

# Boundary density and Voronoi set estimation for irregular sets

Raphaël Lachièze-Rey, Sergio Vega

► **To cite this version:**

Raphaël Lachièze-Rey, Sergio Vega. Boundary density and Voronoi set estimation for irregular sets. MAP5 2015-03. to appear in Trans. AMS. 2015. <hal-01105205v3>

**HAL Id: hal-01105205**

**<https://hal.archives-ouvertes.fr/hal-01105205v3>**

Submitted on 23 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Boundary density and Voronoi set estimation for irregular sets

Raphaël Lachieze-Rey\* and Sergio Vega †

November 23, 2015

## Abstract

In this paper, we study the inner and outer boundary densities of some sets with self-similar boundary having Minkowski dimension  $s > d - 1$  in  $\mathbb{R}^d$ . These quantities turn out to be crucial in some problems of set estimation, as we show here for the Voronoi approximation of the set with a random input constituted by  $n$  iid points in some larger bounded domain. We prove that some classes of such sets have positive inner and outer boundary density, and therefore satisfy Berry-Esseen bounds in  $n^{-s/2d}$  for Kolmogorov distance. The Von Koch flake serves as an example, and a set with Cantor boundary as a counter-example. We also give the almost sure rate of convergence of Hausdorff distance between the set and its approximation.

**Keywords** Voronoi approximation; Set estimation; Minkowski dimension; Berry-Esseen bounds; self-similar sets

**MSC 2010 Classification** Primary 60D05, 60F05, 28A80, Secondary 28A78, 49Q15

---

\*raphael.lachieze-rey@parisdescartes.fr, Laboratoire MAP5 (UMR CNRS 8145), Université Paris Descartes, Sorbonne Paris Cité

†Laboratoire MAP5 (UMR CNRS 8145), Université Paris Descartes, Sorbonne Paris Cité

**Notations** We designate by  $d(.,.)$  the Euclidean distance between points or subsets of  $\mathbb{R}^d$ . The closure, the interior, the topological boundary and the diameter of a set  $A \subset \mathbb{R}^d$  are designated by  $\text{cl}(A)$ ,  $\text{int}(A)$ ,  $\partial A$ ,  $\text{diam}(A)$  respectively. The open Euclidean ball with center  $x$  and radius  $r$  in  $\mathbb{R}^d$  is noted  $B(x, r)$ .

Given two sets  $A, B$ , we write  $A+B$  for  $\{c \in \mathbb{R}^d \mid c = a+b, a \in A, b \in B\}$ . The Hausdorff distance between  $A$  and  $B$  is designated by  $d_H(A, B)$ , that is

$$d_H(A, B) = \inf\{r > 0 : A \subset B + B(0, r), B \subset A + B(0, r)\}.$$

$\text{Vol}$  is the  $d$ -dimensional Lebesgue measure and  $\kappa_d$  is the volume of the Euclidean unit ball. For  $s > 0$ ,  $\mathcal{H}^s$  is the  $s$ -dimensional Hausdorff measure on  $\mathbb{R}^d$ .

Throughout the paper,  $K \subset \mathbb{R}^d$  is a non-empty compact set with positive volume. The letters  $c, C$  are reserved to indicate positive constants that depend only on fixed parameters like  $K$  or  $d$ , and which value may change from line to line.

## Background

Set estimation theory is a topic of nonparametric statistics where an unknown set  $K$  is estimated, based on partial random information. The random input generally consists in a finite sample  $\chi$  of points, either IID variables [8, 19] or a Poisson point process [13, 14, 20]. Based on the information of which of those points belong or not to  $K$ , one can reconstruct a random approximation  $K_\chi$  of  $K$  and study the asymptotic quality of the approximation. See the recent survey [15, Chap. 11] about related works in nonparametric statistics.

The results generally require the set to be smooth in some sense. In the literature, the set under study is assumed to be convex [20, 23],  $r$ -convex [9, 21], to have volume polynomial expansion [2], positive reach, or a  $(d-1)$ -rectifiable boundary [14]. Another class of regularity assumptions usually needed is that of *sliding ball* or *rolling ball* conditions ([7, 24, 25]). The most common form of this condition is that in every point  $x$  of the boundary, there must be a ball touching  $x$  and contained either in  $K$ , in  $K^c$ , or both.

In those works, the random approximation model  $K_\chi$  can be the union of balls centred in the points of  $\chi$  with well tuned radius going to 0, a level set of the sum of appropriately scaled kernels centred on the random points, or else. Recently, a different model has been used in stochastic geometry,

based on the Voronoi tessellation associated with  $\chi$ . One defines  $K_\chi$  as the union of all Voronoi cells which centers lie in  $K$ , assuming that points of  $\chi$  fall indifferently inside and outside  $K$ , as  $K$  is unknown. This is equivalent to defining  $K_\chi$  as the set of points that are closer to  $\chi \cap K$  than to  $\chi \cap K^c$ .

This elegant model presents practical advantages in set estimation. For volume estimation the bias and standard deviation rates of the Voronoi approximation seem to be best among all estimators of which the authors are aware of, and hold under almost no assumption on  $K$ . Regarding shape estimation, Voronoi approximation also consistently estimates  $K$  and  $\partial K$  in the sense of the Hausdorff distance (Proposition 3), and here again convergence rates and necessary assumptions compare favourably to those of other estimators (see Theorem 4 and the following Remarks).

An heuristic explanation of these features is that the estimator naturally fills in regions inside  $K$  where the sample  $\chi$  is sparse, without need for convexity-like assumptions on  $K$  [21] or parameter tuning [3, 8, 10].

The reader will find a more formal presentation of Voronoi approximation along with a summary of existing results [5, 13, 14, 20, 23] in Section 2.

## Approach and main results

This work was inspired and is closely related to [17], in which a central limit theorem and variance asymptotics for  $\text{Vol}(K_\chi)$  were obtained for binomial input under very weak assumptions on  $K$ . Here we slightly enhance their central limit theorem by showing that  $\text{Vol}(K_\chi)$  can be recentered by  $\text{Vol}(K)$  instead of  $\mathbf{E}(\text{Vol}(K_\chi))$ . Explicitly, for suitable  $K$ , we have for each  $\varepsilon > 0$  a constant  $C_{\varepsilon>0}$  such that

$$\sup_{t \in \mathbb{R}} \left| \mathbf{P} \left( \frac{\text{Vol}(K_{\chi_n}) - \text{Vol}(K)}{\sqrt{\text{Var}(\text{Vol}(K_{\chi_n}))}} \geq t \right) - \mathbf{P}(N \geq t) \right| \leq C_\varepsilon n^{-s/2d} \log(n)^{4-s/d+\varepsilon} \quad (1)$$

where  $s$  is the Minkowski dimension of  $\partial K$  (see Section 1.2).

We also show that with Poisson input we have the almost sure convergence rates for the Hausdorff distance

$$c \leq \liminf_{n \rightarrow +\infty} \frac{d_H(K, K_{\chi_n})}{(n^{-1} \ln(n))^{1/d}} \leq \limsup_{n \rightarrow +\infty} \frac{d_H(K, K_{\chi_n})}{(n^{-1} \ln(n))^{1/d}} \leq C, \quad (2)$$

$$c \leq \liminf_{n \rightarrow +\infty} \frac{d_H(\partial K, \partial K_{\chi_n})}{(n^{-1} \ln(n))^{1/d}} \leq \limsup_{n \rightarrow +\infty} \frac{d_H(\partial K, \partial K_{\chi_n})}{(n^{-1} \ln(n))^{1/d}} \leq C, \quad (3)$$

thus answering a query raised in [13] and extending the results obtained in [5].

The assumptions on  $K$  necessary for (1),(2) and (3) to hold are worth of interest on their own. They are broad enough to allow for irregular  $K$ , a feature which few estimators possess and is useful in some applications (see [8, 14], and references therein). Also, they are not specific to Voronoi approximation, and might be crucial for other estimators. They are mainly concerned with the densities of  $K$  at radius  $r$  in  $x$ , defined by

$$f_r^K(x) = \frac{\text{Vol}(K \cap B(x, r))}{\text{Vol}(B(x, r))},$$

$$f_r^{K^c}(x) = \frac{\text{Vol}(K^c \cap B(x, r))}{\text{Vol}(B(x, r))}.$$

For ease of notation, we shall simply write  $f_r$  for  $f_r^K$  and  $g_r$  for  $f_r^{K^c}$ ,  $K$  being implicit in all of the paper. Boundary densities have already appeared in set estimation theory [6, 5, 8], where a set  $K$  is said to be *standard* whenever  $f_r \geq \varepsilon$  on  $K$  for some fixed  $\varepsilon > 0$  and all small enough  $r$ . Here we shall prefer to specify *inner standard* since we are also interested in cases where the inequality  $g_r > \varepsilon$  holds. In the latter case,  $K$  is said to be *outer standard*, and if  $K$  is both inner and outer standard  $K$  will be said to be *bi-standard*. The condition on  $K$  for (3) to hold is essentially bi-standardness, which is a usual assumption in set estimation [8, Theorem 1].

