ON THE ESTIMATION OF THE MEAN DENSITY OF RANDOM CLOSED SETS

F. Camerlenghi[a], V. Capasso[b], E. Villa[b,*]

[a]*Dept. of Mathematics, Università degli Studi di Pavia, via Ferrata 1, 27100 Pavia, Italy*
[b]*Dept. of Mathematics, Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy*

Abstract

Many real phenomena may be modelled as random closed sets in $\mathbb{R}^d$, of different Hausdorff dimensions. Of particular interest are cases in which their Hausdorff dimension, say n, is strictly less than d, such as fiber processes, boundaries of germ-grain models, and n-facets of random tessellations. A crucial problem is the estimation of pointwise mean densities of absolutely continuous, and spatially inhomogeneous random sets, as defined by the authors in a series of recent papers. While the case $n = 0$ (random vectors, point processes, etc.) has been, and still is, the subject of extensive literature, in this paper we face the general case of any $n < d$; pointwise density estimators which extend the notion of kernel density estimators for random vectors are analyzed, together with a previously proposed estimator based on the notion of Minkowski content. In a series of papers, the authors have established the mathematical framework for obtaining suitable approximations of such mean densities. Here we study the unbiasedness and consistency properties, and identify optimal bandwidths for all proposed estimators, under sufficient regularity conditions. We show how some known results in literature follow as particular cases. A series of examples throughout the paper, both non-stationary, and stationary, are provided to illustrate various relevant situations.

Keywords: density estimator, kernel estimate, stochastic geometry, random closed set, Hausdorff dimension, Minkowski content
2010 MSC: Primary 62G07; Secondary 60D05, 28A75

1. Introduction

Given an Euclidean space $\mathbb{R}^d$, the problem of the evaluation and the estimation of the mean density of lower dimensional random closed sets (i.e. with Hausdorff dimension less than d), such as fibre processes and surfaces of full dimensional random sets, has been of great interest in many different scientific and technological fields over the last decades [7, 18]; recent areas of interest include pattern recognition and image analysis [43, 24], computer vision [48], medicine [1, 13, 14, 15], material science [12], etc.

*Corresponding author
Email addresses: `federico.camerlenghi01@ateneopv.it` (F. Camerlenghi), `vincenzo.capasso@unimi.it` (V. Capasso), `elena.villa@unimi.it` (E. Villa)

The papers [11, 16] offer examples of the intrinsic relevance of local approximation of mean densities of random closed sets with lower Hausdorff dimension in stochastic homogenization problems arising in applications.

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

$$\Theta : (\Omega, \mathfrak{F}) \longrightarrow (\mathbb{F}, \sigma_{\mathbb{F}}),$$

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*, that is the topology generated by the set system

$$\{\mathcal{F}_G : G \in \mathcal{G}\} \cup \{\mathcal{F}^C : C \in \mathcal{C}\}$$

where $\mathcal{G}$ and $\mathcal{C}$ are the system of the open and compact subsets of $\mathbb{R}^d$, respectively (e.g., see [38]). We say that a random closed set $\Theta : (\Omega, \mathfrak{F}) \to (\mathbb{F}, \sigma_{\mathbb{F}})$ satisfies a certain property (e.g., Θ has Hausdorff dimension n) if Θ satisfies that property $\mathbb{P}$-a.s.; throughout the paper we shall deal with countably $\mathcal{H}^n$-rectifiable random closed sets (we denote by $\mathcal{H}^n$ the n-dimensional Hausdorff measure).
Let Θ_n be a set of locally finite $\mathcal{H}^n$-measure; then it induces a random measure μ_{Θ_n} defined by

$$\mu_{\Theta_n}(A) := \mathcal{H}^n(\Theta_n \cap A), \qquad A \in \mathcal{B}_{\mathbb{R}^d},$$

($\mathcal{B}_{\mathbb{R}^d}$ is the Borel σ-algebra of $\mathbb{R}^d$), and the corresponding expected measure

$$\mathbb{E}[\mu_{\Theta_n}](A) := \mathbb{E}[\mathcal{H}^n(\Theta_n \cap A)], \qquad A \in \mathcal{B}_{\mathbb{R}^d}.$$

For a discussion of the measurability of the random variables $\mu_{\Theta_n}(A)$, we refer to [6, 55].
Whenever the measure $\mathbb{E}[\mu_{\Theta_n}]$ 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. In this case we say that the random set Θ_n is *absolutely continuous in mean*, and we shall denote its mean density by λ_{Θ_n} [17, 19].

The aim of the present paper consists of providing a rigorous mathematical background for the estimation of the mean density of a random closed set Θ_n, of Hausdorff dimension n less than d, based on an i.i.d. sample $\Theta_n^1, ..., \Theta_n^N$ for Θ_n. In particular we will analyze two different mean density estimators and their statistical properties, the first of which is a direct extension of the kernel estimators of probability densities of random vectors, while the second one, already introduced in [51] and [52], is based on the notion of n-dimensional Minkowski content of sets.
We have felt of interest to report here a discussion about an additional density estimator that naturally derives from the Besicovitch derivation theorem (see e.g. [4]); anyway we observe that it can be seen as a particular case of the kernel estimator.

As in the classical literature referring to the case of random variables, we have paid a particular attention to the identification of an optimal bandwidth, for a given sample size N.
We will show how the theory developed here extends the classical one for absolutely continuous random variables and random vectors [40, 42] (for a general treatment see e.g. [45, 46, 9]), and for point processes (see e.g. [26], [20, page 629], and the recent paper [50]). See also [23] and [54] for a survey of additional foundational papers.

The required mathematical background regarding the global and local approximation of mean densities of random closed sets has been carried out with great detail in a series of papers by Capasso and Villa (see [2, 17, 18, 19, 52], and references therein).

In Section 2 we recall basic definitions and relevant notations, leaving to the Appendix a concise account of basic classical results. In Section 3 we present the main statistical properties of the kernel type estimator, and face the problem of the identification of an optimal bandwidth. Section 4 is devoted to the "Minkowski content"-based estimator. For a better readability of the main results, we have left the proofs of the main theorems to Section 7. As a simple example of applicability of the results presented here, we consider in Section 5 an inhomogeneous Boolean model of segments already introduced in previous literature (see e.g. [52, Example 2] and [7, page 86]); we provide explicit expressions of the optimal bandwidth r_N associated both to $\widehat{\lambda}^{\nu,N}_{\Theta_n}$ and to $\widehat{\lambda}^{\mu,N}_{\Theta_n}$. Hints for further analysis are provided in the concluding remarks (Section 6).

2. Basic notation and definitions

Throughout the paper $\mathcal{H}^n$ is the n-dimensional Hausdorff measure, $\mathrm{d}x$ stands for $\mathcal{H}^d(\mathrm{d}x)$, and $\mathcal{B}_{\mathcal{X}}$ is the Borel σ-algebra of any space $\mathcal{X}$. $B_r(x)$, b_n and $\mathbf{S}^{d-1}$ will denote the closed ball with centre x and radius $r>0$, the volume of the unit ball in $\mathbb{R}^n$ and the unit sphere in $\mathbb{R}^d$, respectively. For any function f, $\mathrm{disc} f$ will denote the set of its discontinuity points.

In Appendix A.2 basics on point process theory are recalled; in particular we recall that every random closed set in $\mathbb{R}^d$ can be represented as a germ-grain model; therefore we shall consider here 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:

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

We remind that, whenever Φ is a marked Poisson point process, Θ is said to be a *Boolean model.* In this paper we shall denote by Θ_n a random closed set in $\mathbb{R}^d$ with integer dimension $0 \le n < d$, represented as in (1), where Φ has intensity measure $\Lambda(\mathrm{d}(x,s)) = \lambda(x,s)\mathrm{d}xQ(\mathrm{d}s)$ and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,s,y,t)\mathrm{d}x\mathrm{d}yQ_{[2]}(\mathrm{d}(s,t))$, while the grains Z_i are *countably $\mathcal{H}^n$-rectifiable.* (For a brief summary on basic notions of geometric measure theory, see Appendix A.3.)

Within the mathematical framework provided in [2] and in [52, Theorem 7], regularity assumptions on Θ_n have been given, ensuring a local approximation of its mean density $\lambda_{\Theta_n}(x)$. Since such assumptions are instrumental throughout this paper, we report here a key result proven in [52].

Theorem 1. *Let Θ_n be a random closed set in $\mathbb{R}^d$ with integer Hausdorff dimension $0 \le n < d$ as in* (1)*, where Φ has intensity measure $\Lambda(\mathrm{d}(x,s)) = \lambda(x,s)\mathrm{d}xQ(\mathrm{d}s)$ and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,s,y,t)\mathrm{d}x\mathrm{d}yQ_{[2]}(\mathrm{d}(s,t))$ such that the following assumptions are fulfilled:*

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

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

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

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

$$\sup_{x \in K_{\oplus \mathrm{diam}(Z(s))}} \lambda(x, s) \leq \widetilde{\xi}_K(s)$$

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

$$\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) \widetilde{\xi}_K(s) Q(\mathrm{d}s) < \infty$$

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

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

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

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

Then

$$\lambda_{\Theta_n}(x) = \lim_{r \downarrow 0} \frac{\mathbb{P}(x \in \Theta_{n_{\oplus r}})}{b_{d-n} r^{d-n}}, \qquad \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d. \tag{3}$$

For a discussion on the above assumptions we refer to [52, Sec. 3.1]; we recall just here that if Θ_n is a Boolean model, then (A3) is a consequence of (A1) and (A2). It is worth recalling also that by [52, Remark 4] and [52, Proposition 5], if Θ_n is a random closed set as above satisfying assumption (A1), then $\mathbb{E}[\mu_{\Theta_n}]$ is locally bounded and absolutely continuous with respect to $\mathcal{H}^d$, with density

$$\lambda_{\Theta_n}(x) = \int_{\mathbf{K}} \int_{x - Z(s)} \lambda(y, s) \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s), \quad \text{for } \mathcal{H}^d\text{-a.e } x \in \mathbb{R}^d. \tag{4}$$

We may notice that, if $n = 0$ and $\Theta_0 = X$ is an absolutely continuous random vector with pdf f_X, we have $\mathbb{E}[\mu_{\Theta_0}(A)] = \mathbb{E}[\mathcal{H}^0(X \cap A)] = \mathbb{P}(X \in A) = \int_A f_X(y) \mathrm{d}y$, for any Borel set $A \subset \mathbb{R}^d$, therefore $\lambda_{\Theta_0} \equiv f_X$, and, as a consequence, Eq. (3) reduces to

$$\lambda_{\Theta_0}(x) = \lim_{r \downarrow 0} \frac{\mathbb{P}(x \in X_{\oplus r})}{b_d r^d} = \lim_{r \downarrow 0} \frac{\mathbb{P}(X \in B_r(x))}{b_d r^d} = f_X(x), \qquad \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d, . \tag{5}$$

It is then well known that (5) leads to either the histogram, or to the kernel estimators for the pdf f_X (see Appendix A.1). By taking all the above into account, we will deal with two kinds of estimators for the mean density $\lambda_{\Theta_n}(x)$, for any $0 \leq n < d$; the first one as an extension of classical kernel density estimators, and the other one based on (3), as in [52] (see also Appendix A.3). Namely, given an i.i.d. random sample $\Theta_n^1, \ldots, \Theta_n^N$ of the random closed set Θ_n, we will analyze here the following estimators:

- **Kernel estimator** $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$
  as a natural extension of the kernel estimator for the probability density of a random vector

(see (A.4)), we define as *kernel estimator of the mean density* $\lambda_{\Theta_n}(x)$ *of* Θ_n*, at a point* $x \in \mathbb{R}^d$, the function

$$\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x) := \frac{1}{N}\sum_{i=1}^{N} k_{r_N} * \mathcal{H}^n_{|\Theta_n^i}(x) = \frac{1}{Nr_N^d}\sum_{i=1}^{N}\int_{\Theta_n^i} k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(dy), \tag{6}$$

where k is a multivariate kernel on $\mathbb{R}^d$;

- **"Minkowski content"-based estimator** $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$
  as a natural byproduct of (3), we define as *"Minkowski content"- based estimator of the mean density* $\lambda_{\Theta_n}(x)$ *of* Θ_n*, at a point* $x \in \mathbb{R}^d$, the function (firstly introduced in [51])

$$\widehat{\lambda}^{\mu,N}_{\Theta_n}(x) := \frac{\sum_{i=1}^{N} \mathbf{1}_{\Theta_n^i \cap B_{r_N}(x) \neq \emptyset}}{N b_{d-n} r_N^{d-n}}. \tag{7}$$

- **Natural estimator** $\widehat{\lambda}^{\nu,N}_{\Theta_n}(x)$.
  For the sake of completeness, we explicitly mention an additional estimator, directly deriving from the definition of density of a measure. Namely, we may notice that, if $\mathbb{E}[\mu_{\Theta_n}] \ll \mathcal{H}^d$, by the Besicovitch derivation theorem (e.g., see [4, Theorem 2.22]), we get

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

  therefore, given an i.i.d. sample $\Theta_n^1, ..., \Theta_n^N$ for Θ_n, the above approximation suggests to define as *natural estimator* $\widehat{\lambda}^{\nu,N}_{\Theta_n}(x)$ *of the mean density* $\lambda_{\Theta_n}(x)$ *of* Θ_n*, at a point* $x \in \mathbb{R}^d$, the function

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

Remark 2. *The well known kernel density estimators for random vectors, as defined in* (A.4)*, follows now as the particular case of* $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$ *for random closed sets of Hausdorff dimension* $n = 0$.

Furthermore, we may easily recognize that $\widehat{\lambda}^{\nu,N}_{\Theta_n}(x)$ *can be obtained as a particular case of the kernel estimator* $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$ *by choosing as kernel the function*

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

It is not difficult to realize that such an argument does not apply to the "Minkowski content"-based estimator defined in (7)*, which seems not reducible to a particular case of kernel estimators. Anyhow it is worth noticing that both* $\widehat{\lambda}^{\nu,N}_{\Theta_n}(x)$ *and* $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$ *reduce to the usual histogram density estimator if* $n = 0$*, and* $\Theta_0 = X$ *is a real random variable*

$$\widehat{\lambda}^{\nu,N}_{X}(x) = \widehat{\lambda}^{\mu,N}_{X}(x) = \frac{1}{N 2 r_N}\sum_{i=1}^{N} \mathbf{1}_{[x-r_N, x+r_N]}(X_i),$$

where $X_1, \ldots, X_N$ *is an i.i.d. random sample for* X*. In the* d*-dimensional case, both kinds of estimators coincide with the so-called* naive *kernel estimator (the one given by* $k(z) = \frac{1}{b_d}\mathbf{1}_{B_1(0)}(z)$*), in the case* $n = 0$*, when* Θ_0 *is a random vector.*

In [52, Corollary 13], it has already been proven that, under the assumptions (A1), (A2) and (A3), $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$ is asymptotically unbiased and weakly consistent, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$, if r_N is such that

$$\lim_{N\to\infty} r_N = 0 \quad \text{and} \qquad \lim_{N\to\infty} N r_N^{d-n} = \infty. \tag{8}$$

In order to prove relevant statistical properties of the estimators of the mean densities of random closed sets introduced above, and to face the problem of finding the corresponding optimal bandwidths, we shall require that the involved sets satisfy analogous regularity conditions.

3. Statistical properties and optimal bandwidths of $\widehat{\boldsymbol{\lambda}}^{\boldsymbol{\kappa},\boldsymbol{N}}_{\boldsymbol{\Theta_n}}(\boldsymbol{x})$

As anticipated in the introduction, the proofs of the main results are left to Section 7, so to let the reader follow the "leit motiv" of our treatment.

It is worth noticing that according to [43, page 589] "It is well known that the value of the bandwidth is of critical importance, while the shape of the kernel function has little practical impact". This may justify our own choice of kernels of easier analytical treatment, while concentrating on the analysis of optimal bandwidths.

3.1. Bias and variance

Theorem 3 (Asymptotic unbiasedness). *Let Θ_n satisfy assumptions (A1) and (A2), and $\{\Theta^i_n\}_{i\in\mathbb{N}}$ be a sequence of random closed sets i.i.d. as Θ_n. Let k be a kernel with compact support (defined as in Definition A.2); then the kernel density estimator $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$ of $\lambda_{\Theta_n}(x)$ defined by (6) is asymptotically unbiased, i.e.*

$$\lim_{N\to\infty} \mathbb{E}[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)] = \lambda_{\Theta_n}(x), \qquad \textit{for } \mathcal{H}^d\textit{-a.e. } x \in \mathbb{R}^d,$$

if $\lim_{N\to\infty} r_N = 0$.

