ has been proposed in Villa (2014):

$$\widehat{\lambda}_{\Theta_n}^{\mu,N}(x) := \frac{\sum_{i=1}^N \mathbf{1}_{\Theta_n^i \cap B_{r_N}(x) \neq \emptyset}}{N b_{d-n} r_N^{d-n}}. \quad (6)$$

The statistical properties of the “Minkowski content”-based estimator $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$, and of the kernel estimator $\widehat{\lambda}_{\Theta_n}^{k,N}(x)$ (and so of $\widehat{\lambda}_{\Theta_n}^{v,N}(x)$ as well, by Remark 1), can be summarized as follows (we refer to Camerlenghi *et al.*, 2014 for a) and b), and to Villa, 2014, Corollary 13 for c)) :

Theorem 2 Assume that Θ_n satisfies assumptions (A1) and (A2), that $r_N \rightarrow 0$, as $N \rightarrow \infty$, and that k is a kernel with compact support. Then, for almost every $x \in \mathbb{R}^d$,

- a) $\widehat{\lambda}_{\Theta_n}^{k,N}(x)$ is asymptotically unbiased,
- b) $\widehat{\lambda}_{\Theta_n}^{k,N}(x)$ is weakly consistent if $(\overline{A1})$ and $(\overline{A3})$, and

$$\lim_{N \rightarrow \infty} N r_N^{d-n} = \infty \quad (7)$$

hold,

- c) $\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)$ is asymptotically unbiased and weakly consistent if (A3) and Eq. 7 hold.

OPTIMAL BANDWIDTH

A problem of statistical interest is to find an optimal bandwidth r_N . By proceeding along the same lines as what is commonly done for the kernel density

estimator $\widehat{f}_X^N(x)$ of the pdf $f_X(x)$ of a random variable X (where r_N is defined as the quantity which minimizes the asymptotic mean square error (AMSE) of $\widehat{f}_X^N(x)$), in Camerlenghi *et al.* (2014) optimal bandwidths for the proposed estimators have been provided. In order to do this, asymptotic approximations of bias and variance are needed.

OPTIMAL BANDWIDTH FOR $\widehat{\lambda}_{\Theta_n}^{k,N}$

The next theorem provides asymptotic approximations for the bias and the variance of the kernel estimator.

Theorem 3 (Camerlenghi *et al.*, 2014) In addition to the hypotheses in b) of Theorem 2, we assume that the kernel k is continuous and (A2bis) holds for $|\alpha| = 2$. Then, for \mathcal{H}^d -a.e. $x \in \mathbb{R}^d$,

$$\begin{aligned} \text{Bias}(\widehat{\lambda}_{\Theta_n}^{k,N}(x)) &= C_{\text{Bias}}(x) r_N^2 + o(r_N^2) \\ \text{Var}(\widehat{\lambda}_{\Theta_n}^{k,N}(x)) &= \frac{C_{\text{Var}}(x)}{N r_N^{d-n}} + o\left(\frac{1}{N r_N^{d-n}}\right), \end{aligned}$$

with

$$\begin{aligned} C_{\text{Bias}}(x) &:= \sum_{|\alpha|=2} \frac{1}{\alpha!} \int_{\mathbb{R}^d} k(z) z^\alpha dz \\ &\cdot \int_{\mathbf{K}} \int_{x-Z(s)} D_y^\alpha f(y,s) \mathcal{H}^n(dy) Q(ds), \end{aligned}$$

and

$$\begin{aligned} C_{\text{Var}}(x) &:= \int_{\mathbf{K}} \int_{\mathbb{R}^d} \int_{x-Z(s)} \int_{\pi_y^{x,s}} \\ &k(z) k(z+w) f(y,s) \mathcal{H}^n(dw) \mathcal{H}^n(dy) dz Q(ds), \end{aligned}$$

where $\pi_y^{x,s} \in \mathbf{G}_n$ is the approximate tangent space to $x - Z(s)$ at $y \in x - Z(s)$.

For the notion of approximate tangent space to a \mathcal{H}^n -rectifiable compact set A of \mathbb{R}^d at a point $x \in A$ we refer to Ambrosio *et al.* (2000, Definition 2.79).

By remembering that

$$\text{MSE}(\widehat{\lambda}_{\Theta_n}^{k,N}(x)) = [\text{Bias}(\widehat{\lambda}_{\Theta_n}^{k,N}(x))]^2 + \text{Var}(\widehat{\lambda}_{\Theta_n}^{k,N}(x)),$$

it follows that the asymptotic approximation of the mean square error of $\widehat{\lambda}_{\Theta_n}^{k,N}(x)$ is given by

$$\text{AMSE}(\widehat{\lambda}_{\Theta_n}^{k,N}(x)) = C_{\text{Bias}}^2(x) r_N^4 + \frac{1}{N r_N^{d-n}} C_{\text{Var}}(x),$$

so that, for any sufficiently large sample size N ,

$$\begin{aligned} r_N^{\text{o,AMSE}}(x) &:= \arg \min_{r_N} \text{AMSE}(r_N) \\ &= {}^{4+d-n} \sqrt{\frac{(d-n)C_{\text{Var}}(x)}{4NC_{\text{Bias}}^2(x)}}, \quad \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d. \end{aligned} \quad (8)$$

We also remind that a criterion for obtaining a uniform choice of the optimal bandwidth is based on the *integrated mean square error* (MISE) (Hardle, 1991), so defined:

$$\text{MISE}[\widehat{\lambda}_{\Theta_n}^{\kappa,N}(W)] := \int_W \text{MSE}[\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)] dx,$$

for any compact $W \subset \mathbb{R}^d$. Under the same assumptions, by the asymptotic approximation of the *integrated mean square error* (AMISE) of $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(W)$, as $N \rightarrow +\infty$, for any compact $W \subset \mathbb{R}^d$ and any given sufficiently large sample size N , we obtain

$$\begin{aligned} r_{N,W}^{\text{o,AMISE}} &:= \arg \min_{r_N} \text{AMISE}(\widehat{\lambda}_{\Theta_n}^{\kappa,N}(W)) \\ &= {}^{4+d-n} \sqrt{\frac{(d-n) \int_W C_{\text{Var}}(x) dx}{4N \int_W C_{\text{Bias}}^2(x) dx}}. \end{aligned} \quad (9)$$

We observe that the cases in which $C_{\text{Bias}}(x) = 0$, might complicate the identification of an optimal bandwidth (we refer to Camerlenghi *et al.* (2014), and to Schucany (1989) for a more detailed discussion). A case of particular interest where $C_{\text{Bias}}(x) = 0$, is given by assuming that Θ_n is stationary (see Section *The case of stationary Θ_n*).