The requirement for (1) to hold seems to be new in set estimation theory. It consists in a positive  $\liminf$  as  $r \rightarrow 0$  of the quantity

$$\frac{1}{\text{Vol}(\partial K_r)} \int_{K^c} f_r = \frac{1}{\text{Vol}(\partial K_r)} \int_K g_r,$$

where  $\partial K_r$  is the set of points within distance  $r$  from  $\partial K$ . See Assumption 2 and Proposition 2 for precise statements and equivalent assertions. The above quantity measures the interpenetration of  $K$  and  $K^c$  along their common boundary, since the greater it is, the more homogenously  $K$  and  $K^c$  are distributed along  $\partial K$ . This lead us to name the condition of Assumption 2 the *boundary permeability condition*.

Study of densities on the boundary is also related with works in geometric measure theory. Points for which  $\lim f_r(x)$  is 0 or 1 are considered resp. as

the measure-theoretic exterior and interior of  $K$ , while other points constitute  $\partial^*K$  the *essential boundary* of  $K$ . Federer [1, Th.3.61] proved that if  $K$  is a measurable set with finite measure-theoretic perimeter then  $f_r \rightarrow 1/2$  on most of the essential boundary.

We address here the question of whether comparable results hold if  $\partial K$  is an irregular set, with self-similar features. In general, such boundaries have a Hausdorff dimension  $s > d - 1$  and don't have finite perimeter. But, because of self-similarity, the densities  $f_r, g_r$  should nevertheless have continuous and somehow periodical fluctuations in  $r$ , and therefore a positive infimum. This is confirmed by Theorem 1, which gives, for  $K$  with self-similar boundary, a set of conditions under which  $f_r > \varepsilon$  on the boundary uniformly in  $r > 0$ . It is even proved that a ball with radius  $cr$  for some  $c > 0$  can be rolled inside or outside the boundary, staying within a distance  $r$  from the boundary, but not touching it (otherwise self-similar boundaries would be excluded). Theorem 1 applies for instance to the Von Koch flake in dimension 2, which is therefore well-behaved under Voronoi approximation and satisfies (1), (2) and (3).

Some sets with self-similar boundary do not fall under the scope of this result, and we also give example of a self-similar set  $K_{\text{cantor}}$  with Cantor-like self-similar boundary not satisfying the boundary permeability condition. Simulations we ran suggest that this irregularity of  $K_{\text{cantor}}$ 's boundary indeed reflects on the behaviour of its Voronoi approximation and prevents the variance of the estimator from satisfying an asymptotic power law like in (10). This suggests that the boundary permeability condition is indeed significant in set estimation and not merely a contingent constraint due to the methods used to obtain (1).

## Plan

The plan of the paper is as follows. In Section 1, we recall basic facts and definitions about self-similar sets, especially regarding upper and lower Minkowski contents. We then give conditions under which sets with self-similar boundaries are standard. Voronoi approximation is formally introduced in Section 2. We then derive the volume normal approximation for sets with well-behaved boundaries, as well as Hausdorff distance results. We also develop the counter example  $K_{\text{cantor}}$  that satisfies neither the hypotheses of Theorem 1 nor the volume approximation variance asymptotics (10).

# 1 Self-similar sets

## 1.1 Self-similar set theory

This subsection contains a review of some classic results of self-similar set theory. A more precise treatment of the subject and most of the results stated here can be found in [11]. Broadly speaking, a set is self-similar when arbitrarily small copies of the set can be found in the neighbourhood of any of its points. This suggests that a self-similar set should be associated with a family of similitudes.

Let  $\{\phi_i, i \in I\}$  be a finite set of contracting similitudes. Such a set is called an *iterated function system*. Define the following set transformation

$$\begin{aligned}\psi : \mathcal{P}(\mathbb{R}^d) &\longrightarrow \mathcal{P}(\mathbb{R}^d) \\ E &\longmapsto \bigcup_i \phi_i(E).\end{aligned}$$

It is easily seen that  $\psi$  is contracting for the Hausdorff metric, which happens to be complete on  $\mathcal{K}^d$ , the class of non-empty compact sets of  $\mathbb{R}^d$ . By a fixed point theorem, there is an unique set  $E \in \mathcal{K}^d$  satisfying  $\psi(E) = E$ , which is by definition the self-similar set associated with the  $\phi_i$ .

If there is a bounded open set  $U$  such as  $\psi(U) = \bigcup \phi_i(U) \subset U$  with the union disjoint, then necessarily  $E \subset \text{cl}(U)$  and the  $\phi_i$  are said to satisfy the *open set condition*. Schief proved in [22] that we can pick  $U$  so that  $U \cap E$  is not empty. This stronger assumption is referred to as the *strong open set condition* in the literature.

The similarity dimension of  $E$  is the unique  $s$  satisfying

$$\sum \lambda_i^s = 1$$

where  $\lambda_i$  is the stretching factor of  $\phi_i$ . When the open set condition holds, this similarity dimension is also the Hausdorff dimension and the Minkowski dimension of  $E$ . Furthermore,  $E$ 's upper and lower  $s$ -dimensional Minkowski contents (see Subsection 1.2) are finite and positive. This is an easy and probably known result, but since we have not found it explicitly stated and separately proven in the literature, we will do so here in Proposition 1. We will need the following classical lemmata, that we prove for completeness.

**Lemma 1.** *Let  $(U_i)$  be a collection of disjoint open sets in  $\mathbb{R}^d$  such that each  $U_i$  contains a ball of radius  $c_1 r$  and is contained in a ball of radius  $c_2 r$ . Then any ball of radius  $r$  intersects at most  $(1 + 2c_2)^d c_1^{-d}$  of the sets  $\text{cl}(U_i)$ .*

*Proof.* Let  $B$  be a ball of center  $x$  and radius  $r$ . If some  $\text{cl}(U_i)$  intersects  $B$  then  $\text{cl}(U_i)$  is contained in the ball  $B'$  of center  $x$  and radius  $r(1 + 2c_2)$ . If  $q$  different  $\text{cl}(U_i)$  intersect  $B$  then there are  $q$  disjoint balls of radius  $c_1 r$  inside  $B'$ , and by comparing volumes  $q \leq (1 + 2c_2)^d c_1^{-d}$ .  $\square$

**Lemma 2.** *Suppose that  $E$  and the  $\phi_i$  satisfy the open set condition with  $U$ . Then for every  $r < 1$  we can find a finite set  $\mathcal{A}$  of similarities  $\Phi_k$  with ratios  $\Lambda_k$  such that*

1. *The  $\Phi_k$  are composites of the  $\phi_i$ .*
2. *The  $\Phi_k(E)$  cover  $E$ .*
3. *The  $\Phi_k(U)$  are disjoint.*
4.  *$\sum \Lambda_k^s = 1$  where  $s$  is the similarity dimension of  $E$ .*
5.  *$\min_i(\lambda_i)r \leq \Lambda_k < r$  for all  $k$ .*

*Proof.* We give an algorithmic proof. Initialise at step 0 with  $\mathcal{A} = \{Id\}$ . At step  $n$  replace every  $\Phi \in \mathcal{A}$  with ratio greater than  $r$  by the similarities  $\Phi \circ \phi_i, i \in I$ . Stop when the process becomes stationary, which will happen no later than step  $\lceil \ln(r) / \ln(\max(\lambda_i)) \rceil$ .