Proof. See Section 7. □

In order to prove also the weak consistency of $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$, we require the following additional regularity assumptions on Θ_n, closely related to the assumptions (A1) and (A3),

($\overline{A1}$) 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 $\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s))Q(\mathrm{d}s) < \infty$ and

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

for some $\gamma, \widetilde{\gamma} > 0$ independent of y and s;

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

$$\sup_{y\in \overline{C}_{\oplus \operatorname{diam} Z(t)}} \ \sup_{x\in C_{\oplus \operatorname{diam} Z(s)}} g(x,s,y,t) \le \xi_{C,\overline{C}}(s,t)$$

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

$$\int_{\mathbf{K}^2} \mathcal{H}^n(\Xi(s))\mathcal{H}^n(\Xi(t))\xi_{C,\overline{C}}(s,t)Q_{[2]}(\mathrm{d}s,\mathrm{d}t) < \infty. \tag{9}$$

Note that ($\overline{A1}$) is just the assumption (A1), together with the condition that the grains of the germ-grain model Θ_n are n-Alfhors regular; moreover, the assumptions (A1) and ($\overline{A1}$) might be regarded as the stochastic version of (A.7) and (A.8), respectively, well known in geometric measure theory (see Appendix A.3).

The following proposition tells us the relationship between the assumptions ($\overline{A3}$) and (A3).

Proposition 4. *If (A1) is satisfied, then* (9) $\Longrightarrow$ (2).

Proof. See Section 7. □

Remark 5 (Particular case: Boolean models). *If Θ_n is a Boolean model with intensity measure $\Lambda(\mathrm{d}(x,s)) = \lambda(x,s)\mathrm{d}xQ(\mathrm{d}s)$, then $g(x,s,y,t) = \lambda(x,s)\lambda(y,t)$ and $Q_{[2]}(\mathrm{d}(s,t)) = Q(\mathrm{d}s)Q(\mathrm{d}t)$; thus it is easy to check that* (9) *holds with $\xi_{C,\overline{C}}(s,t) := \widetilde{\xi}_C(s)\widetilde{\xi}_{\overline{C}}(t)$, and so ($\overline{A3}$) is implied by (A2).*

We are now ready to provide an upper bound for the variance of the kernel estimator $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$.

Theorem 6 (Upper bound for the variance). *Let Θ_n satisfy assumptions ($\overline{A1}$), (A2) and ($\overline{A3}$), and $\{\Theta_n^i\}_{i\in\mathbb{N}}$ be a sequence of random closed sets, i.i.d. as Θ_n. Let k be a kernel with compact support,* $\mathrm{supp}(k) \subseteq B_R(0)$, *defined as in Definition A.2; then, for N sufficiently large so that $r_N \leq \min\{1, 1/2R\}$,*

$$Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) \leq \frac{M\widetilde{\gamma}2^nR^n}{Nr_N^{d-n}}\int_{\boldsymbol{K}}\mathcal{H}^n(\Xi(s))\widetilde{\xi}_{B_R(x)}(s)Q(\mathrm{d}s) + \frac{1}{N}\int_{\boldsymbol{K}^2}\mathcal{H}^n(\Xi(s))\mathcal{H}^n(\Xi(\widetilde{s}))\xi_{B_R(x),B_R(x)}(s,\widetilde{s})Q_{[2]}(\mathrm{d}s,\mathrm{d}\widetilde{s}).$$

Proof. See Section 7. □

By Theorem 3 and Theorem 6, we may directly state the following

Corollary 7. *Let Θ_n satisfy assumptions ($\overline{A1}$), (A2) and ($\overline{A3}$), and $\{\Theta_n^i\}_{i\in\mathbb{N}}$ be a sequence of random closed sets, i.i.d. as Θ_n. Let k be a kernel with compact support,* $\mathrm{supp}(k) \subseteq B_R(0)$, *defined as in Definition A.2; if r_N is such that*

$$\lim_{N\to\infty} r_N = 0 \quad \text{and} \quad \lim_{N\to\infty} Nr_N^{d-n} = \infty, \tag{10}$$

then the kernel density estimator $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$ of $\lambda_{\Theta_n}(x)$, defined by (6), *is weakly consistent for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$.*

Note that the above conditions on r_N are the same required for the weak consistency of the "Minkowski content"-based estimator $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$ (see (8)).

3.2. Optimal bandwidths

As mentioned above, a crucial problem of statistical interest is the choice of an optimal bandwidth r_N. As usual, we will look for an r_N which minimizes the *asymptotic mean square error*

(AMSE); the optimal bandwidth known in literature for the kernel density estimation of a random variable will follow here as a particular case (see Corollary 12).

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

$$\begin{aligned} |\alpha| &:= \alpha_1 + \cdots + \alpha_d \\ \alpha! &:= \alpha_1! \cdots \alpha_d! \\ y^\alpha &:= y_1^{\alpha_1} \cdots y_d^{\alpha_d} \\ D_y^\alpha \lambda(y,s) &:= \frac{\partial^{|\alpha|}\lambda(y,s)}{\partial y_1^{\alpha_1} \cdots \partial y_d^{\alpha_d}}; \end{aligned}$$

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

$$\mathcal{D}^{(\alpha)}(s) := \mathrm{disc}(D_y^\alpha \lambda(y,s)), \qquad \mathcal{D}(s) := \mathrm{disc}(\lambda(\cdot, s)).$$

The (well known in geometric measure theory) notion of *approximate tangent space* to any $\mathcal{H}^n$-rectifiable compact set A of $\mathbb{R}^d$ at a point $x \in A$ (see Appendix A.3), will arise in the approximation of the variance of $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$. In the following theorem we assume that $\lambda(\cdot, s)$ is twice differentiable at least; for sake of simplicity, we also assume that k is continuous (the non-continuous case is discussed in Remark 9).
Note that the assumption ($A2$bis) below will play the same role of the assumption ($A2$), and it is trivially satisfied if $D_y^\alpha \lambda(y,s)$ are bounded.

Theorem 8 (Main theorem). *In addition to the hypotheses of Theorem 6, we assume that the kernel k is continuous, and that the following assumption is fulfilled, for $|\alpha| = 2$,*

($A2$bis) *for any $s \in \boldsymbol{K}$, $\mathcal{H}^n(\mathcal{D}^{(\alpha)}(s)) = 0$ and $D_y^\alpha \lambda(y,s)$ is locally bounded such that for any compact $C \subset \mathbb{R}^d$*

$$\sup_{y \in C_{\oplus \mathrm{diam} Z(s)}} |D_y^\alpha \lambda(y,s)| \le \widetilde{\xi}_C^{(\alpha)}(s)$$

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

$$\int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) \widetilde{\xi}_C^{(\alpha)}(s) Q(\mathrm{d}s) < \infty.$$

Then, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$,

$$Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) = C_{Bias}(x) r_N^2 + o(r_N^2) \tag{11}$$

$$Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) = \frac{C_{Var}(x)}{N r_N^{d-n}} + o(\frac{1}{N r_N^{d-n}}), \tag{12}$$

with

$$C_{Bias}(x) := \sum_{|\alpha|=2} \frac{1}{\alpha!} \int_{\mathbb{R}^d} k(z) z^\alpha \mathrm{d}z \int_{\boldsymbol{K}} \int_{x-Z(s)} D_y^\alpha \lambda(y,s) \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s), \tag{13}$$

$$C_{Var}(x) := \int_{\mathbf{K}} \int_{\mathbb{R}^d} \int_{x-Z(s)} \int_{\pi_y^{x,s}} k(z) k(z+w) \lambda(y,s) \mathcal{H}^n(\mathrm{d}w) \mathcal{H}^n(\mathrm{d}y) \mathrm{d}z Q(\mathrm{d}s), \tag{14}$$

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

Proof. See Section 7. $\square$

Remark 9. *By Theorem A.5 in Appendix A.3, it follows that the limit in* (33) *in the proof of the above theorem (and so its assertion well) holds also in the case the kernel k is not necessarily continuous, provided that*

$$\mathcal{H}^n_{|\pi_y^{x,s}}(\mathrm{disc}(k(z+\cdot)) = 0, \tag{15}$$

for any $s \in \mathbf{K}$, $z \in \mathrm{supp}(k)$, and $\mathcal{H}^n$-a.e. $y \in x - Z(s)$. Such a condition is trivially fulfilled in several cases of interest in applications; for instance, if $Z(s)$ is a sufficiently regular curve in $\mathbb{R}^d$, it follows that $\pi_y^{x,s}$ is a line in $\mathbb{R}^d$ for any $s \in \mathbf{K}$, $x \in \mathbb{R}^d$, and $\mathcal{H}^1$-a.e. $y \in x - Z(s)$, and so (15) *is satisfied by the kernel $k(z) = \frac{1}{b_d}\mathbf{1}_{B_1(0)}(z)$, which is of particular interest among all kernels with compact support (see Remark 2).*

It may be proved that the the above results hold for kernels k with non-compact support too, provided that the assumptions $(A2)$ and $(\overline{A3})$ are replaced by suitable integrability conditions on k, λ and g which allow to apply again the Dominated Convergence Theorem.
Actually, in practical applications it is commonly assumed that λ, its partial derivatives D_x^α, and g are bounded; under such assumptions, if k has compact support, $(A2)$ and $(\overline{A3})$ simplify. More precisely, the integrability conditions expressed in (A2) and (A2bis) are trivially satisfied by $(\overline{A1})$; moreover, $(\overline{A3})$ is a consequence of $(\overline{A1})$ whenever $Q_{[2]}(\mathrm{d}s, \mathrm{d}t) = Q(\mathrm{d}s)Q(\mathrm{d}t)$. Examples of point processes having bounded both intensity and second moment density are provided in [52, Example 2]. A relevant particular case of bounded intensity λ is discussed in the Section 3.3.3.

3.2.1. Pointwise optimal bandwidth

The *mean square error* $MSE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))$ of $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$, defined as usual by

$$MSE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) := \mathbb{E}[(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x) - \lambda_{\Theta_n}(x))^2],$$

is given by

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

so that, from (11) and (12), the following asymptotic approximation of the mean square error follows

$$MSE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) = C^2_{Bias}(x) r_N^4 + \frac{1}{N r_N^{d-n}} C_{Var}(x) + o(r_N^4) + o\Big(\frac{1}{N r_N^{d-n}}\Big), \quad \text{as } N \to +\infty. \tag{16}$$

Hence

$$AMSE(r_N) = C^2_{Bias}(x) r_N^4 + \frac{1}{N r_N^{d-n}} C_{Var}(x).$$

Thus, by defining the optimal bandwidth $r_N^{\mathrm{o,AMSE}}(x)$ associated to a point $x \in \mathbb{R}^d$ as

$$r_N^{\mathrm{o,AMSE}}(x) := \arg\min_{r_N} AMSE(r_N),$$

it is easy to obtain

$$r_N^{\mathrm{o,AMSE}}(x) = \sqrt[4+d-n]{\frac{(d-n)C_{Var}(x)}{4NC^2_{Bias}(x)}}, \quad \mathcal{H}^d\text{-a.e.}\, x \in \mathbb{R}^d, \tag{17}$$

for any fixed sample size N (sufficiently large so to guarantee the asymptotic properties of $\widehat{\lambda}^{\kappa,N}_{\Theta_n}$), provided that $C_{Bias}(x) \neq 0$.

We may observe that, if $C_{Bias}(x) = 0$, the above equation does not apply; in such a case one should use additional terms in the bias expansion. Of course this might complicate the solution for the optimal bandwidth; we refer to [44] for a more detailed discussion of such a problem in the particular case of random variables. We point out that a case of particular interest which cannot be solved by adding additional terms in the bias expansion, is the one in which the germ grain process Θ_n is stationary, that is when λ is constant (see Section 3.3.3).

In accordance with Theorem 6, by (10) and (16) we have that MSE goes to 0 as $N \to \infty$; hence, by (17), the optimal bandwidth goes to 0 as $N^{-1/(4+d-n)}$.
We also point out that $r_N^{\mathrm{o,MSE}}(x)$ depends on the unknown intensity λ and the probability measure Q. As for the case $n = 0$, which has been recalled in Appendix A.1, methods for estimating the intensity λ and the probability measure Q are required. This problem is a subject for further investigation. The reader may refer to [20, p. 770 and followings], [39, p. 77 and followings] and [47] for related inference problems regarding Boolean models.

3.2.2. Uniform optimal bandwidth

It is of interest to provide a *uniform* (or *global*) optimal bandwidth for the estimate of the mean density λ_{Θ_n}, in any given window W, a compact subset of $\mathbb{R}^d$. As for the random variables, we introduce the integrated mean square error of $\widehat{\lambda}^{\kappa,N}_{\Theta_n}$ in W, so defined

$$\begin{aligned} MISE[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(W)] &:= \int_W MSE[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)]\mathrm{d}x \\ &= \int_W [Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))]^2\mathrm{d}x + \int_W Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))\mathrm{d}x. \end{aligned} \tag{18}$$

By proceeding along the same lines of Theorem 8, we can prove the following.

Proposition 10. *Under the hypotheses of Theorem 8, for any compact set W in $\mathbb{R}^d$,*

$$MISE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(W)) = c^2_{Bias}(x)r_N^4 + \frac{1}{Nr_N^{d-n}}c_{Var}(x) + o(r_N^4) + o\Big(\frac{1}{Nr_N^{d-n}}\Big), \quad \text{as } N \to +\infty,$$

where

$$c^2_{Bias} := \int_W C^2_{Bias}(x)\mathrm{d}x, \qquad \text{and} \qquad c_{Var} := \int_W C_{Var}(x)\mathrm{d}x.$$

Proof. See Section 7. □

By the proposition above, we get the following asymptotic approximation of $MISE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(W))$

$$AMISE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(W)) := c^2_{Bias}(x)r_N^4 + \frac{1}{Nr_N^{d-n}}c_{Var}(x), \quad \text{as } N \to +\infty.$$

Thus, by defining the uniform optimal bandwidth $r_{N,W}^{\mathrm{o,AMISE}}$ in W as

$$r_{N,W}^{\mathrm{o,AMISE}} := \arg\min_{r_N} AMISE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(W)),$$

we may easily derive

$$r_{N,W}^{\mathrm{o,AMISE}} = \sqrt[4+d-n]{\frac{(d-n)c_{Var}}{4Nc_{Bias}^2}}, \tag{19}$$

for any given sample size N, sufficiently large.

Remark 11. *We point out that we need to introduce a compact window W in the definition of the MISE of the kernel estimator $\widehat{\lambda}_{\Theta_n}^{\kappa,N}$ because, in general, $\int_A MSE[\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)]\mathrm{d}x = \infty$, for unbounded subsets A of $\mathbb{R}^d$. As a matter of fact, in the particular case of a point process $\Theta_0 = \Psi$ in $\mathbb{R}^d$ (see Section 3.3.2 below), we may notice that, for instance, in order to evaluate c_{Var} it emerges the evaluation of $\mathbb{E}[\Psi(W)]$, i.e. the mean number of points in W, which may diverge for non compact windows W. Clearly then, in the particular case of random variables (discussed in Section 3.3.1) the above results hold also for non-necessarily compact subsets of $\mathbb{R}^d$; indeed a random vector X in $\mathbb{R}^d$ may be seen as a point process Ψ with only one point, and so $\mathbb{E}[\Psi(A)] = \mathbb{P}(X \in A) \leq 1$ for any $A \subseteq \mathbb{R}^d$.*

3.3. Particular cases

3.3.1. Random variables

The next corollary shows how already known results for kernel density estimates of the pdf of absolutely continuous random variables (see Appendix A.1), follow as particular cases.