OPTIMAL BANDWIDTH FOR $\widehat{\lambda}_{\Theta_n}^{\nu,N}$

By Remark 1 we know that $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x) = \widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$ with the kernel $k(z) = \frac{1}{b_d} \mathbf{1}_{B_1(0)}(z)$. The hypothesis of continuity of k in Theorem 3 can be weakened (Camerlenghi *et al.*, 2014), provided that

$$\mathcal{H}_{\perp \pi_y}^n(\text{disc}(k(z+\cdot))) = 0, \quad (10)$$

for any $s \in \mathbf{K}$, $z \in \text{supp}(k)$, and \mathcal{H}^n -a.e. $y \in x - Z(s)$. Such a condition is trivially fulfilled in several cases of interest in applications. Therefore the general formulas for the pointwise and global optimal bandwidth r_N given in Eq. 8 and Eq. 9, respectively, apply for $\widehat{\lambda}_{\Theta_n}^{\nu,N}(x)$ too, provided that $\mathcal{H}_{\perp \pi_y}^n(\text{disc}(\frac{1}{b_d} \mathbf{1}_{B_1(0)}(z+\cdot))) = 0$, for any $s \in \mathbf{K}$, $z \in B_1(0)$, and \mathcal{H}^n -a.e. $y \in x - Z(s)$.

As an example we consider an inhomogeneous Boolean model of segments of the type $[0, l] \times \{0\}$ in \mathbb{R}^2 , with random length $l \sim U(0, L)$ (we have chosen $L = 0.2$ for numerical studies) in the compact window $W = [0, 1]^2$, where the underlying Poisson point process has intensity $f(x_1, x_2) = cx_1^2$. From Eq. 2 it follows that

$$\lambda_{\Theta_1}(x_1, x_2) = \frac{1}{12}L^3c - \frac{1}{3}L^2cx_1 + \frac{1}{2}Lcx_1^2,$$

and by Eq. 8

$$r_N^{\text{o,AMSE}}(x_1, x_2) = \sqrt[5]{\frac{256 \left[\frac{1}{12}L^3 - \frac{1}{3}L^2x_1 + \frac{1}{2}Lx_1^2 \right]}{3NcL^2\pi^2}}.$$

Fig. 1 shows, for $c = 700$, the natural estimator for this type of random closed set as a function of the bandwidth, expressed in pixel (1 pixel = 0.0029). To carry out the numerical experiments, we have studied the estimator at a fixed point (0.5, 0.5) of the compact window $W = [0, 1]^2$.

For $N = 10$, Fig. 1a shows that the choice $r = r_N^{\text{o,AMSE}}$ provides a good estimation of the theoretical mean density; in fact, for this value of the bandwidth, $|\widehat{\lambda}_{\Theta_1}^{\nu,N}(0.5, 0.5) - \lambda_{\Theta_1}(0.5, 0.5)| = 0.2973$. For $N = 100$, Fig. 1b shows that the theoretical optimal value of r is still one of the best choices for the estimation of the mean density. One may notice that the estimation improves with respect to the case $N = 10$; in fact for $r = r_{100}^{\text{o,AMSE}}$ we have $|\widehat{\lambda}_{\Theta_1}^{\nu,N}(0.5, 0.5) - \lambda_{\Theta_1}(0.5, 0.5)| = 0.0614$. We conclude that the optimal bandwidth is one of the best choices for the estimation of the mean density, and as $N \rightarrow \infty$ the estimation improves. Finally we may observe that the natural estimator has good stability properties with respect to the choice of the bandwidth.

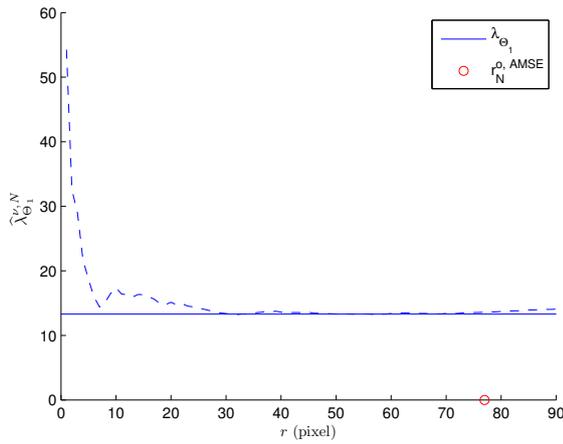
OPTIMAL BANDWIDTH FOR $\widehat{\lambda}_{\Theta_n}^{\mu,N}$

It is not difficult to see that

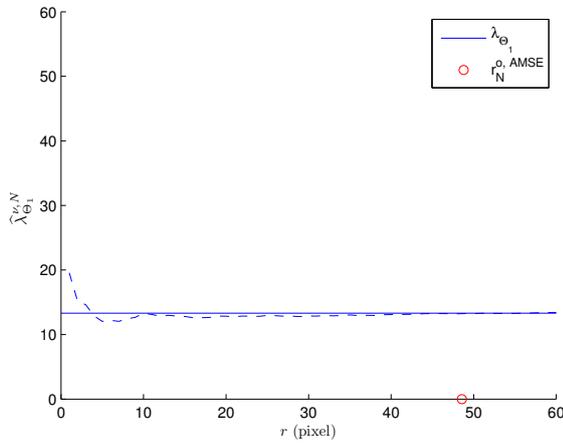
$$\text{Bias}(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = \frac{\mathbb{P}(x \in \Theta_{n \oplus r_N})}{b_{d-n}r^{d-n}} - \lambda_{\Theta_n}(x), \quad (11)$$

$$\text{Var}(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x)) = \frac{\lambda_{\Theta_n}(x)}{Nr_N^{d-n}b_{d-n}} + o\left(\frac{1}{Nr_N^{d-n}}\right). \quad (12)$$

Therefore, it would be necessary to provide a Taylor series expansion of $\text{Bias}(\widehat{\lambda}_{\Theta_n}^{\mu,N}(x))$, or equivalently of $\mathbb{P}(x \in \Theta_{n \oplus r_N})$.



(a)



(b)

Fig. 1. Comparison of the natural estimator and the theoretical value $(\lambda_{\Theta_1}(0.5,0.5) = 13.30)$ at point $(0.5,0.5)$ for an inhomogeneous Boolean model of segments with intensity $f(x_1,x_2) = 700x_1^2$. In (a) $N = 10$; for $r_{10}^{o,AMSE} \approx 77\text{pixel}(0.2973)$ $\hat{\lambda}_{\Theta_1}^{v,N}(0.5,0.5) = 13.5973$. In (b) $N = 100$; for $r_{100}^{o,AMSE} \approx 49\text{pixel}(0.1425)$ $\hat{\lambda}_{\Theta_1}^{v,N}(0.5,0.5) = 13.2386$.