Obviously, point 1 is satisfied. We will prove the next three points by induction. At step 0, all of  $E$  is covered by the  $\Phi_k(E)$ , the  $\Phi_k(U)$  are disjoint, and the  $\Lambda_k^s$  sum up to 1. The first property is preserved when  $\Phi$  is replaced by the  $\Phi \circ \phi_i$ , since  $\Phi(E) = \Phi(\psi(E)) = \bigcup \Phi \circ \phi_i(E)$ . Likewise, the  $\Phi \circ \phi_i(U)$  are disjoint one from each other because  $\Phi$  is one-to-one, and disjoint from the other  $\Phi_k(U)$  because  $\bigcup \Phi \circ \phi_i(U) = \Phi(\psi(U)) \subset \Phi(U)$ , which yields point 3. For point 4 note that if  $\Phi$  has ratio  $\Lambda$ , then the  $\Phi \circ \phi_i$  have ratios  $\Lambda \lambda_i$  and  $\Lambda^s = \Lambda^s \sum \lambda_i^s = \sum (\Lambda \lambda_i)^s$  so the sum of the  $\Lambda_k^s$  remains unchanged by the substitution. Finally, since  $r < 1$ , every final set of the process has an ancestor with ratio greater than  $r$ . This gives the lower bound for point 5; the upper bound comes from the fact that the process ends.  $\square$



**Remark 1.** The process in the proof of Lemma 2 is often resumed by the formula

$$\mathcal{A} = \{\phi_{i_1} \circ \phi_{i_2} \dots \circ \phi_{i_n} \mid \prod_{k=1}^n \lambda_{i_k} < r \leq \prod_{k=1}^{n-1} \lambda_{i_k}\}.$$

## 1.2 Minkowski contents of self-similar sets

Recall that the  $s$ -dimensional lower Minkowski content of a non-empty bounded set  $E \subset \mathbb{R}^d$  can be defined as

$$\liminf_{r>0} \frac{\text{Vol}(E + B(0, r))}{r^{d-s}}.$$

Similarly, the  $s$ -dimensional upper Minkowski content of  $E$  is

$$\limsup_{r>0} \frac{\text{Vol}(E + B(0, r))}{r^{d-s}}.$$

In this paper, when both contents are finite and positive, we will simply say that  $E$  has upper and lower Minkowski contents. This leaves no ambiguity on the choice of  $s$ , since in that case  $s$  is necessarily the Minkowski dimension of  $E$ , i.e

$$s = d - \lim_{r \rightarrow 0} \frac{\ln(\text{Vol}(E + B(0, r)))}{\ln(r)}.$$

We show below that self-similar sets always have upper and lower Minkowski contents. One can find an alternative proof for the lower content in [12, Paragraph 2.4], it can also be considered a consequence of  $\mathcal{H}^s(E) > 0$ , like suggested in [18].

**Proposition 1.** *Let  $E$  be a self-similar set satisfying the open set condition with similarity dimension  $s$ . Then  $E$  has finite positive  $s$ -dimensional upper and lower Minkowski contents.*

*Proof.* As before, let  $\phi_i$  be the generating similarities of  $E$ ,  $\lambda_i$  their ratios,  $\psi : A \mapsto \bigcup \phi_i(A)$  the associated set transformation, and  $U$  the open set of the open set condition. Choose any  $0 < r < 1$  and define the  $\Phi_k, \Lambda_k$  as in Lemma 2. Finally, write  $E_k = \Phi_k(E), U_k = \Phi_k(U)$ .

We approximate  $E + B(0, r)$  by the sets  $E_k + B(0, r)$ , who are similar to the  $E + \Phi_k^{-1}(B(0, r))$ . By construction  $\Phi_k^{-1}(B(0, r))$  is a ball with a radius belonging to  $[1, (\min_i \lambda_i)^{-1}]$ , so that

$$\text{Vol}(B(0, 1)) \leq \text{Vol}(E + \Phi_k^{-1}(B(0, r))) \leq \text{Vol}(B(0, \text{diam}(E) + (\min_i \lambda_i)^{-1})),$$

because  $E$  is not empty. Applying  $\Phi_k$  gives

$$c\Lambda_k^d \leq \text{Vol}(E_k + B(0, r)) \leq C\Lambda_k^d.$$

Since  $E + B(0, r) \subset \bigcup_k E_k + B(0, r)$  and  $\sum \Lambda_k^s = 1$  we immediately get the upper bound

$$\begin{aligned} \text{Vol}(E + B(0, r)) &\leq \sum \text{Vol}(E_k + B(0, r)) \\ &\leq \sum C\Lambda_k^d \\ &\leq C \sum \Lambda_k^s r^{d-s} \\ &\leq Cr^{d-s}. \end{aligned}$$

For the lower bound we apply Lemma 1 to the disjoint  $U_k$ . Since  $U$  is open we can put some ball of radius  $c_1$  in  $U$ , and conversely we can put  $U$  in some ball of radius  $c_2$ , since  $U$  is bounded. This means that each of the  $U_k$  contains a ball of radius  $r \min_i(\lambda_i)c_1 \leq \Lambda_k c_1$  and is contained in a ball of radius  $rc_2 \geq \Lambda_k c_2$ . So for any  $x \in E + B(0, r)$ ,  $B(x, r)$  intersects at most  $q$  of the  $E_k$  (since  $E_k \subset \text{cl}(U_k)$ ) with  $q$  a positive integer independent of  $r$  and  $x$ . This can be rewritten  $\mathbf{1}_{E+B(0,r)} \geq \frac{1}{q} \sum \mathbf{1}_{E_k+B(0,r)}$ . Integrating we get  $\text{Vol}(E + B(0, r)) \geq \frac{1}{q} \sum \text{Vol}(E_k + B(0, r))$  so that

$$\begin{aligned} \text{Vol}(E + B(0, r)) &\geq \frac{1}{q} \sum \text{Vol}(E_k + B(0, r)) \\ &\geq \frac{c}{q} \sum \Lambda_k^d \\ &\geq \frac{c}{q} (\min_i \lambda_i)^{d-s} \sum \Lambda_k^s r^{d-s} \\ &\geq cr^{d-s}. \end{aligned}$$

□

### 1.3 Boundary regularity

In order to formulate our result, we introduce the notion of *proper* and *improper* points. A point  $x \in \mathbb{R}^d$  is proper to  $K$  if  $\text{Vol}(O \cap K) > 0$  for every neighbourhood  $O$  of  $x$ , it is improper to  $K$  otherwise. In other words, the set  $K^{\text{prop}}$  of proper points of  $K$  is the support of the measure  $\text{Vol}(K \cap \cdot)$ . Further use of proper points will be made in Section 2.2. We can already note that  $K$  must have no improper points if we want a positive lower bound for the  $f_r$  on  $K$ .

Our result holds for self-similar subsets  $E$  of  $\partial K$  satisfying the following assumption:

**Assumption 1.** *A self-similar subset  $E$  of  $\partial K$  satisfies the strong open set condition with some set  $U$  such that  $U \cap \partial K \subset E$  and  $U \setminus \partial K$  has finitely many connected components.*

This assumption can be justified heuristically: if  $E$  cuts its neighbourhood into infinitely many connected components, then because of self-similarity it also does so locally, and  $K$  and  $K^c$  are too disconnected to contain the balls mentioned in Theorem 1. Example 2 will show that these concerns are legitimate.

**Theorem 1.** *Let  $K$  be a non-empty compact set with no improper points and  $\text{Vol}(\partial K) = 0$ . Let  $E$  be a self-similar subset of  $\partial K$  for which Assumption 1 holds. Then there are constants  $\delta, \varepsilon > 0$  such that, for all  $r < \delta, x \in E$ , both  $B(x, r) \cap K^c$  and  $B(x, r) \cap K$  contain a ball of radius  $\varepsilon r$ .*

*Proof.* Let  $\phi_i$  be the generating similarities of  $E$ ,  $\lambda_i$  their ratios,  $\psi : A \mapsto \bigcup \phi_i(A)$  the associated set transformation. Denote by  $V_j$  the connected components of  $U \setminus E$ . Since there are finitely many of them, we can suppose they all contain a ball of radius  $\tau > 0$ . Fix any  $0 < r < 1$  and  $x \in E$ .

Lemma 2 shows that there is a similarity  $\Phi$  with ratio  $\Lambda$  such that  $\min(\lambda_i)r \leq \Lambda < r$  and  $x \in \Phi(E)$ . It follows that  $\Phi(U) \subset B(x, r)$ . We also have  $\Phi(U) \cap \partial K = \Phi(U) \cap E = \Phi(U \cap E)$ . Indeed, for any point  $x'$  of  $E$  outside  $\Phi(E)$  there is another similarity  $\Phi'$  of Lemma 2 such that  $x' \in \Phi'(E)$  and  $\Phi'(U) \cap \Phi(U) = \emptyset$ , which implies  $\text{cl}(\Phi'(U)) \cap \Phi(U) = \Phi'(\text{cl}(U)) \cap \Phi(U) = \Phi'(E) \cap \Phi(U) = \emptyset$  so that  $x' \notin \Phi(U)$ .