Corollary 12. *Let X be a random variable with pdf $f_X \in C^2$; then the well-known pointwise optimal bandwidth given in* (A.2)*, and the well-known global optimal bandwidth given in* (A.3)*, follow now by* (17) *and by* (19)*, respectively, as a particular case.*

Proof. We observed that if $n = 0$ and $\Theta_0 = X$ is an absolutely continuous random vector with pdf f_X, then $\lambda_{\Theta_0} \equiv f_X$. In order to apply the above results, let us consider X as the trivial germ-grain process driven by the marked point process $\Phi = \{(X, s)\}$ in $\mathbb{R}$ with mark space $\mathbf{K} = \mathbb{R}$, consisting of one point (X) only, with grain $Z(s) := s$, and intensity measure $\Lambda(\mathrm{d}(y, s)) = f(y)\mathrm{d}y\delta_0(s)\mathrm{d}s$ (hence $\lambda_X(x) = f(x)$, in accordance with (4), as expected). It is clear that the hypotheses of Theorem 8 are fulfilled, being $(\overline{A1})$, $(A2)$, $(\overline{A3})$ and $(A2\text{bis})$ trivially satisfied by choosing $\Xi(s) = Z(s)$, and $\gamma = \widetilde{\gamma} = 1$, and observing that f, f' and f'' are continuous by hypothesis, and $g \equiv 0$. Thus,

$$C_{Bias}(x) \stackrel{(13)}{=} \frac{1}{2}\int_{\mathbb{R}} k(z)z^2\mathrm{d}z \int_{\mathbb{R}}\int_{x-s} f''(y)\mathcal{H}^0(\mathrm{d}y)\delta_0(s)\mathrm{d}s = \frac{1}{2}f''(x)\int_{\mathbb{R}} k(z)z^2\mathrm{d}z,$$

$$C_{Var}(x) \stackrel{(14)}{=} \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{x-s}\int_0 k(z)k(z+w)f(y)\mathcal{H}^0(\mathrm{d}w)\mathcal{H}^0(\mathrm{d}y)\mathrm{d}z\delta_0(s)\mathrm{d}s = f(x)\int_{\mathbb{R}} k^2(z)\mathrm{d}z,$$

so that, by (17) with $d = 1$, $n = 0$, we reobtain

$$r_N^{\mathrm{o,AMSE}}(x) = \sqrt[5]{\frac{C_{Var}(x)}{4NC_{Bias}^2(x)}} = (\mathrm{A.2}).$$

With regard to the global optimal bandwidth, it is sufficient to observe that, in this case,

$$4Nc_{Bias}^2 = 4N\int_{\mathbb{R}}\Big(\frac{1}{2}f''(x)\int_{\mathbb{R}}k(z)z^2\mathrm{d}z\Big)^2\mathrm{d}x = N(c_2\|f''\|)^2,$$

$$c_{Var} = \int_{\mathbb{R}}f(x)\int_{\mathbb{R}}k^2(z)\mathrm{d}z\,\mathrm{d}x = c_1.$$

□

3.3.2. Point processes

Since a point process Ψ in $\mathbb{R}^d$ with intensity f may be regarded as a particular random closed set of dimension $n=0$ with mean density $\lambda_\Psi = f$, we may apply the above results to provide kernel density estimators, as well as the optimal bandwidth, of the intensity of point processes too.

Corollary 13. *Let $\{\Psi^i\}_{i\in\mathbb{N}}$ be a sequence of point processes in $\mathbb{R}^d$, i.i.d. as Ψ, with intensity $\lambda_\Psi\in C^2$, and locally bounded second moment density g, and let k be a kernel with compact support, continuous in 0. Then the kernel density estimator $\widehat{\lambda}_\Psi^{\kappa,N}(x)$ of $\lambda_\Psi(x)$, so defined*

$$\widehat{\lambda}_\Psi^{\kappa,N}(x) = \frac{1}{Nr_N^d}\sum_{i=1}^{N}\sum_{x_j\in\Psi^i}k\Big(\frac{x-x_j}{r_N}\Big), \tag{20}$$

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

$$\lim_{N\to\infty} r_N = 0 \quad \textit{and} \qquad \lim_{N\to\infty} Nr_N^d = \infty.$$

Moreover, the pointwise optimal bandwidth $r_N^{\mathrm{o,AMSE}}(x)$ minimizing the AMSE is given by

$$r_N^{\mathrm{o,AMSE}}(x) = \sqrt[4+d]{\frac{d\lambda_\Psi(x)\int_{\mathbb{R}^d}k^2(z)\mathrm{d}z}{4N\Big(\sum_{|\alpha|=2}\frac{1}{\alpha!}D_x^\alpha\lambda_\Psi(x)\int_{\mathbb{R}^d}k(z)z^\alpha\mathrm{d}z\Big)^2}}, \quad \textit{for } \mathcal{H}^d\textit{-a.e. } x\in\mathbb{R}^d, \tag{21}$$

whereas, for any compact window $W\subset\mathbb{R}^d$, a uniform optimal bandwidth $r_{N,W}^{\mathrm{o,AMISE}}$ is given by

$$r_{N,W}^{\mathrm{o,AMISE}} = \sqrt[4+d]{\frac{d\mathbb{E}[\Psi(W)]\int_{\mathbb{R}^d}k^2(z)\mathrm{d}z}{4N\int_W\Big(\sum_{|\alpha|=2}\frac{1}{\alpha!}D_y^\alpha\lambda_\Psi(x)\int_{\mathbb{R}^d}k(z)z^\alpha\mathrm{d}z\Big)^2\mathrm{d}x}}. \tag{22}$$

Proof. By proceeding along the same lines of the proof of Corollary 12, Ψ might be seen as a trivial marked point process with mark space $\mathbf{K}=\mathbb{R}$ having intensity measure $\Lambda(\mathrm{d}(y,s)) = \lambda_\Psi(y)\mathrm{d}y\delta_0(s)\mathrm{d}s$, and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,y)\mathrm{d}x\mathrm{d}y\delta_0(s)\delta_0(t)\mathrm{d}s\mathrm{d}t$. Then, the asymptotic properties of the kernel estimator (20) directly follow by Theorem 6.

By Remark 9, observing that $\pi_y^{x,s} = \{0\}$, and that by hypothesis the kernel k is continuous in 0, we may claim that the assertion of Theorem 8 holds with

$$\begin{aligned} C_{Bias}(x) &= \sum_{|\alpha|=2} \frac{1}{\alpha!} D_y^{\alpha} \lambda_{\Psi}(y) \int_{\mathbb{R}^d} k(z) z^{\alpha} \mathrm{d}z, \\ C_{Var}(x) &= \lambda_{\Psi}(x) \int_{\mathbb{R}^d} k^2(z) \mathrm{d}z, \end{aligned}$$

so that, by (17), we get the optimal bandwidth defined in (21).
The equality in (22) directly follows by (19) and the proof of Corollary 13, having observed that

$$\int_W \lambda_{\Psi}(x) \mathrm{d}x = \mathbb{E}[\Psi(W)].$$

$\square$

Remark 14. *Let us notice that by choosing* $k(z) := \frac{1}{b_d} \mathbf{1}_{B_1(0)}(z)$ *in* (20)*, with* $N = 1$*, we reobtain the well-known classic and widely used Berman-Diggle estimator [26, 8, 50]*

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

3.3.3. Homogeneous case: unbiased estimators

Let us assume that $\Lambda(\mathrm{d}(x,s)) = c\mathrm{d}xQ(\mathrm{d}s)$; i.e. $\lambda(x,s) \equiv c$ for any $(x,s) \in \mathbb{R}^d \times \mathbf{K}$; then $\Phi = \{(x_i, s_i)\}_{i \in \mathbb{N}}$ is an independent marking of the marginal process $\{x_i\}_{i \in \mathbb{N}}$, which is stationary, and so Θ_n is a stationary random closed set as well. Notice that the very particular case of a real random variable $\Theta_0 = X$ (see also Corollary 12) does not make sense, since λ cannot be a nontrivial constant on the whole real line, whereas the case $\lambda =$ const on a compact set corresponds to the case of a random variable uniformly distributed on that set. This might be a reason why such a case has not been taken into account in the usual kernel density estimation theory; on the other hand the homogeneous case is of particular interest in random sets theory (even for a point process Θ_0, for which we still have $n = 0$).
Under the assumption that Θ_n is stationary, a first important result is that the kernel density estimator $\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x)$ is now unbiased for any bandwidth r, and independent of x; namely Theorem 3 simplifies as follows.

Proposition 15. *Let* Θ_n *be a random closed set in* $\mathbb{R}^d$ *with integer dimension* $0 \leq n < d$*, represented as in* (1)*, where* Φ *has intensity measure* $\Lambda(\mathrm{d}(x,s)) = c\mathrm{d}xQ(\mathrm{d}s)$*, such that Assumption (A1) is fulfilled, and let* $\{\Theta_n^i\}_{i \in \mathbb{N}}$ *be a sequence of random closed sets, i.i.d. as* Θ_n*. Then the kernel density estimator* $\widehat{\lambda}_{\Theta_n}^{\kappa,N}$ *of* λ_{Θ_n}*, defined by*

$$\widehat{\lambda}_{\Theta_n}^{\kappa,N}(x) := \frac{1}{N} \sum_{i=1}^{N} k_r * \mathcal{H}^n_{|_{\Theta_n^i}}(0) = \frac{1}{N r^d} \sum_{i=1}^{N} \int_{\Theta_n^i} k\Big(\frac{-y}{r}\Big) \mathcal{H}^n(dy), \qquad \text{for } \mathcal{H}^d\text{-a.e. } x \in \mathbb{R}^d \tag{23}$$

is unbiased for any bandwidth $r > 0$*, and any sample size* N*.*

Proof. First of all let us notice that Assumption (A2) is trivially satisfied, and so, by the proof of Theorem 3, we have that $\lambda_{\Theta_n}(x) = c\mathbb{E}[\mathcal{H}^n(Z)] =: \lambda_{\Theta_n} \in \mathbb{R}_+$, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$. Then, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$, we may take $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x) \equiv \widehat{\lambda}^{\kappa,N}_{\Theta_n}(0)$, as in (6), with r independent of N (and so defined as in (23)), and observe that

$$\mathbb{E}[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)] \overset{(28)}{=} c \int_{\mathbf{K}} \mathcal{H}^n(Z(s))Q(\mathrm{d}s) = \lambda_{\Theta_n}.$$

□

As a byproduct of the above proposition, we have that Corollary 7 simplifies now as follows.

Corollary 16. *Let Θ_n satisfy the hypotheses of Proposition 15, with (A1) replaced by $(\overline{A1})$, and Assumption $(\overline{A3})$. Let $\{\Theta^i_n\}_{i\in\mathbb{N}}$ be a sequence of random closed sets, i.i.d. as Θ_n, and let k be a kernel with compact support. Then the kernel density estimator $\widehat{\lambda}^{\kappa,N}_{\Theta_n}$ of λ_{Θ_n} defined by (23) is strongly consistent for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$, as $N \to \infty$.*

Proof. By defining $Y_i := k_r * \mathcal{H}^n_{|\Theta^i_n}(0)$, $i = 1, 2, \dots$, we have that $Y_1, Y_2, \dots$ are i.i.d. as $Y = k_r * \mathcal{H}^n_{|\Theta_n}(0)$; we have $\mathbb{E}[Y] = \lambda_{\Theta_n}$, and, by Theorem 6, $Var(Y) < \infty$. Then the SLLN implies that $\widehat{\lambda}^{\kappa,N}_{\Theta_n} \to \lambda_{\Theta_n}$, a.s. as $N \to \infty$. □

It is clear that in this case $Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}) = 0$ for any fixed sample size N, and for any bandwidth r; equivalently,

$$MSE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}) = Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}) = \frac{\mathbb{E}\big[(k_r * \mathcal{H}^n_{|\Theta_n}(0))^2\big] - (\lambda_{\Theta_n})^2}{N},$$

therefore the optimal bandwidth which minimizes the MSE has to minimize the variance: for any fixed sample size N, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$,

$$r^{\mathrm{o,MSE}}(x) := \arg\min_r MSE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}) = \arg\min_r Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}) = \arg\min_r \mathbb{E}\big[(k_r * \mathcal{H}^n_{|\Theta_n}(0))^2\big].$$

By the proof of Theorem 6 we know that

$$\mathbb{E}\big[(k_r * \mathcal{H}^n_{|\Theta_n}(0))^2\big] = I_1(r) + I_2(r),$$

with

$$I_1(r) \overset{(31)}{:=} \frac{c}{r^d} \int_{\mathbf{K}} \int_{Z(s)} \int_{Z(s)} \int_{\mathbb{R}^d} k(z)k\Big(z + \frac{y - \widetilde{y}}{r}\Big)\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)\mathcal{H}^n(\mathrm{d}\widetilde{y})Q(\mathrm{d}s)$$

$$I_2(r) := \int_{\mathbf{K}^2} \int_{Z(s)} \int_{Z(\widetilde{s})} \int_{\mathbb{R}^{2d}} k(z)k(\widetilde{z})g(-y - rz, s, -\widetilde{y} - r\widetilde{z}, \widetilde{s})\mathrm{d}\widetilde{z}\mathrm{d}z\mathcal{H}^n(\mathrm{d}\widetilde{y})\mathcal{H}^n(\mathrm{d}y)Q_{[2]}(\mathrm{d}s, \mathrm{d}\widetilde{s}).$$

Then one has to minimize the function $I(r) := I_1(r) + I_2(r)$.
We discuss here two important cases of particular interest in applications, homogeneous Boolean models, and a - non Boolean - stationary germ-grain model with a cluster point process as germ process.

Particular case: homogenous Boolean models

Let Θ_n be a homogenous Boolean model with intensity measure $\Lambda(\mathrm{d}(x,s)) = c\mathrm{d}xQ(\mathrm{d}s)$ and typical grain Z, such that $\mathbb{E}[(\mathcal{H}^n(Z))^2] < \infty$, and let k be continuous in the interior of its support. Then, it is easy to check (see also Remark 5) that

$$I_2(r) = (c\mathbb{E}[\mathcal{H}^n(Z)])^2 < \infty,$$

and so $r^{\mathrm{o,MSE}} = \arg\min_r I_1(r)$.

By observing now that $I_1(r) > 0$ is continuous for any $r > 0$ (by using for instance the Dominated Convergence Theorem),

$$\lim_{r\to\infty} I_1(r) \leq \lim_{r\to\infty} \frac{Mc}{r^d} \int_{\mathbf{K}} \mathcal{H}^n(Z(s))^2 Q(\mathrm{d}s) = 0,$$

and, by the proof of Theorem 8,

$$\begin{aligned}\lim_{r\downarrow 0} I_1(r) &= \lim_{r\downarrow 0} \frac{1}{r^{d-n}} r^{d-n} I_1(r)\\ &= c\int_{\mathbf{K}}\int_{\mathbb{R}^d}\int_{-Z(s)}\int_{\pi_y^{0,s}} k(z)k(z+w)\mathcal{H}^n(\mathrm{d}w)\mathcal{H}^n(\mathrm{d}y)\mathrm{d}zQ(\mathrm{d}s)\lim_{r\downarrow 0}\frac{1}{r^{d-n}} = +\infty,\end{aligned}$$

we conclude that, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$, $r^{\mathrm{o,MSE}} = +\infty$.

Remark 17. *As a particular case, if Ψ is a homogenous Poisson point process in $\mathbb{R}^d$ with intensity $\lambda_\Psi > 0$, the optimal bandwidth $r^{\mathrm{o,MSE}}$ of the kernel intensity estimator $\widehat{\lambda}_\Psi^{\kappa,N}(x) = \widehat{\lambda}_\Psi^{\kappa,N}(0)$ for any $x \in \mathbb{R}^d$, defined by (20), is $r^{\mathrm{o,MSE}} = +\infty$, in accordance with both intuition and known results in literature (e.g., see [47, p. 46], [25, p. 34], [39, p. 8]); in particular, if W is the* observation window *of any realization of the process (and so $N = 1$), and $|W|$ its volume, we reobtain (by choosing $k(z) := \frac{1}{b_d}\mathbf{1}_{B_1(0)}(z)$) that the best unbiased estimator of the intensity λ_Ψ of Ψ is given by*

$$\widehat{\lambda}_\Psi = \frac{\Psi(W)}{|W|}, \qquad \textit{with } |W| \to \infty.$$

Note that, being Ψ stationary, $\widehat{\lambda}_\Psi$ defined above is equivalent to the following estimator introduced in [25, p. 34]

$$\widehat{\lambda}_\Psi^D := \frac{\sum_{i=1}^N \Psi(W_i)}{N|W|}, \qquad \textit{with } N \to \infty,$$

where $W_1, \ldots, W_N$, are disjoint compact subsets of $\mathbb{R}^d$, each with volume $|W|$; let us also observe that

$$\widehat{\lambda}_\Psi^D = \widehat{\lambda}_\Psi^N := \frac{\sum_{i=1}^N \Psi^i(W)}{N|W|}, \qquad \textit{with } N \to \infty,$$

where W is a given observation window, and $\Psi^1, \ldots, \Psi^N$ is an i.i.d. sample of Ψ, and so we conclude that $\widehat{\lambda}_\Psi$, $\widehat{\lambda}_\Psi^D$ and $\widehat{\lambda}_\Psi^N$ provide equivalent ways to estimate λ_Ψ.