A particular class of germ-grain models Θ_n for which an explicit expression of $\mathbb{P}(x \in \Theta_{n \oplus r_N})$ is available, is the class of Boolean models; in that case we get:

$$\mathbb{P}(x \in \Theta_{n \oplus r_N}) = 1 - \exp \left\{ - \int_{\mathbf{K}} \int_{x-Z(s) \oplus r_N} f(y,s) dy Q(ds) \right\}. \quad (13)$$

For numerical experiments of $\hat{\lambda}_{\Theta_n}^{\mu,N}$ consider the Boolean model of segments analyzed in Section

Optimal bandwidth for $\hat{\lambda}_{\Theta_n}^{v,N}$.

Since a general formula for the optimal bandwidth is not yet available in the literature, we will minimize the asymptotic approximation of the mean square error directly in this particular example. To this aim, a standard calculation of the integral in Eq. 13 leads to:

$$\mathbb{P}(x \in \Theta_{2 \oplus r_N}) = 1 - \exp \left\{ - cLr_N \left(\frac{1}{6}L^2 - \frac{2}{3}Lx_1 + x_1^2 - \frac{1}{6}L\pi r_N + \frac{1}{2} \left(\frac{4}{3}r_N^2 + r_N\pi x_1 \right) \right) \right\}; \quad (14)$$

then, by a Taylor series expansion of the exponential term in Eq. 14, we get that $Bias(\hat{\lambda}_{\Theta_2}^{\mu,N}(x)) = C_B r_N + o(r_N)$, where

$$C_B(x_1,x_2) := -\frac{1}{12}L^2\pi c + \frac{1}{4}\pi x_1 cL - \frac{1}{4}c^2L^2 \left(\frac{1}{6}L^2 - \frac{2}{3}Lx_1 + x_1^2 \right)^2,$$

and so $AMSE(\hat{\lambda}_{\Theta_2}^{\mu,N}(x)) = C_B^2 r_N^2 + \lambda_{\Theta_1}(x_1,x_2)/2Nr_N$; thus it follows

$$r_N^{o,AMSE}(x_1,x_2) = \sqrt[3]{\frac{\lambda_{\Theta_1}(x_1,x_2)}{4NC_B^2(x_1,x_2)}}.$$

Fig. 2 shows the estimator $\hat{\lambda}_{\Theta_n}^{\mu,N}$ at point $(0.5,0.5)$, as a function of the bandwidth r (in pixel) for two different sample sizes ($N = 10$, and $N = 100$). In Fig. 2a, where the sample size N is equal to 10, we can observe that, near the optimal bandwidth, $\hat{\lambda}_{\Theta_1}^{\mu,N}$ provides a very good estimation of the mean density; in fact for the optimal value of r we have $|\hat{\lambda}_{\Theta_1}^{\mu,N}(0.5,0.5) - \lambda_{\Theta_1}(0.5,0.5)| = 1.8556$. In Fig. 2b, where $N = 100$, the estimation improves; indeed $|\hat{\lambda}_{\Theta_1}^{\mu,N}(0.5,0.5) - \lambda_{\Theta_1}(0.5,0.5)| = 0.70$, for r equal to the asymptotic optimal bandwidth. We can conclude that the optimal bandwidth leads to a good estimation of the mean density, which improves as N increases.

Finally observe that as $r \rightarrow +\infty$ the estimator decreases as the function $\frac{1}{2r}$ (Fig. 3), in accordance with the definition of the estimator.

We wish to mention that general results for the optimal bandwidth for inhomogeneous (and not-necessarily Boolean) models of random closed sets are

not yet available in literature.

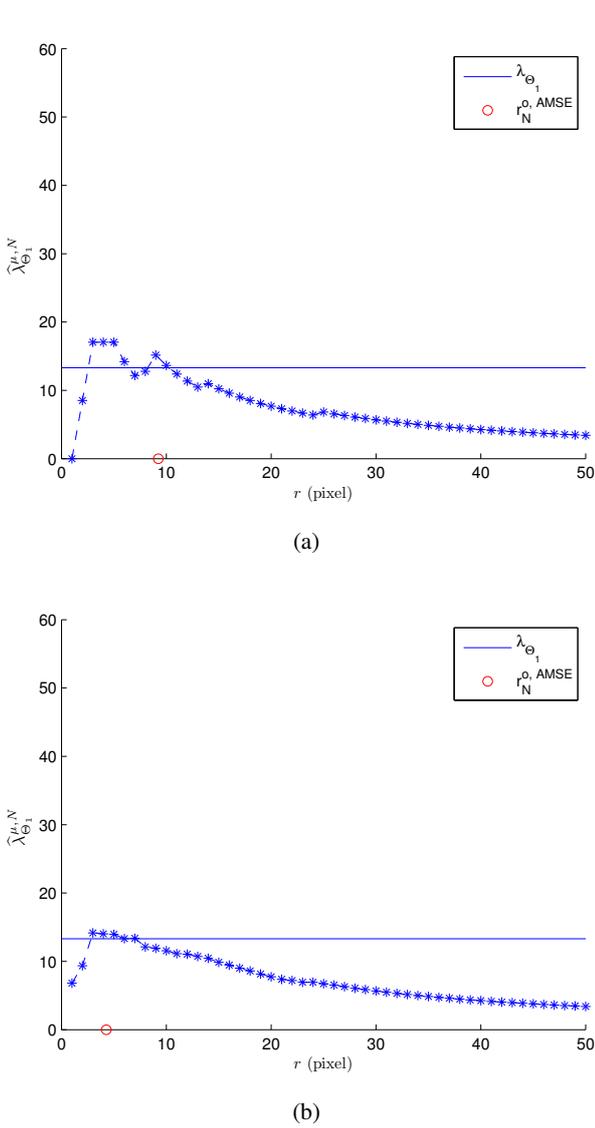


Fig. 2. Comparison of the “Minkowski content”-based estimator and the theoretical value ($\lambda_{\Theta_1}(0.5,0.5) = 13.3$) at point $(0.5,0.5)$ for an inhomogeneous Boolean model of segments with intensity $f(x_1,x_2) = 700x_1^2$. In (a) $N = 10$; for $r_{10}^{o,AMSE} \approx 9 \text{ pixel}(0.0271)$ $\hat{\lambda}_{\Theta_1}^{\mu,N}(0.5,0.5) = 15.1556$. In (b) $N = 100$; for $r_{100}^{o,AMSE} \approx 4 \text{ pixel}(0.0126)$ $\hat{\lambda}_{\Theta_1}^{\mu,N}(0.5,0.5) = 14$.

PARTICULAR CASES

As a confirmation of the validity of our results, in this section we wish to present particular cases (for $n = 0$, and for stationary Θ_n) which have already been treated in literature.