Consequently, for all  $j$ ,  $\Phi(V_j)$  has no intersection with  $\partial K$ . So  $\Phi(V_j) \cap \text{int}(K)$  and  $\Phi(V_j) \cap K^c$  are two disjoint open set sets who cover  $\Phi(V_j)$ , and we must have either  $\Phi(V_j) \subset K$  or  $\Phi(V_j) \subset K^c$ .

Since there is a point  $y$  in  $\Phi(U) \cap E$  and  $K$  has no improper points, we must have  $\text{Vol}(K \cap U), \text{Vol}(K^c \cap U) > 0$ . Because  $\text{Vol}(E) = 0$ , this can only happen if one of the  $\Phi(V_j)$  is included in  $K^c$  and another in  $K$ . Hence  $B(x, r) \cap K, B(x, r) \cap K^c$  each contain a ball of radius  $\Lambda\tau$ . Since  $\Lambda \geq \min(\lambda_i)r$ , the conclusion of the theorem holds with  $\varepsilon = \min(\lambda_i)\tau$ .  $\square$

**Remark 2.** The theorem implies that  $K, K^c$  have lower density bounds on  $E$ . More precisely, for appropriate  $\delta, \varepsilon > 0$

$$\forall x \in E, r < \delta, \quad f_r(x), g_r(x) \geq \varepsilon. \quad (4)$$

This weaker statement is enough for our purposes regarding Voronoi approximation.

We show below that the Von Koch flake provides a concrete example of an irregular set satisfying the hypotheses of Theorem 1.

**Example 1.** Let  $E$  be the self-similar set associated with the direct similarities  $\phi_i : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  sending  $S = A_0A_4$  to  $a_i = A_{i-1}A_i$ , for  $i = 1, 2, 3, 4$ , in the configuration of Figure 1.1. Such sets  $E$  are called Von Koch curves. Looking at the iterates  $\psi^{(n)}(S)$  in Figure 1.2 gives an idea of the general form of the Von Koch curve and of why it is said to be self-similar.

Note that the  $\psi^{(n)}(S)$  are curves, i.e the images of continuous mappings  $\gamma_n : [0, 1] \rightarrow \mathbb{R}^2$ . The  $\gamma_n$  can be chosen to be a Cauchy sequence for the uniform distance between curves in  $\mathbb{R}^2$ . Hence their limit  $\gamma$  is also a continuous mapping,  $\gamma([0, 1])$  is compact and has distance 0 with  $E$  in the Hausdorff metric, so  $\gamma([0, 1]) = E$  which proves that the Van Koch curve is, indeed, a curve. It can also be shown to be a non-intersecting curve (the image of an injective continuous mapping from  $[0, 1]$  into  $\mathbb{R}^2$ ).

With a similar reasoning, if we stick three Von Koch curves of same size as in Figure 1.3, we get a closed non-intersecting curve  $\mathcal{C}$ . Jordan's curve theorem says  $\mathbb{R}^2 \setminus \mathcal{C}$  has exactly two connected components who both have

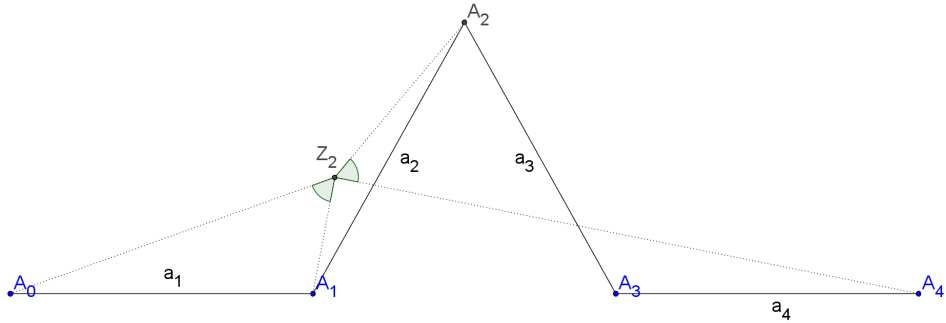


Figure 1.1: The generating similitudes of the Von Koch curve.  $Z_2$  is the center of the similarity  $\phi_2$ .

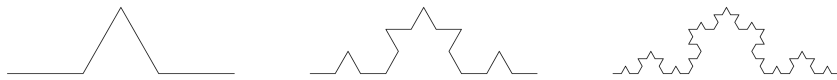


Figure 1.2: The sets  $\psi^{(1)}(S)$ ,  $\psi^{(2)}(S)$ ,  $\psi^{(3)}(S)$ .

$\mathcal{C}$  as topological boundary. The closure  $K$  of the bounded component is a compact set with no improper points satisfying  $\partial K = \mathcal{C}$ .  $K$  is called a Von Koch flake.

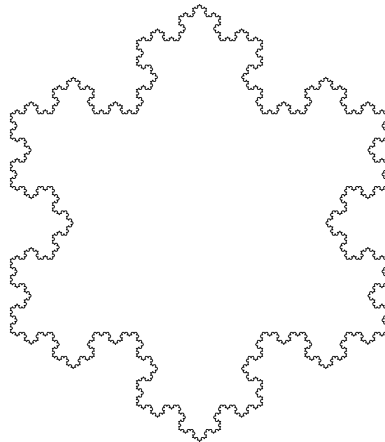


Figure 1.3: The boundary of the Von Koch flake  $K$ .

Now, construct kites  $C_1, C_2, C_3$  on each of the Von Koch curves  $E_1, E_2, E_3$  making  $\partial K$  as in Figure 1.4. It is easy to see that as long as the two equal angles of the lower triangle are flat enough,  $C_i \cap \partial K = C_i \cap E_i$ . Furthermore, applying Jordan's curve theorem to the  $E_i$  and the two upper (resp. lower) edges of the corresponding  $C_i$  shows that the  $C_i \setminus E_i$  have exactly two connected components. The strong open set condition is also satisfied, so Theorem 1 can be applied three times to obtain lower bounds for  $f_r$  and  $g_r$  on  $\partial K$ .

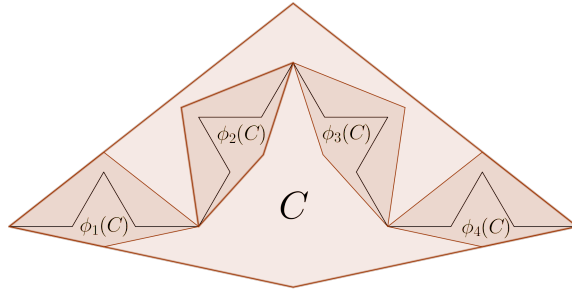


Figure 1.4: Assumption 1 is satisfied with the kite  $C$ .

Theorem 1 is only concerned with the behaviour of  $f_r$  and  $g_r$  on  $\partial K$ , whereas standardness assumption require lower bounds on all of  $K$  and  $K^c$  respectively. The following lemma takes care of this issue.

**Lemma 3.** *If for all  $r < \delta$  we have  $f_r \geq \varepsilon$  on  $\partial K$ , then for all  $r < \delta$  we have  $f_r \geq 2^{-d}\varepsilon$  on  $K$ . The same result holds if  $f_r$  is replaced by  $g_r$  and  $K$  by  $K^c$ .*

*Proof.* If  $x$  is in  $\partial K_{r/2}$  then  $B(x, r)$  contains a ball of radius  $r/2$  centered on  $x' \in \partial K$ , so  $f_r(x) = \text{Vol}(K \cap B(x, r))\kappa_d^{-1}r^{-d} \geq \text{Vol}(K \cap B(x', r/2))\kappa_d^{-1}r^{-d} \geq \varepsilon 2^{-d}$ . If  $x$  is in  $K \setminus \partial K_{r/2}$  then the ball  $B(x, r/2)$  is contained in  $K$  so that  $f_r(x) \geq 2^{-d} \geq \varepsilon 2^{-d}$ . So in all cases, if  $x \in K$  then  $f_r(x) \geq \varepsilon 2^{-d}$ . Replacing  $K$  by  $K^c$  gives the result regarding  $g_r$ .  $\square$

## 2 Voronoi approximation

In this section,  $\chi$  is a locally finite point process, and  $n \geq 1$ . If  $\chi = \chi_n = \{X_1, X_2, \dots, X_n\}$ , where the  $X_i$  are iid random points uniformly distributed

over  $[0, 1]^d$ , we speak of binomial input; if  $\chi = \chi'_\lambda$  is a homogenous Poisson point process of intensity  $\lambda > 0$  we speak of Poisson input.