An example of stationary non Boolean germ-grain model
Let Θ_n be a random closed set in $\mathbb{R}^2$ with integer dimension $0 \leq n < 2$, represented as in (1), such that Φ is an independent marking of the germ point process $\widetilde{\Psi} = \{x_i\}$, which is assumed to be a *Matèrn cluster process* in $\mathbb{R}^2$ (e.g., see [5]), whose parent process is a homogeneous Poisson process with intensity α, and each cluster consists of $\mathfrak{M} \sim$ Poisson(m) points independent and uniformly distributed in the ball $B_R(x)$, where x is the centre of a cluster. It follows that Θ_n is stationary with intensity measure $\Lambda(\mathrm{d}(x,s)) = m\alpha \mathrm{d}x Q(\mathrm{d}s)$, and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = \Big(\alpha^2 m^2 + \alpha m^2 \dfrac{\mathcal{H}^2(B_R(x)\cap B_R(y))}{\pi R^4}\Big)\mathrm{d}x\mathrm{d}yQ(\mathrm{d}s)Q(\mathrm{d}t)$.
We assume that the typical grain Z satisfies the regularity assumption ($\overline{A1}$). Note that $g(x,s,y,t) = \alpha^2m^2 + \alpha m^2 \frac{\mathcal{H}^2(B_R(x)\cap B_R(y))}{\pi R^4} \leq \alpha^2 m^2 + \alpha m^2/(\pi R^2)$, and so Assumption ($\overline{A3}$) is fulfilled; thus Corollary 16 applies.
Similarly to the homogeneous Boolean model case, if k is continuous in the interior of its support, $k \leq M$, and $\mathbb{E}[(\mathcal{H}^n(Z))^2] < \infty$, it is easy to check that $I(r)$ is continuous and $\lim_{r\downarrow 0} I(r) = +\infty$.
Let us notice that

$$
\begin{aligned}
I_2(r) \quad &= \quad \alpha^2m^2(\mathbb{E}[\mathcal{H}^n(Z)])^2 + \frac{\alpha m^2}{\pi R^4}\int_{\mathbf{K}^2}\int_{Z(s)}\int_{Z(\widetilde{s})}\int_{\mathbb{R}^{2d}} k(z)k(\widetilde{z}) \\
&\qquad \mathcal{H}^2(B_R(-y-rz)\cap B_R(-\widetilde{y}-r\widetilde{z}))\mathrm{d}\widetilde{z}\mathrm{d}z\mathcal{H}^n(\mathrm{d}\widetilde{y})\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)Q(\mathrm{d}\widetilde{s}) \\
&= \quad \alpha^2m^2(\mathbb{E}[\mathcal{H}^n(Z)])^2 + \frac{\alpha m^2}{\pi R^4}\int_{\mathbf{K}^2}\int_{Z(s)}\int_{Z(\widetilde{s})}\int_{\mathbb{R}^{d}}\int_{B_{\frac{2R}{r}}(\frac{y-\widetilde{y}}{r})} k(z)k(z+w) \\
&\qquad \mathcal{H}^2(B_R(0)\cap B_R((||\widetilde{y}-y+rw||,0)))\mathrm{d}w\mathrm{d}z\mathcal{H}^n(\mathrm{d}\widetilde{y})\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)Q(\mathrm{d}\widetilde{s}) \\
&\underset{(u:=\frac{r}{2R}w+\frac{\widetilde{y}-y}{2R})}{\leq} \quad \alpha^2m^2(\mathbb{E}[\mathcal{H}^n(Z)])^2 + \frac{4\alpha m^2}{\pi R^2 r^2}(\mathbb{E}[\mathcal{H}^n(Z)])^2\int_{B_1(0)} M\mathcal{H}^2(B_R(0))\mathrm{d}u \\
&= \quad (\mathbb{E}[\mathcal{H}^n(Z)])^2(\alpha^2m^2 + \frac{4M\pi\alpha m^2}{r^2}) \longrightarrow (\mathbb{E}[\mathcal{H}^n(Z)])^2\alpha^2m^2, \quad \text{as } r\to\infty.
\end{aligned}
$$

Finally, being $I(r) = I_1(r) + I_2(r) > (\mathbb{E}[\mathcal{H}^n(Z)])^2\alpha^2m^2$ for any $r > 0$, and $\lim_{r\to\infty} I(r) = (\mathbb{E}[\mathcal{H}^n(Z)])^2\alpha^2m^2$, we may conclude that, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$, $r^{\mathrm{o,MSE}}(x) := \arg\min_r MSE(\widehat{\lambda}^{\kappa,N}_{\Theta_n}) = +\infty$.

4. Optimal bandwidths of $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$: Boolean models case

As mentioned at the end of Section 2, the "Minkowski content"-based estimator $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$ of $\lambda_{\Theta_n}(x)$. defined in (7) is asymptotically unbiased and weakly consistent if the bandwidth r_N is such that $\lim_{N\to\infty} r_N = 0$ and $\lim_{N\to\infty} N r_N^{d-n} = \infty$. It is then worth to carry on the analysis of such an estimator too, by facing the left open problem of finding an optimal bandwidth for $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$ (see [51, Sec. 6]).

A first difficulty arises, due to the fact that it does not seem possible to get a Taylor series expansion of $Bias(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x))$ without any further assumption on the distribution of Θ; indeed it is easy to check that, for $\mathcal{H}^d$-a.e. $x \in \mathbb{R}^d$,

$$Bias(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)) = \frac{\mathbb{P}(x \in \Theta_{n_{\oplus r_N}})}{b_{d-n}r^{d-n}} - \lambda_{\Theta_n}(x) \tag{24}$$

$$Var(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)) = \frac{\mathbb{P}(x \in \Theta_{n\oplus r_N})(1-\mathbb{P}(x \in \Theta_{n_{\oplus r_N}}))}{N(b_{d-n}r_N^{d-n})^2} = \frac{\lambda_{\Theta_n}(x)}{N r_N^{d-n} b_{d-n}} + o\Big(\frac{1}{N r_N^{d-n}}\Big). \tag{25}$$

Then it is clear that, if we wish to proceed as in the previous sections of this paper, an explicit expression for $\mathbb{P}(x \in \Theta_{n\oplus r_N})$ is required for evaluating the MSE; this is the reason why, for the time being, we restrict our analysis to particular classes of germ-grain models Θ_n which allow to explicit the $Bias(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x))$.

Let Θ_n be a Boolean model with intensity measure $\Lambda(\mathrm{d}(y,s)) = \lambda(y,s)\mathrm{d}yQ(\mathrm{d}s)$; then

$$\mathbb{P}(x \in \Theta_{n_{\oplus r_N}}) = 1 - \exp\Big\{-\int_{\mathbf{K}}\int_{x-Z(s)_{\oplus r_N}} \lambda(y,s)\mathrm{d}yQ(\mathrm{d}s)\Big\},$$

and so a Taylor expansion of $\int_{x-Z(s)_{\oplus r_N}} \lambda(y,s)\mathrm{d}y$ is needed. General expressions are not available in literature so far, therefore such a problem is still open. We only mention here that a possible solution might follow, under suitable regularity assumptions on the grains, by an application of the general Steiner-type formula for closed sets (see [35, Theorem 2.1]). Nevertheless, if the "shape" of the grains is known, it is possible to evaluate directly the integral above (see, for instance, the example discussed in Section 5). We may notice that the case in which Θ is homogeneous and grains have positive reach might be handled in a more direct way, by using the well known polynomial expansion of the volume of an enlarged compact set with positive reach [29]. We remind that the reach of a compact set $A \subset \mathbb{R}^d$ is defined by

$$\mathrm{reach}(A) := \inf_{a\in A} \sup\{r > 0 \,:\, B_r(a) \subset \mathrm{Unp}(A)\},$$

where $\mathrm{Unp}(A) := \{x \in \mathbb{R}^d \,:\, \exists!\, a \in A \text{ such that } \mathrm{dist}(x,A) = ||a-x||\}$ is the set of points having a unique projection on A. For any compact set $A \subset \mathbb{R}^d$ with positive reach, the *total curvature measures* $\Phi_i(A) \in \mathbb{R}$ for $i = 1,\dots,d-1$, introduced in [29], are well defined, and the following global Steiner formula holds

$$\mathcal{H}^d(A_{\oplus r}) = \sum_{i=0}^{d} r^{d-i} b_{d-i}\Phi_i(A), \quad \forall r < \mathrm{reach}(A). \tag{26}$$

We also point out that if $\dim(A) = n$, then $\Phi_i(A) = 0$, for any $i > n$, and $\Phi_n(A) = \mathcal{H}^n(A)$. Then, the following assertion is easily proved.

Proposition 18. *Let Θ_n be a Boolean model with intensity measure $\Lambda(\mathrm{d}(y,s)) = c\mathrm{d}yQ(\mathrm{d}s)$, satisfying Assumption (A1), and such that, for any $s \in \mathbf{K}$, $\mathrm{reach}Z(s) > R$, 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 the estimator* (7) *is given by*

$$r_N^{\mathrm{o,AMSE}} := \begin{cases} \sqrt[3]{\dfrac{c\mathbb{E}[\mathcal{H}^n(Z)]}{N\big(\pi c\mathbb{E}[\Phi_{n-1}(Z)] - 2(c\mathbb{E}[\mathcal{H}^n(Z)])^2\big)^2}} & \textit{if } d-n=1, \\ \sqrt[d-n+2]{\dfrac{(d-n)b_{d-n}c\mathbb{E}[\mathcal{H}^n(Z)]}{2N\big(cb_{d-n+1}\mathbb{E}[\Phi_{n-1}(Z)]\big)^2}} & \textit{if } d-n>1, \end{cases} \tag{27}$$

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

Proof. It is easy to check that Θ_n satisfies Assumptions $(A1)$, $(A2)$ and $(A3)$, so that

$$\lim_{N\to\infty} \widehat{\lambda}^{\mu,N}_{\Theta_n}(x) = c\mathbb{E}[\mathcal{H}^n(Z)] = \lambda_{\Theta_n}(x) \qquad \text{for all } x \in \mathbb{R}^d.$$

Moreover, for N is sufficiently large so that $r_N < R$, by (26) we get:

$$\begin{aligned} \lim_{N\to\infty} \frac{Bias(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x))}{r_N} &\overset{(24)}{=} \lim_{N\to\infty} \frac{1}{r_N}\Big(\frac{1-\exp\big\{-c\mathbb{E}[\mathcal{H}^d(Z_{\oplus r_N})]\big\}}{b_{d-n}r_N^{d-n}} - c\mathbb{E}[\mathcal{H}^n(Z)]\Big) \\ &\overset{(26)}{=} \lim_{N\to\infty} \frac{1}{b_{d-n}r_N^{d-n+1}}\Big(cr_N^{d-n+1}b_{d-n+1}\mathbb{E}[\Phi_{n-1}(Z)] + o(r_N^{d-n+1}) \\ &\qquad -\frac{1}{2}c^2b_{d-n}^2(\mathbb{E}[\Phi_n(Z)])^2r_N^{2(d-n)} + o(r_N^{2(d-n)})\Big) \\ &= \begin{cases} \dfrac{\pi c\mathbb{E}[\Phi_{n-1}(Z)]}{2} - (c\mathbb{E}[\mathcal{H}^n(Z)])^2 & \text{, if } d-n=1 \\ \dfrac{cb_{d-n+1}\mathbb{E}[\Phi_{n-1}(Z)]}{b_{d-n}} & \text{, if } d-n>1; \end{cases} \end{aligned}$$

consequently the asymptotic mean square error of $\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)$ is given by

$$AMSE(r_N) \overset{(25)}{=} \begin{cases} \dfrac{\big(\pi c\mathbb{E}[\Phi_{n-1}(Z)] - 2(c\mathbb{E}[\mathcal{H}^n(Z)])^2\big)^2}{4}r_N^2 + \dfrac{c\mathbb{E}[\mathcal{H}^n(Z)]}{2Nr_N} & \text{, if } d-n=1, \\ \Big(\dfrac{cb_{d-n+1}\mathbb{E}[\Phi_{n-1}(Z)]}{b_{d-n}}\Big)^2 r_N^2 + \dfrac{c\mathbb{E}[\mathcal{H}^n(Z)]}{Nr_N^{d-n}b_{d-n}} & \text{, if } d-n>1, \end{cases}$$

which is independent of $x \in \mathbb{R}^d$ as expected, being the process stationary.
By defining now $r_N^{\mathrm{o,AMSE}} := \arg\min_{r_N} AMSE(r_N)$, the assertion follows. □

5. A case studied: inhomogenous segment Boolean model

As a simple example of applicability of the above results, let us consider the segment Boolean model extensively studied in [51, Example 2] (see also [7, page 86]), where an explicit expression for its mean density has been obtained. We provide here the pointwise and the uniform optimal bandwidth $r_N^{\mathrm{o,AMSE}}(x)$ and $r_N^{\mathrm{o,AMISE}}$, respectively associated with both $\widehat{\lambda}^{\nu,N}_{\Theta_n}$ and $\widehat{\lambda}^{\mu,N}_{\Theta_n}$.

Let Θ_1 be an inhomogeneous Boolean model of segments in $\mathbb{R}^2$ with random length L and uniform orientation; so that the mark space is $\mathbf{K} = \mathbb{R}_+ \times [0, 2\pi]$; for all $s = (l, \alpha) \in \mathbf{K}$, let $Z(s) := \{(u, v) \in \mathbb{R}^2 \,:\, u = \tau\cos\alpha,\ v = \tau\sin\alpha,\ \tau \in [0, l]\}$ be the segment with length $l \in \mathbb{R}_+$, and orientation $\alpha \in [0, 2\pi]$. Denoted by $\mathbb{P}_L(\mathrm{d}l)$ the probability law of the random length L, we assume that $\int_{\mathbb{R}_+} l^3\mathbb{P}_L(\mathrm{d}l) < \infty$. Finally the segment process Θ_1, represented as in (1), is driven by the marked Poisson process Φ in $\mathbb{R}^2 \times \mathbf{K}$ having intensity measure $\Lambda(\mathrm{d}y \times \mathrm{d}s) = f(y)\mathrm{d}yQ(\mathrm{d}s)$, with $f(y) = f(y_1, y_2) = y_1^2 + y_2^2$, and $Q(\mathrm{d}s) = \frac{1}{2\pi}\mathrm{d}\alpha\mathbb{P}_L(\mathrm{d}l)$. It is easy to check that the assumptions of all theorems above are fulfilled, and that

$$\lambda_{\Theta_1}(x_1, x_2) = (x_1^2 + x_2^2)\mathbb{E}[L] + \frac{1}{3}\mathbb{E}[L^3], \quad \mathcal{H}^2\text{-a.e. } x = (x_1, x_2) \in \mathbb{R}^2.$$