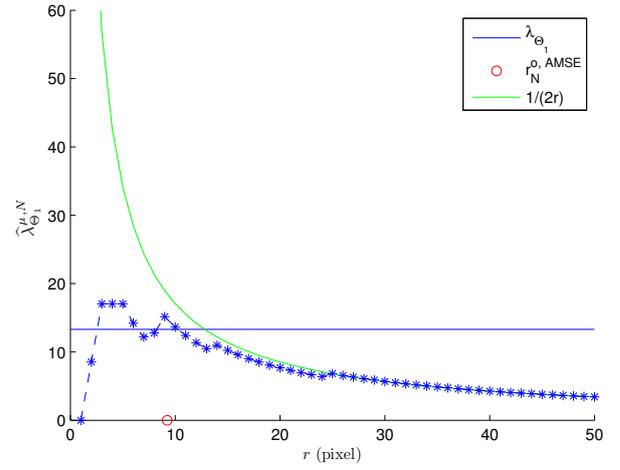


Fig. 3. The “Minkowski content”-based estimator (for $N = 10$) at point $(0.5,0.5)$ for an inhomogeneous Boolean model of segments with intensity $f(x_1,x_2) = 700x_1^2$, compared with the function $\frac{1}{2r}$ (in green).

RANDOM VARIABLES AND POINT PROCESSES ($n=0$)

Let $\Theta_0 \equiv X$ be a continuous random variable with pdf f_X (equivalently, with mean density $\lambda_X = f_X$). In order to apply the above results, X may be considered as the trivial germ-grain process driven by the marked point process $\Phi = \{(X, s)\}$ in \mathbb{R} with mark space $\mathbf{K} = \mathbb{R}^d$, consisting of one point (X) only, with grain $Z(s) := s$, and intensity measure $\Lambda(dy, s) = f(y) dy \delta_0(s) ds$, being δ_0 the usual Dirac delta function in 0. In this case the kernel estimator $\hat{\lambda}_X^{\kappa,N}(x)$ defined in Eq. 4 reduces, as expected, to usual kernel density estimator for random variable well known in classical literature. Known results on the optimal bandwidth (Parzen, 1962; Silverman, 1986; Hardle, 1991, p. 59) follows now as particular case by Eq. 8 and Eq. 9. For a more detailed discussion, see Camerlenghi *et al.* (2014, Section 3.3.1). With regard to the natural estimator $\hat{\lambda}_X^{v,N}(x)$ and the “Minkowski content”-based estimator $\hat{\lambda}_X^{\mu,N}(x)$, we notice that both estimators reduce, in this case, to the usual histogram density estimator, also known in literature as *naive estimator*,

$$\hat{\lambda}_X^{v,N}(x) = \hat{\lambda}_X^{\mu,N}(x) = \frac{1}{N2r_N} \sum_{i=1}^N \mathbf{1}_{[x-r_N, x+r_N]}(X_i),$$

where X_1, \dots, X_N is an i.i.d. random sample for X .

As a more significant example of a random set Θ_0 with dimension $n = 0$, let us consider a point process

Ψ in \mathbb{R}^d with intensity function f_Ψ . In Camerlenghi *et al.* (2014) the following statement has been proven:

Proposition 4 *Let $\{\Psi^i\}_{i \in \mathbb{N}}$ be a sequence of point processes in \mathbb{R}^d , i.i.d. as Ψ , with intensity function $\lambda_\Psi \in C^2$, and locally bounded second moment density g , and let k be a kernel with compact support, continuous at 0. Then the kernel density estimator*

$$\widehat{\lambda}_\Psi^{k,N}(x) = \frac{1}{Nr_N^d} \sum_{i=1}^N \sum_{x_j \in \Psi^i} k\left(\frac{x-x_j}{r_N}\right) \quad (15)$$

of $\lambda_\Psi(x)$ is asymptotically unbiased and weakly consistent for \mathcal{H}^d -a.e. $x \in \mathbb{R}^d$ if r_N is such that

$$\lim_{N \rightarrow \infty} r_N = 0 \quad \text{and} \quad \lim_{N \rightarrow \infty} Nr_N^d = \infty.$$

Moreover, the pointwise optimal bandwidth $r_N^{\text{o,AMSE}}(x)$ is given, for \mathcal{H}^d -a.e. $x \in \mathbb{R}^d$, by

$$r_N^{\text{o,AMSE}}(x) = \sqrt[4+d]{\frac{d\lambda_\Psi(x) \int_{\mathbb{R}^d} k^2(z) dz}{4N \left(\sum_{|\alpha|=2} \frac{1}{\alpha!} D_x^\alpha \lambda_\Psi(x) \int_{\mathbb{R}^d} k(z) z^\alpha dz \right)^2}},$$

while the global optimal bandwidth $r_{N,W}^{\text{o,AMISE}}$ is given, for any compact window $W \subset \mathbb{R}^d$, by

$$r_{N,W}^{\text{o,AMISE}} = \sqrt[4+d]{\frac{d\mathbb{E}[\Psi(W)] \int_{\mathbb{R}^d} k^2(z) dz}{4N \int_W \left(\sum_{|\alpha|=2} \frac{1}{\alpha!} D_y^\alpha \lambda_\Psi(x) \int_{\mathbb{R}^d} k(z) z^\alpha dz \right)^2 dx}}.$$

Note that by choosing $k(z) := \frac{1}{b_d} \mathbf{1}_{B_1(0)}(z)$ in Eq. 15, with $N = 1$, we reobtain the well known classic and widely used Berman-Diggle estimator (Diggle, 1985; Berman and Diggle, 1989; van Lieshout, 2012)

$$\widehat{\lambda}_\Psi^{k,N}(x) = \frac{\Psi(B_r(x))}{b_d r^d}.$$

In order to show numerical results, let Ψ be a Poisson point process in \mathbb{R}^2 with intensity function $\lambda_\Psi(x_1, x_2) = x_1^2 + x_2^2$. We want to estimate the intensity function of Ψ in the compact window $W := [-2, 2]^2$, by means of $\widehat{\lambda}_\Psi^{k,N}(x)$ and $\widehat{\lambda}_\Psi^{v,N}(x)$, for a sample size $N = 1000$.

Fig. 4a shows the first estimator, where we have adopted the kernel of Epanechnikov:

$$k(t) = \begin{cases} \frac{2}{\pi}(1-x_1^2-x_2^2), & \text{if } (x_1, x_2) \in B_1(0) \\ 0, & \text{otherwise} \end{cases}$$

and the optimal bandwidth at each point of estimation, that is

$$r_N^{\text{o,AMSE}}(x_1, x_2) = \sqrt[6]{\frac{6(x_1^2 + x_2^2)}{N\pi}};$$

on the other hand, by Proposition 4, it is easy to see that the uniform optimal bandwidth at all points in W is

$$r_{N,W}^{\text{o,AMISE}} = \sqrt[6]{\frac{64}{N\pi}}.$$

We would like to compare the kernel estimation at (1.8, 1.8), obtained by employing the optimal bandwidth at this point ($\widehat{\lambda}_{\Psi,o}^{k,N}(1.8, 1.8)$), and the corresponding estimation generated by using the uniform optimal bandwidth in W ($\widehat{\lambda}_{\Psi,u}^{k,N}(1.8, 1.8)$). At point (1.8, 1.8) the theoretical value of the intensity function is $\lambda_\Psi(1.8, 1.8) = 6.48$, the optimal bandwidth is $r_{1000}^{\text{o,AMSE}} = 0.4809$ and the corresponding kernel estimation is $\widehat{\lambda}_{\Psi,o}^{k,N}(1.8, 1.8) = 6.4636$ ($|\widehat{\lambda}_{\Psi,o}^{k,N}(1.8, 1.8) - \lambda(1.8, 1.8)| = 0.0164$); instead the uniform optimal bandwidth is $r_{1000,W}^{\text{o,AMISE}} = 0.5226$, and the corresponding estimation $\widehat{\lambda}_{\Psi,u}^{k,N}(1.8, 1.8) = 6.4350$ ($|\widehat{\lambda}_{\Psi,u}^{k,N}(1.8, 1.8) - \lambda(1.8, 1.8)| = 0.045$). Both estimations are accurate, but the first one is better, since it employs the optimal bandwidth at the fixed point (1.8, 1.8).