Define the Voronoi cell  $v_\chi(x)$  of nucleus  $x$  with respect to  $\chi$  as the closed set of points closer to  $x$  than to  $\chi$

$$v_\chi(x) = \{y \in \mathbb{R}^d : \forall x' \in \chi, d(x, y) \leq d(x', y)\}.$$

The Voronoi approximation  $K_\chi$  of  $K$  is the closed set of all points which are closer to  $K \cap \chi$  than to  $K^c \cap \chi$ . Its name comes from the relation

$$K_\chi = \bigcup_{x \in \chi \cap K} v_\chi(x).$$

The volume  $\varphi(\chi) = \text{Vol}(K_\chi)$  first arised in [16] as discriminating statistics in the two-sample problem. These authors proved a strong law of large numbers in dimension 1 for the volume approximation. Explicit rates of convergence in higher dimensions were obtained by Reitzner and Heveling [13], who proved that if  $K$  is convex and compact and  $\chi = \chi'_\lambda$  then

$$\begin{aligned} \mathbf{E}\varphi(\chi) &= \text{Vol}(K), \\ \mathbf{Var}(\varphi(\chi)) &\leq C\lambda^{-1-1/d}S(K), \end{aligned}$$

where  $S(K)$  is the surface area of  $K$ , all constants can be made explicit and depend only on  $d$ . They also studied the quantity  $\varphi_{\text{Per}}(\chi) = \text{Vol}(K \Delta K_\chi)$  to estimate the perimeter, after suitable renormalisation. Reitzner, Spodarev and Zaporozhets [20] extended these results to sets with finite variational perimeter, and also gave upper bounds for  $\mathbf{E}|\varphi(\chi'_\lambda)^q - \text{Vol}(K)^q|$  for  $q \geq 1$ . Schulte [23] obtained a matching lower bound for the variance with convex  $K$ , i.e.  $cS(K)\lambda^{-1-1/d} \leq \mathbf{Var}(\varphi(\chi))$ , and derived the corresponding CLT

$$\frac{\varphi(\chi) - \mathbf{E}\varphi(\chi)}{\sqrt{\mathbf{Var}(\varphi(\chi))}} \xrightarrow{(d)} N.$$

Very recently, Yukich [26] gave quantitative Berry-Esseen bounds for this CLT similar to the ones that are stated here for binomial input.

When dealing with binomial input, which has been less studied than Poisson input, it is necessary to assume that  $K \subset (0, 1)^d$  and redefine  $K_\chi$  as

$$K_\chi = \bigcup_{x \in \chi \cap K} v_\chi(x) \cap [0, 1]^d,$$

in order to avoid trivial complications due to possibly infinite cells. Penrose [19] proved the remarkable fact that for  $\chi = \chi_n$

$$\begin{aligned}\mathbf{E}\varphi(\chi) &\rightarrow \text{Vol}(K), \\ \mathbf{E}(\varphi_{\text{Per}}(\chi)) &\rightarrow 0,\end{aligned}$$

almost surely, with no need for assumptions on  $K$ 's shape.

To further assess the quality of the approximation with binomial input, we must quantify the previous convergence. The unbiasedness of the Poisson case does not occur with binomial input, mainly because of edge effects. Nevertheless those effects seem to decrease exponentially with the distance, like is customary for Voronoi cells. The following result shows that the bias of the estimator  $\varphi(\chi_n)$  decreases geometrically with  $n$ , therefore it is negligible with respect to the standard deviation, as shown in the following sections. Also, it still holds when  $(0, 1)^d$  is replaced by an arbitrary set  $U$  containing  $K$  in its interior.

**Theorem 2.** *Assume that  $K$  is a compact set with positive volume and let  $U$  be an open set containing  $K$ . Let  $\chi_n = \{X_i, 1 \leq i \leq n\}$  be iid uniform variables on  $U$ . Then there is a constant  $0 < c < 1$  depending only on  $K$  and  $d$  such that for  $n \geq 1$ ,*

$$|\mathbf{E}\text{Vol}(K_{\chi_n}) - \text{Vol}(K)| \leq c^n.$$

*Proof.* Let  $\chi_k = \{X_i, i \leq k\}$ . By homogeneity of the problem we can suppose  $\text{Vol}(U) = 1$ . The Voronoi approximation  $K_{\chi_n}$  of  $K$  satisfies

$$\begin{aligned}\mathbf{E}(\text{Vol}(K_{\chi_n})) &= \sum_{i=1}^n \mathbf{E}(1_{X_i \in K} \text{Vol}(v_{\chi_n}(X_i) \cap U)) \\ &= n \mathbf{E}(1_{X_n \in K} \text{Vol}(v_{\chi_{n-1}}(X_n) \cap U)) \\ &= n \int_K \mathbf{E}\text{Vol}(v_{\chi_{n-1}}(x) \cap U) dx.\end{aligned}\tag{5}$$



Take  $0 < r < \frac{1}{2}d(K, U^c)$ . We have for all  $x \in K$

$$\begin{aligned}
\mathbf{E}(\text{Vol}(v_{\chi_{n-1}}(x) \cap U)) &= \mathbf{E}\left(\int_U 1_{y \in v_{\chi_{n-1}}(x)} dy\right) \\
&= \mathbf{E}\left(\int_U 1_{B(y, \|y-x\|) \cap \chi_{n-1} = \emptyset} dy\right) \\
&= \int_U \mathbf{P}(B(y, \|y-x\|) \cap \chi_{n-1} = \emptyset) dy \\
&= \int_U (1 - \text{Vol}(B(y, \|y-x\|) \cap U))^{n-1} dy \\
&= \int_{B(x,r)} (1 - \kappa_d \|y-x\|^d)^{n-1} dy + c_n
\end{aligned}$$

where

$$c_n = \int_{U \setminus B(x,r)} (1 - \text{Vol}(B(y, \|y-x\|) \cap U))^{n-1} dy.$$

For  $y \in U \setminus B(x, r)$ , let  $B_y$  be the ball interiorly tangent to  $B(y, \|y-x\|)$  with center on  $[x, y]$  and radius  $r$ . We have  $B_y \subset B(y, \|y-x\|)$  by construction and  $B_y \subset U$  because  $B_y \subset B(x, 2r)$ . It follows that

$$c_n \leq \int_{U \setminus B(x,r)} (1 - \text{Vol}(B_y))^{n-1} dy = \int_{U \setminus B(x,r)} (1 - \kappa_d r^d)^{n-1} dy \leq c_0^n$$

for some  $0 < c_0 < 1$ , noticing that  $\kappa_d r^d < \text{Vol}(U) \leq 1$  because  $B(x, r) \subset U$ . From there, a polar change of coordinates yields

$$\begin{aligned}
\mathbf{E}(\text{Vol}(v_{\chi_{n-1}}(x) \cap U)) &= \int_0^r d\kappa_d t^{d-1} (1 - \kappa_d t^d)^{n-1} dt + c_n \quad (\text{because a } d\text{-sphere has surface } d\kappa_d) \\
&= \left[ -\frac{(1 - \kappa_d t^d)^n}{n} \right]_0^r + c_n \\
&= \frac{1}{n} + O(c^n)
\end{aligned}$$

for some  $c \in (0, 1)$ . Reporting in (5) yields the result.  $\square$

Recalling that the estimator is unbiased if the underlying sample is Poisson in  $\mathbb{R}^d$ , this pleads in favor of Voronoi approximation against other estimators [10, 21] where the bias is not known and does not seem to be negligible.

## 2.1 Asymptotic normality

This subsection is concerned with the results of [17], where it is shown that with binomial input, the volume approximation  $K_\chi$  is asymptotically normal when the number of points tends to  $\infty$ . Variance asymptotics and upper bounds on the speed of convergence for the Kolmogorov distance are also given.

We begin by stating the boundary regularity condition necessary for these results to hold, which is related to the boundary densities studied in the previous section. As explained in the introduction, it can be seen as a weakened form of the standardness assumption. Define, for all  $r > 0$ , the boundary neighbourhoods

$$\begin{aligned}\partial K_r &= \partial K + B(0, r), \\ \partial K_r^- &= \partial K_r \cap K, \\ \partial K_r^+ &= \partial K_r \cap K^c.\end{aligned}$$

**Assumption 2** (Boundary permeability condition). *A set  $K$  with no improper points satisfies the boundary permeability condition whenever*

$$\liminf_{r>0} \frac{1}{\text{Vol}(\partial K_r)} \left( \int_{\partial K_r^+} f_r^2(x) dx + \int_{\partial K_r^-} g_r^2(x) dx \right) > 0. \quad (6)$$

The following proposition gives a more meaningful equivalent for Assumption 2.