(i) **optimal bandwidth of $\widehat{\lambda}^{\nu,N}_{\Theta_n}$.**
By noticing that $\pi_y^{x,s} = \{(y_1, y_2) \in \mathbb{R}^2 \,:\, y_1 = t\cos\alpha,\ y_2 = t\sin\alpha,\ t \in \mathbb{R}\}$, and that $\mathcal{H}^2(B_1(0) \cap B_1(-(t\cos\alpha, t\sin\alpha))) = \mathcal{H}^2(B_1(0) \cap B_1((t, 0)))$ for any $\alpha \in [0, 2\pi]$, we get

$$\begin{aligned}
C_{Var}(x) &\overset{(14)}{=} \frac{1}{2\pi^3}\int_0^\infty\int_0^{2\pi}\int_{\mathbb{R}^2}\int_0^l\int_{-\infty}^{+\infty}\mathbf{1}_{B_1(0)}(z) \\
&\qquad \mathbf{1}_{B_1(-(t\cos\alpha, t\sin\alpha))}(z)\lambda(x_1 - \tau\cos\alpha, x_2 - \tau\sin\alpha)\mathrm{d}t\mathrm{d}\tau\mathrm{d}z\mathrm{d}\alpha\mathbb{P}_L(\mathrm{d}l) \\
&= \frac{1}{\pi^3}\int_0^\infty\int_0^{2\pi}\int_0^l\int_0^{+2}\mathcal{H}^2(B_1(0) \cap B_1((t, 0)))\mathrm{d}t\lambda(x_1 - \tau\cos\alpha, x_2 - \tau\sin\alpha)\mathrm{d}\tau\mathrm{d}\alpha\mathbb{P}_L(\mathrm{d}l) \\
&= \frac{1}{\pi^3}\int_0^\infty\int_0^{2\pi}\int_0^l\frac{8}{3}((x_1 - \tau\cos\alpha)^2 + (x_2 - \tau\sin\alpha)^2)\mathrm{d}\tau\mathrm{d}\alpha\mathbb{P}_L(\mathrm{d}l) \\
&= \frac{16}{3\pi^2}((x_1^2 + x_2^2)\mathbb{E}[L] + \frac{1}{3}\mathbb{E}[L^3]) \qquad \Big(= \frac{16}{3\pi^2}\lambda_{\Theta_1}(x)\Big),
\end{aligned}$$

and

$$C_{Bias}(x) \overset{(13)}{=} \mathbb{E}[L]\int_{B_1(0)}\frac{1}{\pi}(z_1^2 + z_2^2)\mathrm{d}z = \frac{\mathbb{E}[L]}{2};$$

thus, by Eq (17),

$$r_N^{o,AMSE}(x) = \sqrt[5]{\frac{16(\mathbb{E}L(x_1^2 + x_2^2) + \frac{1}{3}\mathbb{E}L^3)}{3\pi^2 N(\mathbb{E}[L])^2}}.$$

Let us now consider the window $W = [0, 1] \times [0, 1]$; then the uniform optimal bandwidth in W is given by

$$r_N^{o,AMISE} \overset{(19)}{=} \sqrt[5]{\frac{16(\mathbb{E}[L^3] + 2\mathbb{E}[L])}{N9\pi^2(\mathbb{E}[L])^2}}.$$

(ii) **optimal bandwidth of $\widehat{\lambda}^{\mu,N}_{\Theta_n}$.**

Let us observe that

$$
\begin{aligned}
&Bias(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)) \\
&\overset{(24)}{=} \frac{1-\exp\{2r_N\mathbb{E}[L](x_1^2+x_2^2)+\frac{2}{3}r_N\mathbb{E}[L^3]+\frac{\pi}{2}r_N^2(2x_1^2+2x_2^2+\mathbb{E}[L^2])+2\mathbb{E}[L]r_N^3+\frac{\pi}{2}r_N^4\}}{2r_N} \\
&\qquad\qquad\qquad\qquad -(x_1^2+x_2^2)\mathbb{E}[L]+\frac{1}{3}\mathbb{E}[L^3] \\
&= \big(\frac{\pi}{2}(x_1^2+x_2^2)+\frac{\pi}{4}\mathbb{E}[L^2]-(\mathbb{E}[L](x_1^2+x_2^2)+\frac{1}{3}\mathbb{E}[L^3])^2\big)r_N+o(r_N),
\end{aligned}
$$

and

$$
Var(\widehat{\lambda}^{\mu,N}_{\Theta_n}(x)) \overset{(25)}{=} \frac{(x_1^2+x_2^2)\mathbb{E}[L]+\frac{1}{3}\mathbb{E}[L^3]}{N2r_N}+o\Big(\frac{1}{Nr_N}\Big);
$$

then,

$$
AMSE(r_N)=\big(\frac{\pi}{2}(x_1^2+x_2^2)+\frac{\pi}{4}\mathbb{E}[L^2]-(\mathbb{E}[L](x_1^2+x_2^2)+\frac{1}{3}\mathbb{E}[L^3])^2\big)^2r_N^2+\frac{(x_1^2+x_2^2)\mathbb{E}[L]+\frac{1}{3}\mathbb{E}[L^3]}{N2r_N};
$$

and so we get

$$
r_N^{o,AMSE}(x)=\sqrt[3]{\frac{(x_1^2+x_2^2)\mathbb{E}[L]+\frac{1}{3}\mathbb{E}[L^3]}{4N\big(\frac{\pi}{2}(x_1^2+x_2^2)+\frac{\pi}{4}\mathbb{E}L^2-(\mathbb{E}L(x_1^2+x_2^2)+\frac{1}{3}\mathbb{E}L^3)^2\big)^2}}.
$$

Let us now consider the window $W=[0,1]\times[0,1]$; then it is not difficult to obtain

$$
\begin{aligned}
MISE(\widehat{\lambda}^{\mu,N}_{\Theta_1}(W)) &:= \int_W MSE(\widehat{\lambda}^{\mu,N}_{\Theta_1}(x))\mathrm{d}x \\
&= \int_W [Bias(\widehat{\lambda}^{\mu,N}_{\Theta_1}(x))]^2\mathrm{d}x+\int_W Var[\widehat{\lambda}^{\mu,N}_{\Theta_1}(x)]\mathrm{d}x \\
&= \underbrace{r_N^2\int_W C_B^2(x)\mathrm{d}x+\frac{1}{Nr_N}\int_W\frac{\lambda_{\Theta_1}(x)}{2}\mathrm{d}x}_{=:AMISE(\widehat{\lambda}^{\mu,N}_{\Theta_1}(W))}+o(r_N^2)+o\big(\frac{1}{Nr_N}\big),
\end{aligned}
$$

and so

$$
r_N^{o,AMISE}:=\arg\min_{r_N} AMISE(\widehat{\lambda}^{\mu,N}_{\Theta_1}(W))=\sqrt[3]{\frac{\frac{2}{3}\mathbb{E}[L]+\frac{1}{3}\mathbb{E}[L^3]}{4N\int_W C^2_{Bias}(x)\mathrm{d}x}},
$$

where

$$
\begin{aligned}
\int_W C^2_{Bias}(x)\mathrm{d}x &= \frac{1328}{1575}(\mathbb{E}[L])^4-\frac{48}{35}\Big(\frac{\pi}{2}-\frac{2}{3}\mathbb{E}[L]\mathbb{E}[L^3]\Big)(\mathbb{E}[L])^2 \\
&\quad -\frac{56}{45}\Big(\frac{\pi}{4}\mathbb{E}[L^2]-\frac{(\mathbb{E}[L^3])^2}{9}\Big)(\mathbb{E}[L])^2+\frac{28}{45}\Big(\frac{\pi}{2}-\frac{2}{3}\mathbb{E}[L]\mathbb{E}[L^3]\Big)^2 \\
&\quad +\frac{4}{3}\Big(\frac{\pi\mathbb{E}[L^2]}{4}-\frac{(\mathbb{E}[L^3])^2}{9}\Big)\Big(\frac{\pi}{2}-\frac{2}{3}\mathbb{E}[L]\mathbb{E}[L^3]\Big)+\Big(\frac{\pi}{4}\mathbb{E}[L^2]-\frac{1}{9}(\mathbb{E}[L^3])^2\Big)^2.
\end{aligned}
$$

Remark 19. As already mentioned in Section 4, the evaluation of the optimal bandwidth of the estimator $\widehat{\lambda}^{\mu,N}_{\Theta_n}$ simplifies when the Boolean model Θ_n is homogeneous. Indeed, let us consider for instance the above Boolean model of segments, for $f(y) \equiv c > 0$. By observing that reach$Z(s) = \infty$, and $\Phi_0(Z(s)) = 1$, for any $s \in \mathbf{K}$, the optimal bandwidth is then given by

$$r_N^{\mathrm{o,AMSE}} \stackrel{(27)}{=} \sqrt[3]{\frac{c\mathbb{E}[L]}{N(c\pi - 2(c\mathbb{E}[L])^2)^2}}.$$

6. Concluding Remarks

Based on the analysis carried out in the previous sections, we may conclude with the following remarks.

- Kernel estimators: the *pro* for the kernel estimators $\widehat{\lambda}^{\kappa,N}_{\Theta_n}$ proposed in Section 2 extend in a natural way the corresponding kernel estimators for random objects of dimension $n = 0$ (random variables - univariate and multivariate, point processes) to random closed sets of any integer Hausdorff dimension $n < d$, in a space $\mathbb{R}^d$; the *cons* concern the practical applicability of them, due not only to the enhanced computational problems related to the higher dimensionality of the relevant objects, but also due to the problems encountered for the case $n = 0$, in connection with the preliminary estimation of bounds of the relevant parameters of the unknown density. This would require further research about possible extensions of plug-in methods as exploited for the case $n = 0$ (e.g., see [49] and references therein).

- "Minkowski content"-based estimators: the *pro* for the estimators $\widehat{\lambda}^{\mu,N}_{\Theta_n}$ proposed in Section 2 concerns its easy computational evaluation; the *cons* include the analytical difficulty of evaluating an optimal bandwidth.

- Natural estimators: the *pro* for the natural estimators $\widehat{\lambda}^{\nu,N}_{\Theta_n}$ proposed in Section 2 concerns their direct derivation from the Besicovitch Theorem; they generalizes the notion of histogram estimators for the case $n = 0$ (see Remark 2); the *cons* include the nontrivial evaluation of $\mathcal{H}^n(\Theta_n^i \cap B_{r_N}(x))$ for any element Θ_n^i of the sample; for segment processes ($n = 1$) it seems more feasible, but for other sets of dimension $n \geq 1$ it results of higher computational complexity. This problem has been already raised in [41] even for stationary fibre processes in $\mathbb{R}^2$.

  It is clear that we have left open a series of problems towards which we address the attention of readers for further research.

  Here we wish to evidence the need of analyzing the very realistic case in which only one realization of the geometric process is available, and only in a bounded window, as it frequently happens in material science and medicine. This problem is still of great interest in current research, even for the case $n = 0$, in particular for point processes (see e.g. [33, 32] and references therein). In this case there is a need of additional properties for the process, such as stationarity, ergodicity, mixing properties, allowing a possible "increasing domain asymptotics" (see e.g. [20, page 480], [7, page 88] and references therein).

7. Proofs of the main results

Proof of Theorem 3
Lemma 3 in [52] tells us that the event that different grains of Θ_n overlap in a subset of $\mathbb{R}^d$ of positive $\mathcal{H}^n$-measure has null probability; therefore, the following chain of equalities holds:

$$\begin{aligned}
\mathbb{E}[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)] &= \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^{N} k_{r_N} * \mathcal{H}^n_{|\Theta_n^i}(x)\Big] = \mathbb{E}\Big[\frac{1}{r_N^d}\int_{\Theta_n} k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\Big] \\
&= \mathbb{E}\Big[\frac{1}{r_N^d}\sum_{(x_i,s_i)\in\Phi}\int_{(x_i+Z(s_i))} k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\Big] \\
&\overset{(A.5)}{=} \int_{\mathbf{K}}\int_{\mathbb{R}^d}\frac{1}{r_N^d}\int_{w+Z(s)} k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\lambda(w,s)\mathrm{d}wQ(\mathrm{d}s) \\
&= \int_{\mathbf{K}}\int_{\mathbb{R}^d}\int_{Z(s)}\frac{1}{r_N^d}k\Big(\frac{x-(w+y)}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\lambda(w,s)\mathrm{d}wQ(\mathrm{d}s) \\
&= \int_{\mathbb{R}^d}\int_{\mathbf{K}}\int_{x-Z(s)} k(z)\lambda(y-r_Nz,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\mathrm{d}z \qquad (28)
\end{aligned}$$

By hypothesis we know that k has compact support, say $S\subset B_R(0)$ for some $R>0$; besides, by denoting $\mathcal{D}(s) := \mathrm{disc}(\lambda(\cdot,s))$, we have $\mathcal{H}^n(\mathcal{D}(s))=0$ for any $s\in\mathbf{K}$. Therefore,

$$(28) = \int_{\mathbb{R}^d}\int_{\mathbf{K}}\int_{(x-Z(s))\setminus\mathcal{D}(s)} k(z)\lambda(y-r_Nz,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\mathrm{d}z.$$

Note that, for N sufficiently large, $B_R(x)\supset B_{Rr_N}(x)$, and so

$$\sup_{y\in(x-Z(s))\setminus\mathcal{D}(s)}\lambda(y-r_Nz,s) \le \sup_{y\in B_R(x)_{\oplus\mathrm{diam}Z(s)}}\lambda(y,s) \overset{(A2)}{\le} \widetilde{\xi}_{B_R(x)}(s),$$

$$\lim_{N\to\infty} k(z)\lambda(y-r_Nz,s) = k(z)\lambda(y,s) < \infty, \qquad \forall y\in(x-Z(s))\setminus\mathcal{D}(s), \qquad (29)$$

so that, by $(A2)$,

$$\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbf{K}}\int_{(x-Z(s))\setminus\mathcal{D}(s)} k(z)\lambda(y-r_Nz,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\mathrm{d}z \\
&\qquad \le \int_{B_R(0)}\int_{\mathbf{K}}\int_{(x-Z(s))} k(z)\widetilde{\xi}_{B_R(x)}(s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\mathrm{d}z \\
&\qquad\qquad \le \int_{B_R(0)} k(z)\mathrm{d}z\int_{\mathbf{K}}\widetilde{\xi}_{B_R(x)}(s)\mathcal{H}^n(\Xi(s))Q(\mathrm{d}s) < \infty.
\end{aligned}$$

Hence, we may apply the Dominated Convergence Theorem to (28) in order to obtain

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)] &= \lim_{N\to\infty}\int_{\mathbb{R}^d}\int_{\mathbf{K}}\int_{(x-Z(s))\setminus\mathcal{D}(s)} k(z)\lambda(y-r_N z,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\mathrm{d}z\\ &\overset{(29)}{=} \int_{\mathbb{R}^d} k(z)\mathrm{d}z\int_{\mathbf{K}}\int_{(x-Z(s))\setminus\mathcal{D}(s)}\lambda(y,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\\ &= \int_{\mathbf{K}}\int_{(x-Z(s))}\lambda(y,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s) \overset{(4)}{=} \lambda_{\Theta_n}(x),\end{aligned}$$

thus proving that $\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)$ is asymptotically unbiased for $\mathcal{H}^d$-a.e. $x\in\mathbb{R}^d$. □

Proof of Proposition 4

Let $C\subset\mathbb{R}^d$, $a\in\mathbb{R}^d$, and $\xi_{C,B_1(a)}$ satisfy (9). By Remark 4 in [52] we know that (A1) guarantees that

$$\mathcal{H}^d(Z(s)_{\oplus R})\le\begin{cases}\mathcal{H}^n(\Xi(s))\gamma^{-1}2^n4^d b_d R^{d-n} & \text{if } R<2\\ \mathcal{H}^n(\Xi(s))\gamma^{-1}2^n4^d b_d R^{n} & \text{if } R\ge 2\end{cases}. \tag{30}$$

Let $\xi_{a,C}(s,y,t):=\mathbf{1}_{(a-Z(t))_{\oplus 1}}(y)\xi_{C,B_1(a)}(s,t)$, for any $(s,y,t)\in\mathbf{K}\times\mathbb{R}^d\times\mathbf{K}$; then

$$\begin{aligned}\mathbf{1}_{(a-Z(t))_{\oplus 1}}(y)\sup_{x\in C_{\oplus\mathrm{diam}(Z(s))}} g(x,s,y,t) &\le \mathbf{1}_{(a-Z(t))_{\oplus 1}}(y)\sup_{y\in B_1(a)_{\oplus\mathrm{diam}(Z(t))}}\ \sup_{x\in C_{\oplus\mathrm{diam}(Z(s))}} g(x,s,y,t)\\ &\le \xi_{a,C}(s,y,t),\end{aligned}$$

and

$$\begin{aligned}&\int_{\mathbb{R}^d\times\mathbf{K}^2}\mathcal{H}^n(\Xi(s))\xi_{a,C}(s,y,t)\mathrm{d}yQ_{[2]}(\mathrm{d}s,\mathrm{d}t)\\ &\qquad=\int_{\mathbf{K}^2}\mathcal{H}^n(\Xi(s))\mathcal{H}^d((a-Z(t))_{\oplus 1})\xi_{C,B_1(a)}(s,t)Q_{[2]}(\mathrm{d}s,\mathrm{d}t)\\ &\qquad\overset{(30)}{\le}\gamma^{-1}2^n4^d b_d\int_{\mathbf{K}^2}\mathcal{H}^n(\Xi(s))\mathcal{H}^n(\Xi(t))\xi_{C,B_1(a)}(s,t)Q_{[2]}(\mathrm{d}s,\mathrm{d}t)<\infty\end{aligned}$$