Fig. 4b shows the natural estimator in the compact window W , generated by employing the theoretical optimal bandwidth at each point in the compact window:

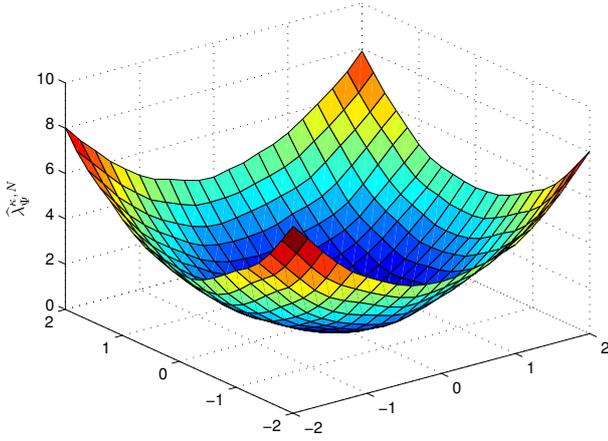
$$r_N^{\text{o,AMSE}}(x_1, x_2) = \sqrt[6]{\frac{2(x_1^2 + x_2^2)}{N\pi}};$$

as in the previous case, it is easy to obtain the uniform optimal bandwidth:

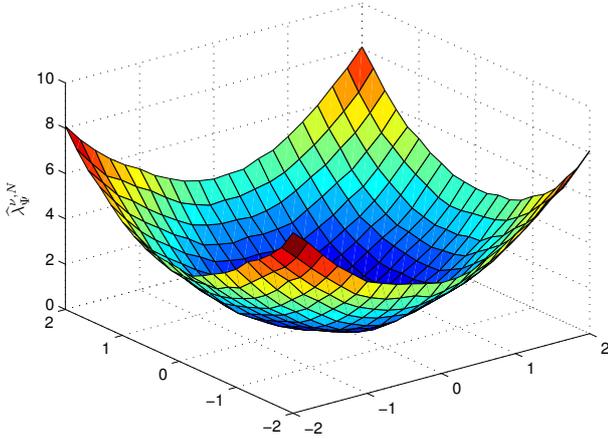
$$r_{N,W}^{\text{o,AMISE}} = \sqrt[6]{\frac{16}{3N\pi}}.$$

As before, we will analyze the behavior of $\widehat{\lambda}_\Psi^{v,N}(1.8, 1.8)$, obtained by employing the optimal bandwidth in the point (1.8, 1.8), and the same estimation generated by using the uniform optimal bandwidth in W . At the chosen point (1.8, 1.8), the optimal bandwidth is $r_{1000}^{\text{o,AMSE}} = 0.4005$ and the corresponding natural estimation $\widehat{\lambda}_{\Psi,o}^{v,N}(1.8, 1.8) =$

6.4821 ($|\widehat{\lambda}_{\Psi,o}^{v,N}(1.8, 1.8) - \lambda(1.8, 1.8)| = 0.0021$); the uniform optimal bandwidth is $r_{1000,W}^{o,AMISE} = 0.3454$, and the corresponding estimation $\widehat{\lambda}_{\Psi,u}^{v,N}(1.8, 1.8) = 6.5133$ ($|\widehat{\lambda}_{\Psi,u}^{v,N}(1.8, 1.8) - \lambda(1.8, 1.8)| = 0.0333$). As before, the first estimation is more accurate, since it employs the optimal bandwidth at the relevant point.



(a)



(b)

Fig. 4. Estimators for the intensity function of a Poisson point process, for $N = 1000$, for two different kernels. (a) shows $\widehat{\lambda}_{\Psi}^{k,N}(x_1, x_2)$, where k is the kernel of Epanechnikov and we have used the optimal bandwidth $r_N^{o,AMISE}(x_1, x_2)$ at each point. (b) shows the natural estimator $\widehat{\lambda}_{\Psi}^{v,N}(x_1, x_2)$; here we have used the optimal bandwidth $r_N^{o,AMISE}(x)$ at each point.

THE CASE OF STATIONARY Θ_n

Let Θ_n be stationary; we assume here that, in the point process representation, $\Phi = \{(x_i, s_i)\}_{i \in \mathbb{N}}$ is an independent marking of the marginal process $\{x_i\}_{i \in \mathbb{N}}$, which is itself stationary, so that $\Lambda(d(x, s)) = c dx Q(ds)$, i.e., $f(x, s) \equiv c$, for any $(x, s) \in \mathbb{R}^d \times \mathbf{K}$. Thus, $\lambda_{\Theta_n}(x) \equiv c \mathbb{E}[\mathcal{H}^n(Z)]$ for \mathcal{H}^d -a.e. $x \in \mathbb{R}^d$, and the optimal bandwidth r_N associated with the proposed estimators will be independent of x as well.

Optimal bandwidth for $\widehat{\lambda}_{\Theta_n}^{k,N}$ and $\widehat{\lambda}_{\Theta_n}^{v,N}$

We point out that in the stationary case a kernel type estimation would be irrelevant, since the intensity of the point process is constant; though we treat this case too in order to show the full compatibility of our approach with the standard one, in which we may just take global “means” in the observation window (see, e.g., Beneš and Rataj (2004) and the next paragraphs for further details).

In (Camerlenghi *et al.*, 2014) the following implications have been shown:

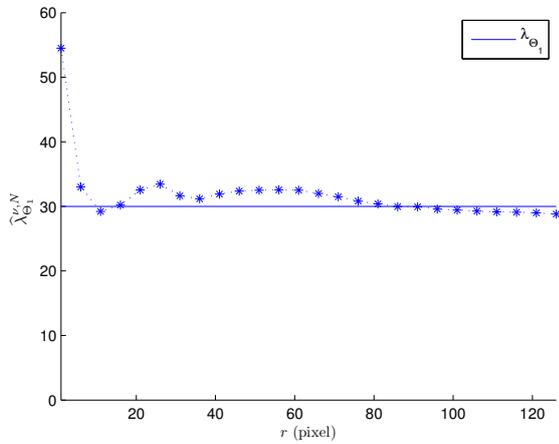
- (A1) $\Rightarrow \text{Bias}(\widehat{\lambda}_{\Theta_n}^{k,N}(x)) = 0$ for any bandwidth $r > 0$, and any sample size N ;
- $(\overline{\text{A1}})$ and $(\overline{\text{A3}}) \Rightarrow \widehat{\lambda}_{\Theta_n}^{k,N}$ is strongly consistent for \mathcal{H}^d -a.e. $x \in \mathbb{R}^d$, as $N \rightarrow \infty$.