**Proposition 2.** *Assumption 2 holds if and only if*

$$\liminf_{r>0} \frac{1}{\text{Vol}(\partial K_r) \kappa_d r^d} \int_{K \times K^c} \mathbf{1}_{\|x-y\| \leq r} dx dy > 0. \quad (7)$$

*Proof.* Let us begin by establishing the relation between the expression of (7) and  $K$ 's boundary densities. By Fubini's theorem

$$\begin{aligned}\int_{K \times K^c} \frac{\mathbf{1}_{\|x-y\| \leq r}}{\kappa_d r^d} dx dy &= \int_K \frac{\text{Vol}(B(x, r) \cap K^c)}{\text{Vol}(B(x, r))} dx \\ &= \int_{K^c} \frac{\text{Vol}(B(x, r) \cap K)}{\text{Vol}(B(x, r))} dx,\end{aligned}$$

which rewrites simply as

$$\int_{K^c} f_r = \frac{1}{\kappa_d r^d} \int_{K \times K^c} \mathbf{1}_{\|x-y\| \leq r} dx dy = \int_K g_r, \quad (8)$$

by definition of boundary densities.

Consider the function  $h_r = \mathbf{1}_K g_r + \mathbf{1}_{K^c} f_r$ . We have  $0 \leq h_r \leq 1$  and  $h_r = 0$  outside of  $\partial K_r$ . Applying the Cauchy-Schwarz inequality gives

$$\int h_r^2 \leq \int h_r \leq \sqrt{\text{Vol}(\partial K_r)} \sqrt{\int h_r^2}$$

which rewrites as

$$\begin{aligned} \frac{1}{\text{Vol}(\partial K_r)} \int_{\partial K_r^+} f_r^2 + \int_{\partial K_r^-} g_r^2 &\leq \frac{1}{\text{Vol}(\partial K_r)} \left( \frac{2}{\kappa_d r^d} \int_{K \times K^c} \mathbf{1}_{\|x-y\| \leq r} dx dy \right) \\ &\leq \sqrt{\frac{1}{\text{Vol}(\partial K_r)} \left( \int_{\partial K_r^+} f_r^2 + \int_{\partial K_r^-} g_r^2 \right)}, \end{aligned}$$

so that clearly (6) and (7) are equivalent.  $\square$

**Remark 1.** If  $K$  is bi-standard with constant  $\varepsilon$  then (6) holds as well with the left hand being greater than  $\varepsilon^2$ . Hence bi-standariness implies the boundary permeability condition.

**Remark 2.** Note that

$$\text{Vol}(\partial K_r^+) \geq \frac{1}{\kappa_d r^d} \int_{K \times K^c} \mathbf{1}_{\|x-y\| \leq r} dx dy,$$

so that  $\text{Vol}(\partial K_r^+) \ll \text{Vol}(\partial K_r)$  prevents (7) from being satisfied. Of course, the same reasoning holds with  $\partial K_r^-$  instead. In other words, it is necessary for the boundary permeability condition to be fulfilled that both sides of the boundary have comparable volumes.

We reproduce below the result derived in [17, Th. 6.1] for Voronoi approximation, modified to measure the distance to the normal of the variable  $\text{Vol}(K_{\chi_n}) - \text{Vol}(K)$ , instead of  $\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})$  like in the original result. This subtlety is important for dealing with practical applications

and obtaining confidence intervals for  $\text{Vol}(K)$ . We deal with Kolmogorov distance, also adapted to confidence intervals, and defined by

$$d_{\mathcal{K}}(U, V) := \sup_{t \in \mathbb{R}} |\mathbf{P}(U \leq t) - \mathbf{P}(V \leq t)|,$$

for any random variables  $U, V$ .

**Theorem 3.** *Let  $K$  be a compact subset of  $(0, 1)^d$ . Assume that for some  $s < d$*

$$0 < \liminf_{r>0} r^{s-d} \text{Vol}(\partial K_r) \leq \limsup_{r>0} r^{s-d} \text{Vol}(\partial K_r) < \infty, \quad (9)$$

and that  $K$  satisfies the boundary permeability condition (Assumption 2). Then

$$0 < \liminf_{r>0} \frac{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}{n^{-2+s/d}} \leq \limsup_{r>0} \frac{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}{n^{-2+s/d}} < \infty, \quad (10)$$

and for all  $\varepsilon > 0$  there is  $C_\varepsilon > 0$  such that for all  $n \geq 1$

$$d_{\mathcal{K}} \left( \frac{\text{Vol}(K_{\chi_n}) - \text{Vol}(K)}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}}, N \right) \leq C_\varepsilon n^{-s/2d} \log(n)^{4-s/d+\varepsilon}, \quad (11)$$

where  $N$  is a standard Gaussian variable.

*Proof.* This result is almost exactly [17, Th. 6.1], except that there it is proved that

$$d_{\mathcal{K}} \left( \frac{\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}}, N \right) \leq C_\varepsilon n^{-s/2d} \log(n)^{4-s/d+\varepsilon}. \quad (12)$$

To have a similar bound involving  $\text{Vol}(K)$  instead of  $\mathbf{E}\varphi(\text{Vol}(K_{\chi_n}))$ , let us first remark that for  $\delta \in \mathbb{R}$ , a random variable  $U$ , and  $V = U + \delta$ ,

$$d_{\mathcal{K}}(V, N) \leq d_{\mathcal{K}}(V, N + \delta) + d_{\mathcal{K}}(N + \delta, N) \leq d_{\mathcal{K}}(U, N) + (2\pi)^{-1/2} |\delta|,$$

since  $d_{\mathcal{K}}(V, N + \delta) = d_{\mathcal{K}}(U, N)$ . It follows that

$$\begin{aligned} d_{\mathcal{K}}\left(\frac{\text{Vol}(K_{\chi_n}) - \text{Vol}(K)}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}}, N\right) &\leq d_{\mathcal{K}}\left(\frac{\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}}, N\right) \\ &\quad + (2\pi)^{-1/2} \left| \frac{\mathbf{E}\text{Vol}(K_{\chi_n}) - \text{Vol}(K)}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}} \right| \\ &\leq d_{\mathcal{K}}\left(\frac{\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}}, N\right) + \frac{O(c^n)}{n^{-1+s/2d}} \end{aligned}$$

for some  $c \in (0, 1)$  by Theorem 2. Reporting the bounds of (12) yields (11).  $\square$

**Remark 3.** The fact that  $[0, 1]^d$  is the support of the random sampling variables does not seem to have a great importance. Uniformity over  $[0, 1]^d$  eases certain estimates in the proof of [17, Th. 6.1] related to stationarity, but is not essential. If the variables are only assumed to have a positive continuous density  $\kappa(x) > 0$  on an open neighborhood of  $\partial K$ , it should be enough for similar results to hold. See Theorem 2, or [19], for rigorous results in this direction.

**Remark 4.** If  $K$  satisfies all the hypotheses of Theorem 3 except the boundary permeability condition, then we have

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left| \mathbf{P}\left(\frac{\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})}{\sqrt{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}} \leq t\right) - \mathbf{P}(N \leq t) \right| \\ \leq C_{\varepsilon} n^{\varepsilon} (\sigma^{-2} n^{-2+s/2d} + \sigma^{-3} n^{-3+s/d} + \sigma^{-4} n^{-4+s/d}) \end{aligned} \tag{13}$$

where  $\sigma^2$  is the variance of  $\text{Vol}(K_{\chi_n})$ . See [17, Th. 6.2] for more details.

**Remark 5.** Set-estimation literature is also concerned with perimeter approximation [15, Sec. 11.2.1]. In the context of Voronoi approximation, the study of the functional  $\text{Vol}(K_{\chi_n} \Delta K)$  has been initiated in [13, 20]. Although the result is not formally stated, a bound of the form (13) for this functional is available using the exact same method. One has to work separately to obtain a variance lower bound. Such a result with Poisson input has been derived very recently in the paper of Yukich [26].

Results regarding the volume of the symmetric difference between the set and its approximation can be used to compare Voronoi approximation with another estimator. Indeed, the bound in  $n^{-1/d}$  given in [13] for  $\mathbf{E}\text{Vol}(K_{\chi_n} \Delta K)$  is better than the bound in  $(nr_n^d)^{-1/2}$  of [4], who use the Devroye-Wise estimator with a smoothing parameter  $r_n \ll n^{-1/(d+1)}$ .