□

Proof of Theorem 6

Let us notice that

$$
\begin{aligned}
&Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) = \frac{\mathbb{E}\big[(k_{r_N} * \mathcal{H}^n_{|\Theta_n}(x))^2\big] - \big(\mathbb{E}[k_{r_N} * \mathcal{H}^n_{|\Theta_n}(x)]\big)^2}{N} \\
&\le \frac{\mathbb{E}\big[(k_{r_N} * \mathcal{H}^n_{|\Theta_n}(x))^2\big]}{N} \\
&= \frac{1}{N}\mathbb{E}\Big[\sum_{(x_i,s_i)\in\Phi}\Big(\int_{x_i+Z(s_i)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\Big)^2\Big] + \\
&\quad \frac{1}{N}\mathbb{E}\Big[\sum_{\substack{(x_i,s_i),(x_j,s_j)\in\Phi,\\ x_i\neq x_j}}\int_{x_i+Z(s_i)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\int_{x_j+Z(s_j)}\frac{1}{r_N^d}k\Big(\frac{x-\widetilde{y}}{r_N}\Big)\mathcal{H}^n(\mathrm{d}\widetilde{y})\Big] \\
&\overset{(A.5),(A.6)}{=} \frac{1}{N}\int_{\mathbf{K}\times\mathbb{R}^d}\Big(\int_{w+Z(s)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\Big)^2\lambda(w,s)\mathrm{d}wQ(\mathrm{d}s) + \frac{1}{N}\int_{(\mathbf{K}\times\mathbb{R}^d)^2} \\
&\quad \Big(\int_{w+Z(s)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\int_{\widetilde{w}+Z(\widetilde{s})}\frac{1}{r_N^d}k\Big(\frac{x-\widetilde{y}}{r_N}\Big)\mathcal{H}^n(\mathrm{d}\widetilde{y})\Big)g(w,s,\widetilde{w},\widetilde{s})\mathrm{d}w\mathrm{d}\widetilde{w}Q_{[2]}(\mathrm{d}s,\mathrm{d}\widetilde{s}).
\end{aligned}
$$

We remind that $k(z) \le M\mathbf{1}_{B_R(0)}(z)$ for any $z\in\mathbb{R}^d$; for N sufficiently large so that $r_N \le \min\{1, 1/2R\}$, we get

$$
\begin{aligned}
&\frac{1}{N}\int_{\mathbf{K}\times\mathbb{R}^d}\Big(\int_{w+Z(s)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\Big)^2\lambda(w,s)\mathrm{d}wQ(\mathrm{d}s) \\
&= \frac{1}{Nr_N^{2d}}\int_{\mathbf{K}}\int_{\mathbb{R}^d}\int_{Z(s)}\int_{Z(s)}k\Big(\frac{x-y-w}{r_N}\Big)k\Big(\frac{x-\widetilde{y}-w}{r_N}\Big)\lambda(w,s)\mathcal{H}^n(\mathrm{d}y)\mathcal{H}^n(\mathrm{d}\widetilde{y})\mathrm{d}wQ(\mathrm{d}s) \\
&= \frac{1}{Nr_N^d}\int_{\mathbf{K}}\int_{Z(s)}\int_{Z(s)}\int_{\mathbb{R}^d}k(z)k\Big(z+\frac{y-\widetilde{y}}{r_N}\Big)\lambda(x-y-r_Nz,s)\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)\mathcal{H}^n(\mathrm{d}\widetilde{y})Q(\mathrm{d}s) \qquad (31)\\
&\le \frac{1}{Nr_N^d}\int_{\mathbf{K}}\int_{\mathbb{R}^d}\int_{Z(s)}k(z)\mathbf{1}_{B_R(0)}(z)\lambda(x-y-r_Nz,s)\int_{Z(s)}M\mathbf{1}_{B_{r_NR}(y+r_Nz)}(\widetilde{y})\mathcal{H}^n(\mathrm{d}\widetilde{y})\mathcal{H}^n(\mathrm{d}y)\mathrm{d}zQ(\mathrm{d}s) \\
&\le \frac{1}{Nr_N^d}\int_{\mathbf{K}}\int_{\mathbb{R}^d}\int_{Z(s)}k(z)\mathbf{1}_{B_R(0)}(z)\sup_{\xi\in x-Z(s)-r_Nz}\lambda(\xi,s)M\mathcal{H}^n(\Xi(s)\cap B_{r_NR}(y+r_Nz))\mathcal{H}^n(\mathrm{d}y)\mathrm{d}zQ(\mathrm{d}s) \\
&\le \frac{1}{Nr_N^d}\int_{\mathbf{K}}\int_{Z(s)}\widetilde{\xi}_{B_R(x)}(s)M\mathcal{H}^n(\Xi(s)\cap B_{2r_NR}(y))\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s) \\
&\overset{(\overline{A1})}{\le} \frac{1}{Nr_N^d}\int_{\mathbf{K}}\int_{Z(s)}\widetilde{\xi}_{B_R(x)}(s)M\widetilde{\gamma}(2r_NR)^n\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s) \\
&\le \frac{M\widetilde{\gamma}2^nR^n}{Nr_N^{d-n}}\int_{\mathbf{K}}\mathcal{H}^n(\Xi(s))\widetilde{\xi}_{B_R(x)}(s)Q(\mathrm{d}s) \overset{(A2)}{<} \infty.
\end{aligned}
$$

Similarly,

$$\frac{1}{N}\int_{(\mathbf{K}\times\mathbb{R}^d)^2}\Big(\int_{w+Z(s)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\int_{\widetilde{w}+Z(\widetilde{s})}\frac{1}{r_N^d}k\Big(\frac{x-\widetilde{y}}{r_N}\Big)\mathcal{H}^n(\mathrm{d}\widetilde{y})\Big)g(w,s,\widetilde{w},\widetilde{s})\mathrm{d}w\mathrm{d}\widetilde{w}Q_{[2]}(\mathrm{d}s,\mathrm{d}\widetilde{s})$$
$$\leq\frac{1}{N}\int_{B_R(0)}k(z)\mathrm{d}z\int_{B_R(0)}k(\widetilde{z})\mathrm{d}\widetilde{z}$$
$$\int_{\mathbf{K}^2}\mathcal{H}^n(\Xi(s))\mathcal{H}^n(\Xi(\widetilde{s}))\sup_{y\in B_R(x)_{\oplus\mathrm{diam}Z(s)}}\ \sup_{\widetilde{y}\in B_R(x)_{\oplus\mathrm{diam}Z(\widetilde{s})}}g(y,s,\widetilde{y},\widetilde{s})Q_{[2]}(\mathrm{d}s,\mathrm{d}\widetilde{s})$$
$$\leq\frac{1}{N}\int_{\mathbf{K}^2}\mathcal{H}^n(\Xi(s))\mathcal{H}^n(\Xi(\widetilde{s}))\xi_{B_R(x),B_R(x)}(s,\widetilde{s})Q_{[2]}(\mathrm{d}s,\mathrm{d}\widetilde{s})\overset{(\overline{A3})}{<}\infty,$$

so that the assertion follows. □

Proof of Theorem 8

Let us observe that $\int_{\mathbb{R}^d}z_ik(z)\mathrm{d}z=0$, being k radially symmetric; thus, by a Taylor series expansion we get:

$$\begin{aligned}
&Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)):=\mathbb{E}[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)]-\lambda_{\Theta_n}(x)=\\
&\overset{(28)}{=}\int_{\mathbf{K}}\int_{Z(s)}\int_{\mathbb{R}^d}k(z)\lambda(x-y-r_Nz,s)\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)-\lambda_{\Theta_n}(x)\\
&=\int_{\mathbf{K}}\int_{Z(s)}\int_{\mathbb{R}^d}k(z)\Big[\lambda(x-y,s)-\sum_{|\alpha|=1}\frac{1}{\alpha!}D^\alpha_x\lambda(x-y,s)z^\alpha r_N\\
&\qquad\qquad+\sum_{|\alpha|=2}\frac{1}{\alpha!}D^\alpha_x\lambda(x-y-\theta r_Nz,s)z^\alpha r_N^2\Big]\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)-\lambda_{\Theta_n}(x)\\
&\overset{(4)}{=}r_N^2\int_{\mathbf{K}}\int_{Z(s)}\int_{\mathbb{R}^d}k(z)\sum_{|\alpha|=2}\frac{1}{\alpha!}D^\alpha_x\lambda(x-y-\theta r_Nz,s)z^\alpha\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s),
\end{aligned}$$

with $\theta\in(0,1)$ depending on x, y, z, s.
By hypothesis, for any fixed $s\in\mathbf{K}$ and $y\notin\bigcup_{|\alpha|=2}\mathcal{D}^{(\alpha)}(s)$, with $\mathcal{H}^n(\cup_{|\alpha|=2}\mathcal{D}^{(\alpha)}(s))=0$, we have that

$$\lim_{N\to\infty}k(z)\mathbf{1}_{x-Z(s)}(y)\sum_{|\alpha|=2}\frac{1}{\alpha!}D^\alpha_x\lambda(y-\theta r_Nz,s)z^\alpha=k(z)\mathbf{1}_{x-Z(s)}(y)\sum_{|\alpha|=2}\frac{1}{\alpha!}D^\alpha_x\lambda(y,s)z^\alpha;$$

moreover, since $k(z)\leq M\mathbf{1}_{B_R(0)}(z)$ for any $z\in\mathbb{R}^d$, for N sufficiently large so that $r_N<1$, the following holds

$$\begin{aligned}
&\int_{\mathbf{K}}\int_{Z(s)}\int_{\mathbb{R}^d}\Big|k(z)\sum_{|\alpha|=2}\frac{1}{\alpha!}D_x^\alpha\lambda(x-y-\theta r_N z,s)z^\alpha\Big|\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\\
&\qquad\le\int_{\mathbf{K}}\int_{Z(s)}\int_{\mathbb{R}^d}M\mathbf{1}_{B_R(0)}(z)\sum_{|\alpha|=2}\frac{1}{\alpha!}\sup_{\xi\in x-Z(s)-\theta r_N z}|D_x^\alpha\lambda(\xi,s)|\cdot|z^\alpha|\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)\\
&\qquad\qquad\overset{(A2\text{bis})}{\le}\sum_{|\alpha|=2}\frac{M}{\alpha!}\int_{B_R(0)}|z^\alpha|\mathrm{d}z\int_{\mathbf{K}}\mathcal{H}^n(\Xi(s))\widetilde{\xi}^{(\alpha)}_{B_R(x)}(s)Q(\mathrm{d}s)<+\infty. \qquad (32)
\end{aligned}$$

By applying now the Dominated Convergence Theorem we get

$$\lim_{N\to\infty}\frac{Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))}{r_N^2}=\sum_{|\alpha|=2}\frac{1}{\alpha!}\int_{\mathbb{R}^d}k(z)z^\alpha\mathrm{d}z\int_{\mathbf{K}}\int_{x-Z(s)}D_y^\alpha\lambda(y,s)\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)=C_{Bias}(x).$$

As far as the variance is concerned, by the proof of Theorem 6 we know that:

$$\begin{aligned}
&Var[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)]\\
&=\frac{\mathbb{E}\big[(k_{r_N}*\mathcal{H}^n_{|\Theta_n}(x))^2\big]-\big(\mathbb{E}[k_{r_N}*\mathcal{H}^n_{|\Theta_n}(x)]\big)^2}{N}\\
&=\frac{1}{N}\int_{\mathbf{K}}\int_{\mathbb{R}^d}\Big(\int_{w+Z(s)}\frac{1}{r_N^d}k\Big(\frac{x-y}{r_N}\Big)\mathcal{H}^n(\mathrm{d}y)\Big)^2\lambda(w,s)\mathrm{d}wQ(\mathrm{d}s)+O\Big(\frac{1}{N}\Big)\\
&\overset{(31)}{=}\frac{1}{Nr_N^d}\int_{\mathbf{K}}\int_{Z(s)}\int_{Z(s)}\int_{\mathbb{R}^d}k(z)k\Big(z+\frac{y-\widetilde{y}}{r_N}\Big)\lambda(x-y-r_Nz,s)\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)\mathcal{H}^n(\mathrm{d}\widetilde{y})Q(\mathrm{d}s)+O\Big(\frac{1}{N}\Big)\\
&=\frac{1}{Nr_N^{d-n}}\int_{\mathbf{K}}\int_{Z(s)}\int_{\mathbb{R}^d}k(z)\lambda(x-y-r_Nz,s)\Big(\int_{\frac{y-Z(s)}{r_N}}k(z+w)\mathcal{H}^n(\mathrm{d}w)\Big)\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)+O\Big(\frac{1}{N}\Big)\\
&=\frac{1}{Nr_N^{d-n}}\int_{\mathbf{K}}\int_{x-Z(s)}\int_{\mathbb{R}^d}k(z)\mathbf{1}_{B_R(0)}(z)\lambda(y-r_Nz,s)\mathbf{1}_{(\mathcal{D}(s))^c}(y)\\
&\qquad\qquad\int_{\frac{(x-Z(s))-y}{r_N}}k(z+w)\mathcal{H}^n(\mathrm{d}w)\mathrm{d}z\mathcal{H}^n(\mathrm{d}y)Q(\mathrm{d}s)+o\Big(\frac{1}{Nr_N^{d-n}}\Big),
\end{aligned}$$

being $\mathrm{supp}(k)\subset B_R(0)$, and $\mathcal{H}^n(\mathcal{D}(s))=0$ for any $s\in\mathbf{K}$, and $d>n$.
Let us observe that, by Theorem A.5, for any $s\in\mathbf{K},\ z\in\mathbb{R}^d$, and $\mathcal{H}^n$-a.e. $y\in x-Z(s)$,

$$\begin{aligned}
&\lim_{N\to\infty}k(z)\mathbf{1}_{B_R(0)}(z)\lambda(y-r_Nz,s)\mathbf{1}_{(\mathcal{D}(s))^c}(y)\int_{\frac{(x-Z(s))-y}{r_N}}k(z+w)\mathcal{H}^n(\mathrm{d}w)\\
&\qquad\qquad=k(z)\mathbf{1}_{B_R(0)}(z)\lambda(y,s)\mathbf{1}_{\mathcal{D}(s)}(y)\int_{\pi_y^{x,s}}k(z+w)\mathcal{H}^n(\mathrm{d}w), \qquad (33)
\end{aligned}$$

having denoted by $\pi_y^{x,s}$ the approximate tangent space to $x-Z(s)$ at y.