It is worth noting that whenever Θ_n is a Boolean model such that $\mathbb{E}[(\mathcal{H}^n(Z))^2] < \infty$, and the kernel k is assumed to be continuous in the interior of its support, then $r^{o,MSE} = +\infty$ (Camerlenghi *et al.*, 2014).

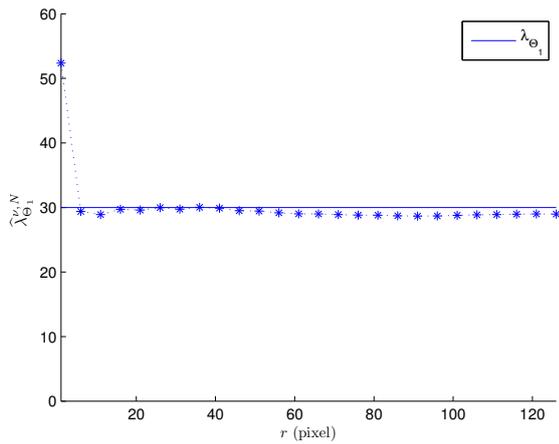
The same conclusions hold for $\widehat{\lambda}_{\Theta_n}^{v,N}$ too, by choosing $k(z) := \frac{1}{b_d} \mathbf{1}_{B_1(0)}(z)$. This is in accordance with both intuition and known results in literature for the optimal bandwidth of the kernel estimators of the intensity of homogeneous Poisson point processes. In particular, if W is the *observation window* of any realization of a homogeneous Poisson point process Ψ in \mathbb{R}^d (and so $N = 1$), and $|W|$ its volume, being $\widehat{\lambda}_{\Psi}^{v,1}(x) \stackrel{(3)}{=} \frac{1}{b_d r^d} \Psi(B_r(0))$ for any $x \in \mathbb{R}^d$, we reobtain that the best unbiased estimator of the intensity λ_{Ψ} of Ψ is given by (taking $r^{o,MSE} = +\infty$) $\widehat{\lambda}_{\Psi} = \Psi(W)/|W|$, with $|W| \rightarrow \infty$.

In order to carry out numerical experiments for $\widehat{\lambda}_{\Theta_n}^{v,N}$ in the stationary case, we have considered a Boolean model of segments of the type $[0, l] \times \{0\}$ with random length $l \sim U(0, L)$, where $L = 0.2$, in the compact window $W = [0, 1]^2$. Furthermore, assume that the underlying Poisson point process has constant intensity $f(x_1, x_2) = c > 0$. It is obvious that $\lambda_{\Theta_1} = cL/2$.

Fig. 5 shows the natural estimator at the fixed point $(0.5,0.5)$, for different values of the bandwidth r (in pixel). Since the optimal bandwidth is $+\infty$, we expect that, as r grows to infinity, the estimation improves, which is confirmed: the estimator seems to stabilize after a certain value of r . By comparing Fig. 5a, where $N = 10$, and b, where $N = 100$, the estimation improves as the sample size grows to infinity; in fact for $N = 100$ the value after which the estimator stabilizes is less than the corresponding value for $N = 10$.



(a)



(b)

Fig. 5. Comparison between the natural estimation and the theoretical value at the point $(0.5,0.5)$ for a homogeneous Boolean model of segments with intensity $f(x_1, x_2) = 300$. In (a) $N = 10$; in (b) $N = 100$. The optimal bandwidth is $+\infty$.

5.2.2 Optimal bandwidth for $\hat{\lambda}_{\Theta_n}^{\mu, N}$

We denote by $\Phi_i(A)$ the i -th total curvature measure of any compact set $A \subset \mathbb{R}^d$ with positive reach, as introduced in Federer (1959), for $i = 0, \dots, d-1$.

Proposition 5 (Camerlenghi et al., 2014) Let Θ_n be a Boolean model with intensity measure $\Lambda(d(y, s)) = c dy Q(ds)$, satisfying Assumption (A1), and such that, for any $s \in \mathbf{K}$, $\text{reach}Z(s) > R$, for some $R > 0$. Let us assume also that $\mathbb{E}[\Phi_i(Z)] < \infty$ for all $i = 0, \dots, n-1$. Then, the optimal bandwidth associated with $\hat{\lambda}_{\Theta_n}^{\mu, N}$ is given by

$$r_N^{\text{o,AMSE}} := \begin{cases} \sqrt[3]{\frac{c\mathbb{E}[\mathcal{H}^n(Z)]}{N(\pi c\mathbb{E}[\Phi_{n-1}(Z)] - 2(c\mathbb{E}[\mathcal{H}^n(Z)])^2)^2}} & \text{if } d-n=1, \\ \sqrt[3]{\frac{(d-n)b_{d-n}c\mathbb{E}[\mathcal{H}^n(Z)]}{2N(cb_{d-n+1}\mathbb{E}[\Phi_{n-1}(Z)])^2}} & \text{if } d-n>1, \end{cases} \quad (16)$$

independent of $x \in \mathbb{R}^d$.

To understand further the behavior of $\hat{\lambda}_{\Theta_n}^{\mu, N}$ in the stationary case, consider the Boolean model of segments of the previous section. It is easy to calculate the optimal bandwidth, that is:

$$r_N^{\text{o,AMSE}} = \sqrt[3]{\frac{c\mathbb{E}L}{N(c\pi - 2c^2(\mathbb{E}L)^2)}}. \quad (17)$$

Fig. 6 shows $\hat{\lambda}_{\Theta_n}^{\mu, N}(0.5,0.5)$ as a function of the bandwidth r (in pixel). In Fig. 6a ($N = 10$) observe that for $r = r_N^{\text{o,AMSE}}$ the estimation approaches very well the theoretical value of the mean density, even if N is low ($|\hat{\lambda}_{\Theta_1}^{\mu, 10}(0.5,0.5) - \lambda_{\Theta_1}(0.5,0.5)| = 2.72$). In Fig. 6b ($N = 100$) the best value of the estimation is achieved when r is equal to the theoretical optimal bandwidth, furthermore as N grows the estimation improves; in fact, for $r = r_{100}^{\text{o,AMSE}}$, $|\hat{\lambda}_{\Theta_1}^{\mu, 100}(0.5,0.5) - \lambda_{\Theta_1}(0.5,0.5)| = 1.5833$. We can conclude that the estimator is optimal for the choice $r = r_N^{\text{o,AMSE}}$ and the estimation improves as the dimension of the sample size diverges to infinity. Finally, observe that as $r \rightarrow +\infty$, the estimator decreases as $\frac{1}{2r}$, which is the same conclusion we have reached for the inhomogeneous case.