The consequences of Theorems 1 and 3 for sets  $K$  with self-similar boundary are immediate, condition (9) automatically holds by Proposition 1.

**Corollary 1.** *Let  $K$  be a compact set such that  $\partial K$  is a finite union of self-similar sets satisfying Assumption 1. Then (10) and (11) hold.*

This corollary applies to the Von Koch flake with  $s = \ln(4)/\ln(3)$  (Example 1). The conclusions of Theorem 3 also apply for instance to the Von Koch anti flake, where three Von Koch curves are stucked together like for building the flake, but here the curves are pointing inwards, and not outwards (Figure 2.1). Assumption 1 is not satisfied on the whole boundary, but it is within an open ball of  $\mathbb{R}^d$  intersecting one and only one of the three curves, and having (4) on a self-similar  $E$  with same Minkowski dimension as  $\partial K$  is actually enough for the boundary permeability condition to hold.

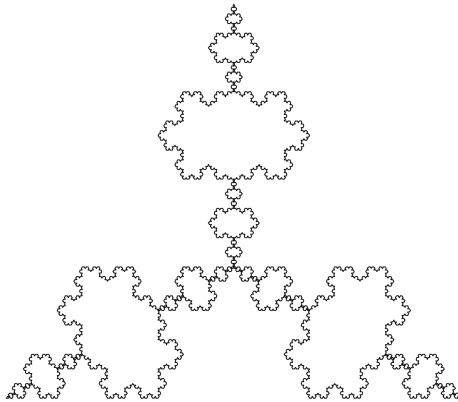


Figure 2.1: The Von Koch antiflake

In Section 2.3 we exhibit an example of a set  $K$  such that  $\partial K$  is self-similar and  $K$  does not satisfy Assumption 1. We run simulations suggesting that (10) is also false. Our theorem gives a set of sufficient conditions, but other

versions should be valid. For instance, the question of whether a compact set  $K \subset \mathbb{R}^2$  whose boundary is a locally self-similar Jordan curve satisfies the conclusions of the theorem above seems to be of interest.

## 2.2 Convergence for the Hausdorff distance

In this subsection we will make use of  $r$ -coverings and  $r$ -packings. Consider a collection  $\mathcal{B}$  of balls having radius  $r$  and centers belonging to some set  $E \subset \mathbb{R}^d$ .  $\mathcal{B}$  is said to be an  $r$ -packing of  $E$  if the balls are disjoint. It is an  $r$ -covering if the balls cover  $E$ .

The size of minimal coverings and maximal packings is closely related to the Minkowski dimension of  $E$ . A necessary and sufficient condition for  $E$  to have upper and lower Minkowski contents is that, for all small enough  $r$ , we can find an  $r$ -covering of  $E$  with less than  $Cr^{-s}$  balls, and an  $r$ -packing of the same set with more than  $cr^{-s}$  balls. More related results can be found in [18].

To estimate with precision  $r$  the shape of a set  $E$  by a point process  $\chi$  it is often necessary to request that every point of  $E$  is at distance less than  $r$  of  $\chi$ . In the context of Voronoi approximation, this is made precise by the following lemma. Note that we only require  $\chi$  to be dense enough near  $\partial A$ . This is, as suggested in the introduction, because Voronoi approximation fills in the interior regions of  $K$  where points of  $\chi$  are scarce.

**Lemma 4.** *Let  $\chi \subset \mathbb{R}^d$  be a locally finite non-empty set.*

1. *If every point  $x$  of  $\text{cl}(\partial K_r^+)$  satisfies  $d(x, \chi) < r$  then  $K_\chi \subset K + B(0, r)$ .*
2. *If every point  $x$  of  $\text{cl}(\partial K_r^+)$  satisfies  $d(x, \chi) < r$  and every point  $x$  of  $\text{cl}(\partial K_r^-)$  satisfies  $d(x, \chi \cap K) < r$  then  $d_H(K, K_\chi) \leq r$ .*
3. *If every point  $x$  of  $\text{cl}(\partial K_r^+)$  satisfies  $d(x, \chi \cap K^c) < r$  and every point  $x$  of  $\text{cl}(\partial K_r^-)$  satisfies  $d(x, \chi \cap K) < r$  then  $d_H(\partial K, \partial K_\chi) \leq r$ .*
4. *If some point  $x \in \partial K$  satisfies  $d(x, \chi \cap K) \geq 3r$  and  $d(x, \chi \cap K^c) \leq r$  then  $d_H(K, K_\chi) \geq r$  and  $d_H(\partial K, \partial K_\chi) \geq r$ .*

*Proof.* We begin with the first point. Suppose  $x \in K_\chi$  satisfies  $d(x, K) \geq r$ . Then there is a point  $c_x \in \chi \cap K$  such that  $x \in v_\chi(c_x)$ . The segment

joining  $c_x$  and  $x$  contains points from  $\partial K$  so we can consider  $x_0$  the point of  $\text{cl}(\partial K_r^+)$  closest to  $x$  on that segment. We have  $d(x_0, \partial K) = r$  and  $x_0 \in K^c$  since otherwise there would be another point of  $\text{cl}(\partial K_r^+)$  closer to  $x$ . As a consequence  $d(x_0, c_x) \geq r$ . But then by assumption there is a point  $y$  of  $\chi$  such that  $d(x_0, y) < r$  and  $c_x$  isn't the point of  $\chi$  closest to  $x$ , which is a contradiction. Hence  $x \in K_\chi$  implies  $d(x, K) < r$  and  $K_\chi \subset K + B(0, r)$ .

Note that in the setting of points 2 and 3 we can apply the previous argument to  $K^c$  instead of  $K$ , the compactity of  $K$  not playing any role in the proof. Along with  $(K_\chi)^c = (K^c)_\chi$  this yields  $K_\chi^c \subset K^c + B(0, r)$ , which reformulates as  $K \setminus \text{cl}(\partial K_r^-) \subset K_\chi$  by taking complements. Hence in both cases we have the inclusions

$$K_\chi \subset K + B(0, r), \quad K_\chi^c \subset K^c + B(0, r),$$

and their reformulations

$$K \setminus \text{cl}(\partial K_r^-) \subset K_\chi, \quad K^c \setminus \text{cl}(\partial K_r^+) \subset K_\chi^c.$$

To prove the second point it is enough to show that  $K \subset K_\chi + B(0, r)$ . Let  $x$  be a point of  $K$ . If  $x \in K \setminus \text{cl}(\partial K_r^-)$  then  $x$  belongs to  $K_\chi$ . And if  $x$  is in  $\text{cl}(\partial K_r^-)$  then there is a point  $y$  of  $\chi \cap K$  such that  $d(x, y) < r$ . In all cases  $x \in K_\chi + B(0, r)$ .

We move on to point 3. The two inclusions  $K^c \setminus \text{cl}(\partial K_r) \subset K_\chi^c$  and  $K \setminus \text{cl}(\partial K_r) \subset K_\chi$  also show that if  $x$  satisfies  $d(x, \partial K) > r$ ,  $x$  is interior to either  $K_\chi$  or  $K_\chi^c$ . Hence  $\partial K_\chi \subset \partial K + \text{cl}(B(0, r))$ . Conversely, for every point  $x$  of  $\partial K$  there are points of both  $\chi \cap K$  and  $\chi \cap K^c$  inside  $B(x, r)$ , so  $B(x, r)$  contains a point of  $\partial K_\chi$ . Hence  $\partial K \subset \partial K_\chi + B(0, r)$  and  $d_H(\partial K, \partial K_\chi) \leq r$ .

Lastly, suppose the requirements of point 4 are met. Let  $y$  be a point of  $\chi \cap B(x, r) \cap K^c$ . Then all of the points in  $B(x, r)$  are closer to  $y$  than to the points outside of  $B(x, 3r)$ . Consequently all points  $B(x, r)$  must lie in Voronoi cells centered in  $K^c$ , and  $x \notin K_\chi + B(0, r)$  so that  $d_H(K, K_\chi) \geq r$ . The fact that  $B(x, r) \subset K_\chi^c$  also implies  $d(x, \partial K_\chi) \geq r$  and  $d_H(\partial K, \partial K_\chi) \geq r$ .  $\square$



Now we apply this lemma to show almost sure convergence of  $K_\chi$  in the sense of the Hausdorff distance. To formulate such a result, the concept of proper points (beginning of Section 1.3) proves to be useful. Improper points are invisible to the Voronoi approximation  $K_\chi$  of  $K$ . Though this has no incidence when measuring volumes, it becomes a nuisance when measuring Hausdorff distances.