Moreover, for N sufficiently large so that $r_N \le \min\{1, \frac{1}{2R}\}$, we get

$$\int_{\mathbf{K}} \int_{x-Z(s)} \int_{\mathbb{R}^d} \Big(k(z) \mathbf{1}_{B_R(0)}(z) \lambda(y - r_N z, s) \mathbf{1}_{(\mathcal{D}(s))^c}(y) \int_{\frac{(x-Z(s))-y}{r_N}} k(z+w) \mathcal{H}^n(\mathrm{d}w) \Big) \mathrm{d}z \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s)$$

$$\le \int_{\mathbf{K}} \int_{x-Z(s)} \int_{\mathbb{R}^d} k(z) \mathbf{1}_{B_R(0)}(z) \sup_{\xi \in x - Z(s) - r_N z} \lambda(\xi, s) \frac{1}{r_N^n} \int_{x-Z(s)} k(z + (\widetilde{y} - y)/r_N) \mathcal{H}^n(\mathrm{d}\widetilde{y}) \mathrm{d}z \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s)$$

$$\le \frac{M^2}{r_N^n} \int_{\mathbf{K}} \int_{x-Z(s)} \int_{\mathbb{R}^d} \mathbf{1}_{B_R(0)}(z) \widetilde{\xi}_{B_R(x)}(s) \int_{x-\Xi(s)} \mathbf{1}_{B_{r_N R}(0)}(\widetilde{y} - y + r_N z) \mathcal{H}^n(\mathrm{d}\widetilde{y}) \mathrm{d}z \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s)$$

$$\le \frac{M^2 b_d R^d}{r_N^n} \int_{\mathbf{K}} \widetilde{\xi}_{B_R(x)}(s) \int_{x-Z(s)} \mathcal{H}^n((x - \Xi(s)) \cap B_{2r_N R}(y)) \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s)$$

$$\overset{(A1)}{\le} 2^n M^2 b_d R^{d+n} \widetilde{\gamma} \int_{\mathbf{K}} \widetilde{\xi}_{B_R(x)}(s) \mathcal{H}^n(\Xi(s)) Q(\mathrm{d}s) \overset{(A2)}{<} \infty.$$

Therefore, by applying the Dominated Convergence Theorem, we get

$$\begin{aligned} &\lim_{N\to\infty} N r_N^{d-n} Var[\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)] \\ &= \lim_{N\to\infty} \int_{\mathbf{K}} \int_{x-Z(s)} \int_{\mathbb{R}^d} k(z) \mathbf{1}_{B_R(0)}(z) \lambda(y - r_N z, s) \mathbf{1}_{(\mathcal{D}(s))^c}(y) \int_{\frac{(x-Z(s))-y}{r_N}} k(z+w) \mathcal{H}^n(\mathrm{d}w) \mathrm{d}z \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s) + o(1) \\ &= \int_{\mathbf{K}} \int_{x-Z(s)} \int_{\mathbb{R}^d} k(z) \mathbf{1}_{B_R(0)}(z) \lambda(y,s) \mathbf{1}_{\mathcal{D}(s)}(y) \int_{\pi_y^{x,s}} k(z+w) \mathcal{H}^n(\mathrm{d}w) \mathrm{d}z \mathcal{H}^n(\mathrm{d}y) Q(\mathrm{d}s) \\ &= \int_{\mathbf{K}} \int_{\mathbb{R}^d} \int_{x-Z(s)} \int_{\pi_y^{x,s}} k(z) k(z+w) \lambda(y,s) \mathcal{H}^n(\mathrm{d}w) \mathcal{H}^n(\mathrm{d}y) \mathrm{d}z Q(\mathrm{d}s) = C_{Var}(x). \end{aligned} \tag{34}$$

$\square$

Proof of Proposition 10

By the proof of Theorem 8, we know that, for N sufficiently large so that $r_N < 1$,

$$\frac{Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))^2}{r_N^4} \overset{(32)}{\le} \Big(\sum_{|\alpha|=2} \frac{M}{\alpha!} \int_{B_R(0)} |z^\alpha| \mathrm{d}z \int_{\mathbf{K}} \mathcal{H}^n(\Xi(s)) \widetilde{\xi}^{(\alpha)}_{W_{\oplus R}}(s) Q(\mathrm{d}s) \Big)^2 < +\infty \quad \mathcal{H}^d\text{-a.e. } x \in W,$$

which is integrable on W; thus, by (11) and the Dominated Convergence Theorem, we get

$$\lim_{N\to\infty} \frac{1}{r_N^4} \int_W [Bias(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))]^2 \mathrm{d}x = \int_W C^2_{Bias}(x) \mathrm{d}x. \tag{35}$$

Similarly, by the proof of Theorem 6, we deduce the following upper bound for the variance, for N sufficiently large so that $r_N \leq \min\{1, 1/(2R)\}$, $\mathcal{H}^d$-a.e. $x \in W$,

$$\begin{aligned} N r_N^{d-n} Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x)) &= r_N^{d-n}\mathbb{E}\big[(k_{r_N} * \mathcal{H}^n_{|\Theta_n}(x))^2\big] - r_N^{d-n}\big(\mathbb{E}[k_{r_N} * \mathcal{H}^n_{|\Theta_n}(x)]\big)^2 \\ &\leq M\widetilde{\gamma}2^n R^n \int_{\mathbf{K}} \mathcal{H}^n(\Xi(s))\widetilde{\xi}_{W\oplus R}(s)Q(\mathrm{d}s) \\ &\quad + r_N^{d-n}\int_{\mathbf{K}^2} \mathcal{H}^n(\Xi(s))\mathcal{H}^n(\Xi(\widetilde{s}))\xi_{W\oplus R, W\oplus R}(s,\widetilde{s})Q_{[2]}(\mathrm{d}s, \mathrm{d}\widetilde{s}) < \infty, \end{aligned}$$

which is integrable on W; thus, by (34) and the Dominated Convergence Theorem, we get

$$\lim_{N\to\infty} N r_N^{d-n}\int_W Var(\widehat{\lambda}^{\kappa,N}_{\Theta_n}(x))\mathrm{d}x = \int_W C_{Var}(x)\mathrm{d}x. \tag{36}$$

The assertion directly follows by (18), (35) and (36). $\square$

Appendix A. Background

Appendix A.1. Kernel density estimation of random variables

Let X be an absolutely continuous real random variable having p.d.f f, and let $\{X_i\}_{i\in\mathbb{N}}$ be a countable sample of X, i.e. a sequence of random variables i.i.d. as X. A (scaled) kernel density estimator $\widehat{f}^N_X(x)$ of $f(x)$, based on a kernel k, is defined as

$$\widehat{f}^N_X(x) := \frac{1}{N r_N}\sum_{i=1}^N k\Big(\frac{x - X_i}{r_N}\Big), \quad x \in \mathbb{R}. \tag{A.1}$$

A kernel is usually taken as a unimodal probability density function on the real line; in this case it is usually assumed that it satisfies the following conditions

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

We may notice that, since k is a kernel, the following holds

$$c_1 := \int k(z)^2 dz < +\infty.$$

We further assume that

$$c_2 := \int_{\mathbb{R}} z^2 k(z)dz \in (0, +\infty).$$

The scaling parameter $r_N \in (0, +\infty)$, also known as the bandwidth of the estimate, is responsible of the smoothness of the kernel estimate; smaller r_N's generate more noisy estimates, while larger r_N's generate smoother estimates.

Advantages of kernel estimates of a density functions are well know in literature [45], [31], [53]; in particular we may remind that, due to the above assumptions on k, we know that

- the kernel estimate $\widehat{f}_X^N$ is a pdf on $\mathbb{R}$;
- the kernel estimate $\widehat{f}_X^N$ inherits the smoothness of the kernel k; i.e. if k is n times continuously differentiable, $\widehat{f}_X^N$ is n times continuously differentiable, too.

The guidelines for choosing an optimal kernel estimator are thus based on the required analytical properties of the kernel k, and on an optimal choice of the bandwidth r_N, so to obtain optimal statistical properties of $\widehat{f}_X^N$, which include unbiasedness, minimal variance, and consistency.

By assuming that the underlying density f and the chosen kernel k are sufficiently smooth, (in particular, provided that $f \in C^2$) it can be shown (see e.g. [45, Chapter 3], [53, p. 20-21]) that

$$Bias(\widehat{f}_X^N(x)) = \frac{1}{2}f''(x)c_2r_N^2 + o(r_N^2), \qquad \text{for } \lim_{N\to\infty} r_N = 0$$

$$Var(\widehat{f}_X^N(x)) = \frac{1}{Nr_N}c_1f(x) + o\big(\frac{1}{Nr_N}\big), \quad \text{for } \lim_{N\to\infty} r_N = 0 \text{ and } \lim_{N\to\infty} Nr_N = \infty.$$

As a consequence, the following assertion is proved in [40].

Theorem A.1. *Let X be an absolutely continuous real random variable having p.d.f f, and let $\{X_i\}_{i\in\mathbb{N}}$ be a countable sample of X, i.e. a sequence of random variables i.i.d. as X. Then the kernel estimate $\widehat{f}_X^N(x)$ of f defined in* (A.1) *is asymptotically unbiased and weakly consistent at all points x at which f is continuous, if r_N is such that*

$$\lim_{N\to\infty} r_N = 0 \quad \textit{and} \qquad \lim_{N\to\infty} Nr_N = \infty.$$

For a given sample size $N \in \mathbb{N}\setminus\{0\}$, a compromise is required about an optimal choice of the bandwidth, since the bias tends to 0 as the bandwidth decreases, but correspondingly the variance diverges. A well known measure of both effects is the pointwise mean square error ($MSE(\widehat{f}_X^N(x))$) defined as follows.

$$MSE(\widehat{f}_X^N(x)) = \mathbb{E}[(\widehat{f}_X^N(x)) - f(x))^2] = [Bias(\widehat{f}_X^N(x))]^2 + Var(\widehat{f}_X^N(x)), \quad x \in \mathbb{R}.$$

We are now ready to obtain an optimal bandwidth for estimating $f(x)$, by minimizing $MSE(\widehat{f}_X^N(x))$. The analysis is simplified by considering the *asymptotic* approximation of the *mean square error*, i.e. the quantity

$$AMSE(\widehat{f}_X^N(x)) := \big(\frac{1}{2}f''(x)c_2r_N^2\big)^2 + \frac{1}{Nr_N}c_1f(x);$$

an *optimal bandwidth* is then given by (e.g.,[31, p. 59], [45, 40])

$$r_N^{o,AMSE}(x) := \arg\min_{r_N} AMSE(\widehat{f}_X^N(x)) = \sqrt[5]{\frac{c_1f(x)}{N(c_2f''(x))^2}}. \tag{A.2}$$

A criterion for obtaining a uniform choice of the optimal bandwidth is based on the *integrated mean square error* ($MISE(\widehat{f}_X^N)$), defined as follows (e.g., see [31, p. 60])

$$MISE(\widehat{f}_X^N) := \int_{\mathbb{R}} MSE(\widehat{f}_X^N(x))dx.$$

By considering, as above, the *asymptotic approximation* of the MISE, i.e. the quantity

$$AMISE(\widehat{f}_X^N) := \frac{1}{Nr_N}c_1 + \frac{r_N^4}{4}(c_2\|f''\|)^2,$$

it follows that a *global optimal bandwidth* is given by

$$r_N^{o,AMISE} := \underset{r_N}{\arg\min} AMISE(\widehat{f}_X^N(x)) = \sqrt[5]{\frac{c_1}{N(c_2\|f''\|)^2}}. \tag{A.3}$$

Equation (A.3) shows in particular the dependence of both optimal bandwidths upon $\|f''\|^2$ which is a measure of the roughness of the unknown pdf f. Methods for estimating $\|f''\|^2$ are discussed in [31] (see also [45, 43, 37, 10] and references therein).

A natural extension to the multivariate case $\mathbb{R}^d$, $d \in \mathbb{N} \setminus 0$ is obtained by introducing a multivariate kernel defined as follows.

Definition A.2 (Multivariate kernel). *A measurable function $k : \mathbb{R}^d \longrightarrow \mathbb{R}$ is said to be a multivariate kernel if it satisfies the following conditions,*

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

Given a countable sample $\{X_i\}_{i\in\mathbb{N}}$ of an absolutely continuous $d-$dimensional random vector X having p.d.f f, the multivariate kernel density estimator of f, based on a chosen kernel k, and scaling parameter $r_N \in (0, +\infty)$, is defined, as in the scalar case, by

$$\widehat{f}_X^N(x) := \frac{1}{Nr_N^d}\sum_{i=1}^N k\left(\frac{x - X_i}{r_N}\right), \quad x \in \mathbb{R}^d. \tag{A.4}$$

By introducing the scaled kernel

$$k_{r_N}(x) := \frac{1}{r_N^d}k\left(\frac{x}{r_N}\right),$$

we may rewrite Eq. (A.4) in the form

$$\widehat{f}_X^N(x) := \frac{1}{Nr_N^d}\sum_{i=1}^N \int_{\mathbb{R}^d} k\Big(\frac{x-y}{r_N}\Big)\varepsilon_{X_i}(dy) = \frac{1}{N}\sum_{i=1}^N k_{r_N} * \mathcal{H}^0_{|X_i}(x),$$

where $*$ stands for the usual convolution product, having noticed that the Dirac measure ε_{X_i}, $i = 1, \dots N$, can be rewritten as the restriction to X_i of the 0-dimensional Hausdorff measure $\mathcal{H}^0$.

Appendix A.2. Point processes and germ-grain representation of random closed sets

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* (see e.g., [5] and references therein). To lighten the presentation, we shall use similar notation to previous works [51, 52]; in particular, we refer to [52, Sec. 2.2] for further details.

A point process $\widetilde{\Phi}$ in $\mathbb{R}^d$ is a locally finite collection $\{\xi_i\}_{i\in\mathbb{N}}$ of random points in $\mathbb{R}^d$; equivalently, it can be seen as a random counting measure, that is a measurable map from a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ into the space of locally finite counting measures on $\mathbb{R}^d$ (e.g., see [21, 36]). We shall always consider simple point processes $\widetilde{\Phi}$, i.e. $\widetilde{\Phi}(\{x\}) \leq 1$ for all $x \in \mathbb{R}^d$. The measure $\widetilde{\Lambda}(A) := \mathbb{E}[\widetilde{\Phi}(A)]$ on $\mathcal{B}_{\mathbb{R}^d}$ is called *intensity measure* of $\widetilde{\Phi}$; whenever it is absolutely continuous with respect to $\mathcal{H}^d$, its density is called *intensity* of $\widetilde{\Phi}$.
Marked point processes generalize the above notion; we recall that a marked point process $\Phi = \{\xi_i, K_i\}_{i\in\mathbb{N}}$ on $\mathbb{R}^d$ with marks in a Polish space $\mathbf{K}$ (a complete and separable metric space) is a point process on $\mathbb{R}^d \times \mathbf{K}$ with the property that the unmarked process $\{\widetilde{\Phi}(B) : B \in \mathcal{B}_{\mathbb{R}^d}\} := \{\Phi(B \times \mathbf{K}) : B \in \mathcal{B}_{\mathbb{R}^d}\}$ is a point process in $\mathbb{R}^d$. The point process $\widetilde{\Phi}$ is called the *underlying point process*, while $\mathbf{K}$ is called the *mark space*; the random element K_i of $\mathbf{K}$ is the *mark associated to the point* $\xi_i \in \widetilde{\Phi}$. Φ is said to be *stationary* if the distribution of $\{\xi_i + x, K_i\}_{i\in\mathbb{N}}$ is independent of $x \in \mathbb{R}^d$. If the marks are independent and identically distributed, and independent of the unmarked point process $\widetilde{\Phi}$, then Φ is said to be an *independent marking* of $\widetilde{\Phi}$.
The intensity measure of Φ, say Λ, is a σ-finite measure on $\mathcal{B}_{\mathbb{R}^d\times\mathbf{K}}$ defined as $\Lambda(B \times L) := \mathbb{E}[\Phi(B \times L)]$, the mean number of points of Φ in B with marks in L. We recall that Campbell's formula for marked point processes reads as follows [5]:

$$\mathbb{E}\Big[\sum_{(x,K)\in\Phi} f(x, K)\Big] = \int_{\mathbb{R}^d\times\mathbf{K}} f(x, K)\Lambda(\mathrm{d}(x, K)). \tag{A.5}$$

A common assumption is that, given a probability measure Q on $\mathbf{K}$, called the distribution of marks, there exists a measurable function $\lambda : \mathbb{R}^d \times \mathbf{K} \to \mathbb{R}_+$ such that $\Lambda(\mathrm{d}(x, K)) = \lambda(x, K)\mathrm{d}xQ(\mathrm{d}K)$; if Φ is stationary, then its intensity measure is of the type $\Lambda = \lambda\nu^d \otimes Q$ for some $\lambda > 0$. If Φ is an independent marking of $\widetilde{\Phi}$, then $\Lambda(\mathrm{d}(x, K)) = \widetilde{\Lambda}(\mathrm{d}x)Q(\mathrm{d}K)$.
Another important measure associated to Φ is the so-called *second factorial moment measure*, $\nu_{[2]}$, defined on $\mathcal{B}_{(\mathbb{R}^d\times\mathbf{K})^2}$ as follows [47]:

$$\int f(x_1, K_1, x_2, K_2)\nu_{[2]}(\mathrm{d}(x_1, K_1, x_2K_2)) = \mathbb{E}\Big[\sum_{\substack{(x_i,K_i),(x_j,K_j)\in\Phi,\\ x_i\neq x_j}} f(x_i, K_i, x_j, K_j)\Big], \tag{A.6}$$

for any non-negative measurable function f on $(\mathbb{R}^d \times \mathbf{K})^2$. Informally, $\nu_{[2]}(\mathrm{d}(x_1, K_1, x_2, K_2))$ represents the joint probability that there are points at two specific locations x_1 and x_2 with marks K_1 and K_2, respectively. Similarly to Λ, we shall assume that there exist a measurable function $g : (\mathbb{R}^d \times \mathbf{K})^2 \to \mathbb{R}_+$, and a probability measure $Q_{[2]}$ on $\mathbf{K}^2$ such that

$$\nu_{[2]}(\mathrm{d}(x_1, K_1, x_2, K_2)) = g(x_1, K_1, x_2, K_2)\mathrm{d}x_1\mathrm{d}x_2Q_{[2]}(\mathrm{d}(K_1, K_2)).$$

We remind that if Φ is a marked Poisson point process (i.e. its underlying point process is a Poisson process on $\mathbb{R}^d$) with intensity measure Λ, then $\nu_{[2]} = \Lambda \otimes \Lambda$. (See [52, Sec. 2.2] for a more complete discussion about $\nu_{[2]}$, and additional references.)