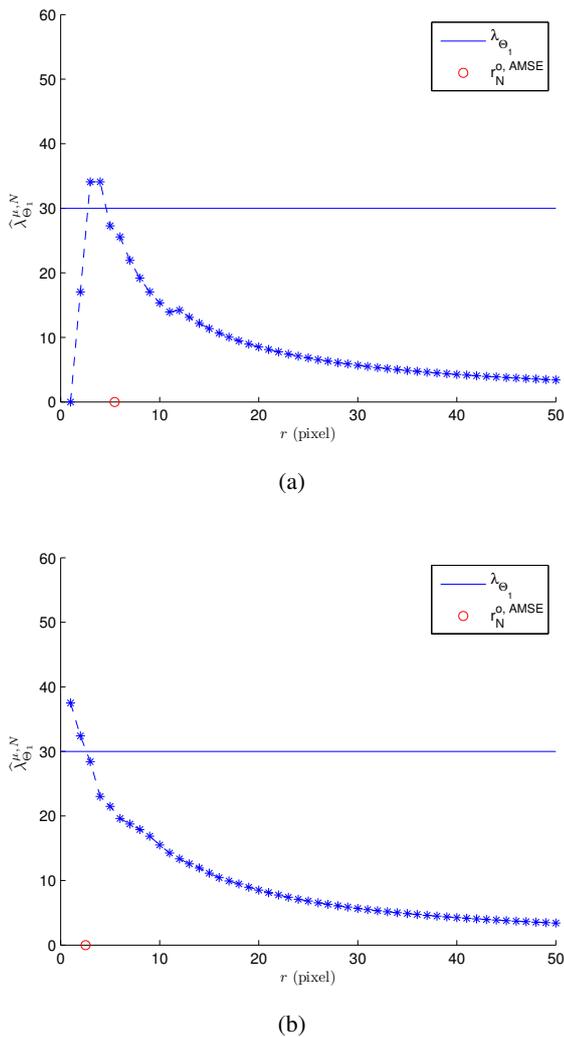


Fig. 6. Comparison of the “Minkowski content”-based estimator and the theoretical value ($\lambda_{\Theta_1}(0.5,0.5) = 30$) at the point $(0.5,0.5)$ for a homogeneous Boolean model of segments with intensity $f(x_1, x_2) = 300$. In (a) we have chosen $N = 10$; for $r_{10}^{o,AMSE} \approx 5 \text{ pixel}(0.016)$, we have $\hat{\lambda}_{\Theta_1}^{\mu,N}(0.5,0.5) = 27.28$. In (b) we have chosen $N = 100$, for $r_{100}^{o,AMSE} \approx 3 \text{ pixel}(0.0074)$, we have $\hat{\lambda}_{\Theta_1}^{\mu,N}(0.5,0.5) = 28.4167$.

CONCLUDING REMARKS

Based on the numerical simulations we may now offer here some comparison about the computational advantages / disadvantages of the estimators proposed by the authors in Villa (2014), and Camerlenghi *et al.* (2014).

From a purely computationally point of view, it emerges the “Minkowski content”-estimator as

the most treatable, as one may easily realize by considering that for this estimator we just need to count relevant pixels of the random object (Eq. 6), while for the kernel estimator a, generally nontrivial, computation of integrals is required (Eq. 4). This is the main reason why we have reduced our numerical simulation to the sole point process case.

The natural estimator, which is a particular case of the kernel estimator, seems to be computationally easier to handle; further for point processes the choice of the kernel does not seem to be of much influence. The stationary case has been extensively studied in the literature. It is worth noticing that the optimal bandwidth for a generic kernel estimator is infinity, whenever Θ_n is a stationary Boolean model, in accordance to well known results in the literature. In applied problems, an infinite optimal bandwidth is equivalent to the choice of an observation window as large as possible.

As far as the behaviour of the proposed estimators with respect to the choice of the bandwidth is concerned, we have in particular realized that the natural estimator results to be more stable; *i.e.*, the “Minkowski content”-based estimators are quite sensitive to the choice of the bandwidth, while for kernel estimators it is only important that the bandwidth has the correct order of magnitude (Fig. 1-2).

For the time being we have not yet taken into account possible edge effects, which require further analysis.

ACKNOWLEDGMENTS

FC and EV are members of the Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM); VC wishes to acknowledge the continued collaboration with ADAMSS, the Milan Interdisciplinary Centre for Advanced Applied Mathematical and Statistical Sciences.

It is our duty and a pleasure to acknowledge the anonymous referees for the accurate reading of the paper, and their valuable comments and suggestions which lead to an effective improvement of the presentation of our results.

REFERENCES

Ambrosio L , Capasso V, Villa E (2009). On the approximation of mean densities of random closed sets. Bernoulli 15:1222–42.

