Journal of Multivariate Analysis 125 (2014) 65–88

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On the estimation of the mean density of random closed sets

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

[a] Department of Mathematics, Università degli Studi di Pavia, via Ferrata 1, 27100 Pavia, Italy

[b] Department of Mathematics, Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy

ARTICLE INFO

Article history:
Received 23 January 2013
Available online 18 December 2013

AMS 2010 subject classifications:
primary 62G07
secondary 60D05
28A75

Keywords:
Density estimator
Kernel estimate
Stochastic geometry
Random closed set
Hausdorff dimension
Minkowski content

ABSTRACT

Many real phenomena may be modeled 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 the literature follow as particular cases. A series of examples throughout the paper, both non-stationary, and stationary, are provided to illustrate various relevant situations.

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 fiber 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 [1,13–15], material science [12], etc.

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

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

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

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.12.003

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

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

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

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

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 $\Theta_n^1, \ldots, \Theta_n^N$ 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 [51,52], 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 [23,54] 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 [2,17–19,52], 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, dx 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 center 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, 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:

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

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 $\Lambda(\mathrm{d}(x,s)) = \lambda(x,s)\mathrm{d}x Q(\mathrm{d}s)$ and second factorial moment measure $\nu_{[2]}(\mathrm{d}(x,s,y,t)) = g(x,s,y,t)\mathrm{d}x\mathrm{d}y Q_{[2]}(\mathrm{d}(s,t))$, 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:*