We also recall that point processes in $\mathcal{C}^d$, the class of compact subsets of $\mathbb{R}^d$, are called *particle process* (e.g., see [5] and references therein), and it is well known that, by a *center map*, any particle process can be transformed into a marked point process Φ on $\mathbb{R}^d$ with marks in $\mathcal{C}^d$, by representing any compact set C as a pair (x, Z), where x may be interpreted as the "location" of C

and $Z := C - x$ the "shape" (or "form") of C (e.g., see [5, p. 192] and [34]). In this case the marked point process $\Phi = \{(\xi_i, Z_i)\}_{i\in\mathbb{N}}$ is also called *germ-grain model.* Thus, every random closed set in $\mathbb{R}^d$ can be represented as a germ-grain model, by a suitable marked point process $\Phi = \{\xi_i, Z_i\}_{i\in\mathbb{N}}$. In many examples and applications the random sets Z_i, $i \in \mathbb{N}$ are uniquely determined by suitable random parameters $S \in \mathbf{K}$. For instance, in the very simple case of random balls, $\mathbf{K} = \mathbb{R}_+$ and S is the radius of a ball centred in the origin; in applications to birth-and-growth processes, in some models $\mathbf{K} = \mathbb{R}^d$ and S is the spatial location of the nucleus (e.g., [2, Example 2]); in segment processes in $\mathbb{R}^2$, $\mathbf{K} = \mathbb{R}_+ \times [0, 2\pi]$ and $S = (L, \alpha)$, where L and α are the random length and orientation of the segment through the origin, respectively (e.g., [51], Example 2); etc.

Appendix A.3. Basic notions of geometric measure theory

We remind that a compact set $A \subset \mathbb{R}^d$ is called *n-rectifiable* ($0 \leq n \leq d-1$ integer) if it can be written as the image of a compact subset of $\mathbb{R}^n$ by a Lipschitz map from $\mathbb{R}^n$ to $\mathbb{R}^d$; more in general, a closed subset A of $\mathbb{R}^d$ is said to be *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. (For definitions and basic properties of Hausdorff measure and rectifiable sets see, e.g., [4, 27, 30].)

We recall that, given a subset A of $\mathbb{R}^d$ and an integer n with $0 \leq n \leq d$, the *n-dimensional Minkowski content of A* is defined as

$$\mathcal{M}^n(A) := \lim_{r\downarrow 0} \frac{\mathcal{H}^d(A_{\oplus r})}{b_{d-n} r^{d-n}},$$

whenever the limit exists finite, where $A_{\oplus r}$ is the parallel set of A at distance $r > 0$, i.e. $A_{\oplus r} := \{s \in \mathbb{R}^d : \text{dist}(x, A) \leq r\}$ (see, e.g. [4]). Well known general results about the existence of the Minkowski content of closed sets in $\mathbb{R}^d$ are related to rectifiability properties of the involved sets; in particular, the following theorem proved in [4, p. 110] provides a quite general condition ensuring the existence of the n-dimensional Minkowski content of any compact subset of $\mathbb{R}^d$. We call Radon measure in $\mathbb{R}^d$ any nonnegative and σ-additive set function defined on $\mathcal{B}_{\mathbb{R}^d}$ which is finite on bounded sets.

Theorem A.3. *Let $A \subset \mathbb{R}^d$ be a countably $\mathcal{H}^n$-rectifiable compact set, and assume that*

$$\eta(B_r(x)) \geq \gamma r^n, \qquad \forall x \in A,\ \forall r \in (0,1) \tag{A.7}$$

holds for some $\gamma > 0$ and some Radon measure $\eta \ll \mathcal{H}^n$ in $\mathbb{R}^d$. Then $\mathcal{M}^n(A) = \mathcal{H}^n(A)$.

Condition (A.7) is a kind of quantitative non-degeneracy condition which prevents A from being too sparse; simple examples show that $\mathcal{M}^n(A)$ can be infinite, and $\mathcal{H}^n(A)$ arbitrarily small, when this condition fails [4, 3]. The above theorem extends (see [4, Theorem 2.106]) the well-known Federer's result [30, p. 275] to countably $\mathcal{H}^n$-rectifiable compact sets; in particular for any n-rectifiable compact set $A \subset \mathbb{R}^d$ there exists a suitable measure η satisfying (A.7) (see [3, Remark 1]). As a consequence, for instance in the case $n = d-1$, the boundary of any convex body or, more in general, of a set with positive reach, and the boundary of a set with Lipschitz boundary satisfy condition (A.7), which in many applications is fulfilled with $\eta(\cdot) = \mathcal{H}^n(\widetilde{A} \cap \cdot)$ for some closed set $\widetilde{A} \supseteq A$ (see [4, p. 111], [3]). A regularity condition closely related to (A.7) is the so-called n-Ahlfors regularity (e.g., see [22, 28]); more precisely, we say that a compact set $A \subset \mathbb{R}^d$ is *n-Ahlfors regular* if there exist $\gamma_1, \gamma_2 > 0$ such that

$$\gamma_1 r^n \leq \mathcal{H}^n(A \cap B_r(x)) \leq \gamma_2 r^n, \qquad \forall x \in A,\ \forall r \in (0,1). \tag{A.8}$$

Finally, we recall the notion of *approximate tangent space*, which arises in the approximation of the variance of the kernel density estimator in Theorem 8.
Let $\mathbf{G}_n$ be the set of unoriented n-dimensional subspaces of $\mathbb{R}^d$, and $C_c(\mathbb{R}^d;\mathbb{R})$ be the space of all the real valued continuous functions with compact support in $\mathbb{R}^d$.

Definition A.4. *[4, Definition 2.79] Let A be a $\mathcal{H}^n$-rectifiable compact set of $\mathbb{R}^d$, and $A_{x,r} := (A-x)/r$, $x \in A$, $r > 0$. We say that $\mu = \mathcal{H}^n_{|A}$ has approximate tangent space $\pi_x \in \boldsymbol{G}_n$ with multiplicity 1 at x, and we write*

$$\mathrm{Tan}^n(\mu, x) = \mathcal{H}^n_{|\pi_x}$$

if $\mathcal{H}^n_{|A_{x,r}}$ locally weakly converge to $\mathcal{H}^n_{|\pi_x}$ as $r \to 0$, i.e.*

$$\lim_{r\to 0} \int_{A_{x,r}} \phi(y)\mathcal{H}^n(\mathrm{d}y) = \int_{\pi_x} \phi(y)\mathcal{H}^n(\mathrm{d}y) \quad \forall \phi \in C_c(\mathbb{R}^d;\mathbb{R}). \tag{A.9}$$

By Theorem 2.83 and Proposition 1.62 in [4] the following holds.

Theorem A.5. *Let A be a $\mathcal{H}^n$-rectifiable compact set of $\mathbb{R}^d$, and let $\mu = \mathcal{H}^n_{|A}$, then μ admits an approximate tangent space with multiplicity* 1 *for $\mathcal{H}^n$-a.e. $x \in A$. Moreover,* (A.9) *holds for any bounded Borel measurable function $\phi : \mathbb{R}^d \to \mathbb{R}$ with compact support such that* $\mathcal{H}^n_{|\pi_x}(\mathrm{disc}(\phi)) = 0$.

Acknowledgments. The research work of VC has been financially supported by the Italian national research project "PRIN 2009RNH97Z-002-2009", funded by MIUR, the Italian Ministery of Instruction, University and Research. The authors would also thank the anonymous referee for his/her valuable comments and suggestions, which helped to improve the readability of the paper.

References

[1] J. Aitchison, J.W. Kay, I.J. Lauder, Statistical Concepts and Applications in Clinical Medicine, Chapman and Hall/CRC, London, 2004.

[2] L. Ambrosio, V. Capasso, E. Villa, On the approximation of mean densities of random closed sets, Bernoulli 15 (2009) 1222–1242.

[3] L. Ambrosio, A. Colesanti, E. Villa, Outer Minkowski content for some classes of closed sets, Math. Ann. 342 (2008) 727–748.

[4] L. Ambrosio, N. Fusco, D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Clarendon Press, Oxford, 2000.

[5] A. Baddeley, I. Barany, R. Schneider, W. Weil, Stochastic Geometry, Lecture Notes in Mathematics 1982, Springer, Berlin, 2007.

[6] A. Baddeley, I.S. Molchanov, On the expected measure of a random set, in: Proceedings of the International Symposium on Advances in Theory and Applications of Random Sets (Fontainebleau, 1996), River Edge, NJ, World Sci. Publishing, 1997, pp. 3–20.

[7] V. Beneš, J. Rataj, Stochastic Geometry: Selected Topics, Kluwer, Dordrecht, 2004.

[8] M. Berman, P. Diggle, Estimating weighted integrals of the second-order intensity of a spatial point process, J. R. Statist. Soc. B 51 (1989) 81–92.

[9] D. Bosq, J.P. Lecoutre, Theorie De L'estimation Fonctionnelle, Economica, Paris, 1987.

[10] Z.I. Botev, J.F. Grotowski, D.P. Kroese, Kernel density estimation via diffusion, Ann. Statist. 38 (2010) 2916–2957.

[11] M. Burger, V. Capasso, L. Pizzocchero, Mesoscale averaging of nucleation and growth models, SIAM J. Multiscale Modelling and Simulation 5 (2006) 564–592.

[12] Mathematical Modelling for Polymer Processing. Polymerization, Crystallization, Manufacturing (V. Capasso, Editor). ECMI Series on Mathematics in Industry Vol 2, Springer Verlag, Heidelberg, 2003.

[13] V. Capasso, E. Dejana, A. Micheletti, Methods of stochastic geometry, and related statistical problems in the analysis and therapy of tumour growth and tumour-driven angiogenesis, in: Selected Topics on Cancer Modelling, (N. Bellomo et al. Eds.), Birkhauser, Boston, 2008, pp. 299–335.

[14] V. Capasso, A. Micheletti, Stochastic geometry and related statistical problems in biomedicine, in: Complex System in Biomedicine, (A. Quarteroni et al. Eds.), Springer, Milano, 2006, pp. 35–69.

[15] V. Capasso, A. Micheletti, D. Morale, Stochastic geometric models and related statistical issues in tumour-induced angiogenesis, Mathematical Biosciences 214 (2008) 20–31.

[16] V. Capasso, D. Morale, Stochastic modelling of tumour-induced angiogenesis, J. Math. Biol. 58 (2009) 219–233.

[17] V. Capasso, E. Villa, On the continuity and absolute continuity of random closed sets, Stoch. An. Appl. 24 (2006) 381–397.

[18] V. Capasso, E. Villa, On mean densities of inhomogeneous geometric processes arising in material sciences and medicine, Image Anal. Setreol. 26 (2007) 23–36.

[19] V. Capasso, E. Villa, On the geometric densities of random closed sets, Stoch. An. Appl. 26 (2008) 784–808.

[20] N.A.C. Cressie, Statistics for Spatial Data, Revised edition, John Wiley and Sons, New York, 1993.

[21] D.J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, Springer Series in Statistics, New York, 1988.

[22] G. David, S. Semmes, Fractured Fractals and Broken Dreams-Self-Similar Geometry through Metric and Measure, Oxford Univ. Press, London, 1997.

[23] P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators, Stat. Inference Stoch. Process. 7 (2004) 225–277.

[24] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer Series in Stochastic Modelling and Applied Probability, New York, 1996.

[25] P.J. Diggle, Statistical analysis of spatial point patterns, Mathematics in Biology, Academic Press, London, 1983.

[26] P.J. Diggle, A kernel method for smoothing point process data, Appl. Statist. 34 (1985) 138–147.

[27] K.J. Falconer, The Geometry of Fractal Sets, Cambridge University Press, Cambridge, 1986.

[28] K.J. Falconer, One-sided multifractal analysis and points of non-differentiability of devil's staircases, Math. Proc. Cambridge Philos. Soc. 136 (2004) 167–174.

[29] H. Federer, Curvature measures, Trans. Amer. Math. Soc. 93 (1959) 418–491.

[30] H. Federer, Geometric Measure Theory, Spriger, Berlin, 1969.

[31] W. Härdle, Smoothing Techniques with Implementation in **S**, Springer-Verlag, New York, 1991.

[32] R. Helmers, I.W. Mangku, R. Zitikis, Consistent estimation of the intensity function of a cyclic Poisson process, J. Multivariate Anal. 84 (2003) 19–39.

[33] R. Helmers, R. Zitikis, On estimation of Poisson intensity functions, Ann. Inst. Statist. Math. 51 (1999) 265–280.

[34] D. Hug, G. Last, W. Weil, A survey on contact distributions, in: Statistical Physics and Spatial Statistics, Lecture Notes in Physics 600, Morphology of Condensed Matter, Physics and Geometry of Spatially Complex Systems, (K. Mecke and D. Stoyan Eds.), Springer, Berlin, 2002, pp. 317–357.

[35] D. Hug, G. Last, W. Weil, A local Steiner-type formula for general closed sets and applications, Math. Z. 246 (2004) 237–272.

[36] A.F. Karr, Point Processes and Their Statistical Inference, Marcel Dekker, New York, 1986.

[37] C.R. Loader, Bandwidth selection: Classical or plug-in, Ann. Statist. 27 (1999) 415–438.

[38] G. Matheron, Random Sets and Integral Geometry, John Wiley & Sons, New York, 1975.

[39] I. Molchanov, Statistics of the Boolean Model for Practitioners and Mathematicians, Chichester, Wiley, 1997.

[40] E. Parzen, On the estimation of a probability density function and the mode, Ann. Math. Statist. 33 (1962) 1065–1076.

[41] P.M.V Rancoita, A. Giusti, A. Micheletti, Intensity estimation of stationary fibre processes from digital images with a learned detector, Image Anal. Setreol. 30 (2011) 167–178.

[42] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Statist. 27 (1956) 832–837.

[43] S.J. Sheather, Density estimation, Statistical Science 19 (2004) 588–597.

[44] W.R. Schucany, Locally optimal window widths for kernel density estimation with large samples, Statist. Probab. Lett. 7 (1989) 401–405.

[45] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.

[46] J.S. Simonoff, Smoothing Methods in Statistics, Springer, New York, 1996.

[47] D. Stoyan, W.S. Kendall, J. Mecke, Stochastic Geometry and its Applications, 2nd edition, John Wiley & Sons, Chichcester, 1995.

[48] R. Szeliski, Computer Vision. Algorithms and Applications, Springer-Verlag, London, 2011.

[49] P.V. van Kerm, Adaptive kernel density estimation, Stata Journal 3 (2003) 148–156.

[50] M.N.M. van Lieshout, On estimation of the intensity function of a point process, Methodology and Computing in Applied Probability 14 (2012) 567–578.

[51] E. Villa, Mean densities and spherical contact distribution function of inhomogeneous Boolean models, Stoch. An. Appl. 28 (2010) 480–504.

[52] E. Villa, On the local approximation of mean densities of random closed sets, Bernoulli, in press (*Available at* http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers).

[53] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman & Hall, London, 1995.

[54] W. Wertz, Statistical Density Estimation: a Survey, Angewandte Statistik und Ökonometrie [Applied Statistics and Econometrics], 13, Vandenhoeck & Ruprecht, Göttingen, 1978.

[55] M. Zähle, Random processes of Hausdorff rectifiable closed sets, Math. Nachr. 108 (1982) 49–72.