- Ambrosio L, Fusco N, Pallara D (2000). Functions of bounded variation and free discontinuity problems. Oxford: Clarendon Press.
- Baddeley A, Barany I, Schneider R, Weil W (2007). Stochastic geometry. Lect Not Math 1982. Berlin: Springer.
- Beneš V, Rataj J (2004). Stochastic geometry: selected topics. Dordrecht: Kluwer.
- Berman M, Diggle P (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *J Roy Statist Soc B* 51:81–92.
- Camerlenghi F, Capasso V, Villa E (2014). On the estimation of the mean density of random closed sets. *J Multivariate Anal* 125:65–88.
- Capasso V, Villa E (2006). On the continuity and absolute continuity of random closed sets. *Stoch An Appl* 24:381–97.
- Capasso V, Villa E (2007). On mean densities of inhomogeneous geometric processes arising in material science and medicine. *Image Anal Stereol* 26:23–36.
- Capasso V, Villa E (2008). On the geometric densities of random closed sets. *Stoch Anal Appl* 26:784–808.
- Daley DJ, Vere-Jones D (1988). An introduction to the theory of point processes. *Probab Appl* 2003. New York: Springer.
- David G, Semmes S (1997). Fractured fractals and broken dreams-self-similar geometry through metric and measure. London: Oxford Univ Press.
- Deheuvels P, Mason DM (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat Inference Stoch Process* 7:225–77.
- Diggle PJ (1985). A kernel method for smoothing point process data. *Appl Statist* 34:138–47.
- Falconer KJ (1985). , The geometry of fractal sets. Cambridge: Cambridge University Press.
- Falconer KJ (2004). One-sided multifractal analysis and points of non-differentiability of devil’s staircases. *Math Proc Cambridge Philos Soc* 136:167–74.
- Federer H (1959). Curvature measures. *Trans Amer Math Soc* 93:418–91.
- Federer H (1969). Geometric measure theory. Berlin: Spriger.
- Härdle W (1991). Smoothing techniques with implementation in S. New York: Springer-Verlag.
- Hug D, Last G, Weil W (2002). A survey on contact distributions. In: Mecke KR, Stoyan D, Eds. Morphology of condensed matter. Physics and geometry of spatially complex systems. Lect Not Phys 600:317–57. Berlin: Springer.
- Hug D, Last G, Weil W (2004). A local Steiner-type formula for general closed sets and applications. *Math Z* 246:237–72.
- Karr AF (1986). Point processes and their statistical inference. New York: Marcel Dekker.
- Matheron G (1975). Random sets and integral geometry. New York: John Wiley & Sons.
- Parzen E (1962). On the estimation of a probability density function and the mode. *Ann Math Statist* 33:1065–76.
- Rosenblatt M (1956). Remarks on some nonparametric estimates of a density function. *Ann Math Statist* 27:832–7.
- Schucany WR (1989). Locally optimal window widths for kernel density estimation with large samples. *Statist Probab Lett* 7:401–5.
- Silverman BW (1986). Density estimation for statistics and data analysis. London: Chapman & Hall.
- Stoyan D, Kendall WS, Mecke J (1995). Stochastic geometry and its application. New York: John Wiley & Sons.
- van Lieshout MNM (2012). On estimation of the intensity function of a point process. *Meth Comput Appl Probab* 14:567-78.
- Villa E (2010). Mean densities and spherical contact distribution function of inhomogeneous Boolean models. *Stoch An Appl* 28:480–504.
- Villa E (2014) On the local approximation of mean densities of random closed sets. *Bernoulli* 20:1–27
- Wertz W (1978). Statistical density estimation: a survey. *Angewandte Statistik und Ökonometrie [Applied statistics and econometrics]* 13 Göttingen: Vandenhoeck & Ruprecht.

APPENDIX

To fix the notation, $\alpha := (\alpha_1, \dots, \alpha_d)$ will be a multi-index of \mathbb{N}_0^d ; we will denote

$$\begin{aligned} |\alpha| &:= \alpha_1 + \dots + \alpha_d \\ \alpha! &:= \alpha_1! \dots \alpha_d! \\ y^\alpha &:= y_1^{\alpha_1} \dots y_d^{\alpha_d} \end{aligned}$$

$$D_y^\alpha f(y, s) := \frac{\partial^{|\alpha|} f(y, s)}{\partial y_1^{\alpha_1} \dots \partial y_d^{\alpha_d}};$$

furthermore, for all $s \in \mathbf{K}$, we will denote

$$\mathcal{D}^{(\alpha)}(s) := \text{disc}(D_y^\alpha f(y, s)), \quad \mathcal{D}(s) := \text{disc}(f(\cdot, s)).$$

Furthermore, we list here the assumptions on Φ which have been adopted in the text.

(A1) for any $(y, s) \in \mathbb{R}^d \times \mathbf{K}$, $y + Z(s)$ is a countably \mathcal{H}^n -rectifiable and compact subset of \mathbb{R}^d , such that there exists a closed set $\Xi(s) \supseteq Z(s)$ such that

$$\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) Q(ds) < \infty \text{ and}$$

$$\mathcal{H}^n(\Xi(s) \cap B_r(x)) \geq \gamma r^n \quad \forall x \in Z(s), \forall r \in (0, 1) \quad (18)$$

for some $\gamma > 0$ independent of s ;

(A1) as (A1), replacing (18) with

$$\gamma r^n \leq \mathcal{H}^n(\Xi(s) \cap B_r(x)) \leq \tilde{\gamma} r^n \quad \forall x \in Z(s), r \in (0, 1)$$

for some $\gamma, \tilde{\gamma} > 0$ independent of s ;

(A2) for any $s \in \mathbf{K}$, $\mathcal{H}^n(\text{disc}(f(\cdot, s))) = 0$ and $f(\cdot, s)$ is locally bounded such that for any compact $K \subset \mathbb{R}^d$

$$\sup_{x \in K_{\oplus \text{diam}(Z(s))}} f(x, s) \leq \tilde{\xi}_K(s)$$

for some $\tilde{\xi}_K(s)$ with

$$\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) \tilde{\xi}_K(s) Q(ds) < \infty;$$

(A2bis) for any $s \in \mathbf{K}$, $\mathcal{H}^n(\mathcal{D}^{(\alpha)}(s)) = 0$ and $D_y^\alpha f(y, s)$ is locally bounded such that, for any compact $C \subset \mathbb{R}^d$,

$$\sup_{y \in C_{\oplus \text{diam}Z(s)}} |D_y^\alpha f(y, s)| \leq \tilde{\xi}_C^{(\alpha)}(s)$$

for some $\tilde{\xi}_C^{(\alpha)}(s)$ with

$$\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) \tilde{\xi}_C^{(\alpha)}(s) Q(ds) < \infty;$$

(A3) for any $(s, y, t) \in \mathbf{K} \times \mathbb{R}^d \times \mathbf{K}$, $\mathcal{H}^n(\text{disc}(g(\cdot, s, y, t))) = 0$ and $g(\cdot, s, y, t)$ is locally bounded such that for any compact $K \subset \mathbb{R}^d$ and $a \in \mathbb{R}^d$,

$$\mathbf{1}_{(a-Z(t))_{\oplus 1}}(y) \sup_{x \in K_{\oplus \text{diam}(Z(s))}} g(x, s, y, t) \leq \xi_{a, K}(s, y, t)$$

for some $\xi_{a, K}(s, y, t)$ with

$$\int_{\mathbb{R}^d \times \mathbf{K}^2} \mathcal{H}^n(\Xi(s)) \xi_{a, K}(s, y, t) dy Q_{[2]}(ds, dt) < \infty. \quad (19)$$

(A3) for any $s, t \in \mathbf{K}$, $g(\cdot, s, \cdot, t)$ is locally bounded such that, for any $C, \bar{C} \subset \mathbb{R}^d$ compact sets:

$$\sup_{y \in \bar{C}_{\oplus \text{diam}Z(t)}} \sup_{x \in C_{\oplus \text{diam}Z(s)}} g(x, s, y, t) \leq \xi_{C, \bar{C}}(s, t)$$

for some $\xi_{C, \bar{C}}(s, t)$ with

$$\int_{\mathbf{K}^2} \mathcal{H}^n(\Xi(s)) \mathcal{H}^n(\Xi(t)) \xi_{C, \bar{C}}(s, t) Q_{[2]}(ds, dt) < \infty. \quad (20)$$

For a discussion on the above assumptions and on how they simplify in certain particular cases, for instance whenever Θ_n is a Boolean model (*i.e.*, Φ is an independently marked Poisson point process), see Villa (2014, Sec. 3.1) and Camerlenghi *et al.* (2014).