The set  $K^{\text{prop}}$  of points proper to  $K$  can be thought of as the complement of the biggest open set  $O$  such that  $\text{Vol}(O \cap K) = 0$ , from which it follows that  $K^{\text{prop}}$  is compact and that  $K_\chi = K_\chi^{\text{prop}}$  a.s.

**Proposition 3.**  $K_{\chi_n} \xrightarrow[n \rightarrow +\infty]{} K^{\text{prop}}$  and  $\partial K_{\chi_n} \xrightarrow[n \rightarrow +\infty]{} \partial K^{\text{prop}}$  almost surely in the sense of the Hausdorff metric for both Poisson and binomial input.

*Proof.* Since  $K_\chi = (K^{\text{prop}})_\chi$  almost surely and  $K^{\text{prop}}$  has no improper points, this is equivalent to the fact that  $K_{\chi_n} \rightarrow K$  and  $\partial K_{\chi_n} \rightarrow \partial K$  almost surely when  $K$  has no improper points. By the Borel-Cantelli lemma it is enough to show that both series

$$\sum_{n \geq 1} \mathbf{P}(d_H(K_{\chi_n}, K) > r), \quad \sum_{n \geq 1} \mathbf{P}(d_H(\partial K_{\chi_n}, \partial K) > r)$$

are convergent for any positive  $r$ .

Consider  $r/2$ -coverings  $\mathcal{B}^+, \mathcal{B}^-$  of  $\text{cl}(\partial K_r^+), \text{cl}(\partial K_r^-)$  respectively. Since both sets are compact, these coverings can be made with finitely many balls. Set  $\mathcal{B} = \mathcal{B}^+ \cup \mathcal{B}^-$  and

$$V = \min \left( \min_{B \in \mathcal{B}^-} \text{Vol}(B \cap K), \min_{B \in \mathcal{B}^+} \text{Vol}(B \cap K^c) \right).$$

Because  $K$  and  $K^c$  have no improper points,  $V > 0$ . If every ball of  $\mathcal{B}^+$  contains a point of  $\chi \cap K^c$  and every ball of  $\mathcal{B}^-$  a point of  $\chi \cap K$ , then the requirements of points 2 and 3 in Lemma 4 are met. The probability of this not happening is bounded by  $|\mathcal{B}|(1 - V)^n$  for binomial input and  $|\mathcal{B}|e^{-nV}$  for Poisson input. In all cases the series associated with  $\mathbf{P}(d_H(K_{\chi_n}, K) > r)$  and  $\mathbf{P}(d_H(\partial K_{\chi_n}, \partial K) > r)$  converge, as required.  $\square$

A refinement of the method above gives an order of magnitude for  $d_H(K, K_\chi)$  with Poisson input, under assumptions on  $\partial K, f_r$  and  $g_r$  resembling those of Theorem 3. This requires better estimations of the probability of the points of Lemma 5 being met, which is the purpose of the following lemma.

**Lemma 5.** *Let  $A, \chi$  be non-empty sets, and  $\mathcal{B}$  a collection of balls centered on  $A$  with radii  $r$ . Write  $\mathcal{B}_\tau$  for the collection of balls having same centers as those of  $\mathcal{B}$  but radius  $\tau r$ , and choose  $\tau_1, \tau_2 > 0$  such that  $\tau_1 + \tau_2 = 1$ . If  $\mathcal{B}_{\tau_1}$  is a  $\tau_1 r$ -covering of  $A$  and every ball of  $\mathcal{B}_{\tau_2}$  contains a point of  $\chi$ , then  $A \subset \chi + B(0, r)$ .*

*Proof.* Let  $x$  be a point of  $A$ . By hypothesis, there is a ball of  $\mathcal{B}$  with center  $c$  such that  $d(x, c) < \tau_1 r$ , and also a point  $y$  of  $\chi$  such that  $d(y, c) < \tau_2 r$ . Hence  $d(x, y) < r(\tau_1 + \tau_2)$  and  $d(x, \chi) < r$ . So indeed every point of  $A$  is at distance less than  $r$  of  $\chi$ .  $\square$

This handy lemma is meant to give probability estimations of events of the type  $A \subset \chi + B(0, r)$ , which are useful outside the context of Voronoi approximation. Typically,  $\chi$  is chosen to be a random point process, and the covering  $\mathcal{B}_{\tau_1}$  is chosen deterministically with as few balls as possible, often  $C\tau_1^{-d}r^{-d}$ . Bounding the probability that a ball of  $\mathcal{B}_{\tau_2}$  does not intersect  $\chi$  then gives an upper bound of the form

$$\mathbf{P}(A \not\subset \chi + B(0, r)) \leq |\mathcal{B}| \max_{B \in \mathcal{B}_{\tau_2}} \mathbf{P}(B \cap \chi = \emptyset).$$

The estimations obtained in such applications are less sensible to the number of balls in  $\mathcal{B}$  than to their size. Hence optimal results are obtained when  $\tau_1$  is small.

For example, the reader may use Lemma 5 to derive [9, Th. 1] and its counterpart for Poisson input, which are concerned with the order of magnitude of  $d_H(K, K \cap \chi)$  with  $\chi$  an homogenous point process. Note that use of Minkowski contents and boundary densities give slightly better bounds, which turn out to be optimal, see Remark 10.

**Theorem 4.** *Suppose that  $\partial K$  has Minkowski dimension  $s > 0$  with upper and lower contents, and that for all  $r$  small enough,*

$$\begin{aligned} f_r &\geq \varepsilon \text{ on } K, \\ g_r &\geq \varepsilon \text{ on } K^c. \end{aligned}$$

*Then we have*

$$\begin{aligned} \mathbf{P} \left( \alpha \leq \frac{d_H(K, K_{\chi'_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta \right) &\xrightarrow{\lambda \rightarrow \infty} 1 \\ \mathbf{P} \left( \alpha \leq \frac{d_H(\partial K, \partial K_{\chi'_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta \right) &\xrightarrow{\lambda \rightarrow \infty} 1 \end{aligned}$$

where  $\chi'_\lambda$  is a Poisson point process of intensity  $\lambda$  and  $\alpha, \beta$  satisfy  $\alpha < \alpha_K, \beta > \beta_K$  with

$$\alpha_K = \frac{1}{3} \left( \frac{s}{d\kappa_d(1-\varepsilon)} \right)^{1/d},$$

$$\beta_K = \left( \frac{s}{d\kappa_d\varepsilon} \right)^{1/d}.$$

*Proof.* The approach of the proof is to tune  $r$  in Lemma 4 in order to have the events of point 3 happen with high probability. We shall only show the assertions regarding  $d_H(\partial K, \partial K_\chi)$ , since the exact same arguments hold with  $d_H(K, K_\chi)$  as well.

We start with the upper bound. For all  $\lambda$  let  $\Omega_\lambda$  be the event where all the requirements from point 3 of Lemma 4 are met with  $\chi = \chi'_\lambda$ ,  $r = r_\lambda = \beta(\lambda^{-1} \ln(\lambda))^{1/d}$ . Hence  $\{d_H(\partial K, \partial K_\chi) > r\} \subset \Omega_\lambda^c$ . We shall show that  $\mathbf{P}(\Omega_\lambda^c) \rightarrow 0$ .

Choose  $\tau_1, \tau_2 < 1$  so that  $\tau_1 + \tau_2 = 1$  and  $\tau_2\beta > \beta_K$ . Let  $\mathcal{B}^+$  be a collection of balls with radius  $r$  and centers on  $\text{cl}(\partial K_r^+)$ . As in Lemma 5, call  $\mathcal{B}_\tau^+$  the collection of balls with same centers as those of  $\mathcal{B}^+$ , but radius  $\tau r$ . Define  $\mathcal{B}^-, \mathcal{B}_\tau^-$  similarly and set  $\mathcal{B} = \mathcal{B}^+ \cup \mathcal{B}^-$ . Note that  $\mathcal{B}$  depends on  $\lambda$ , but  $\tau_1, \tau_2$  do not.

We can and do choose  $\mathcal{B}^+, \mathcal{B}^-$  so that  $\mathcal{B}_{\tau_1}^+, \mathcal{B}_{\tau_1}^-$  are coverings of  $\text{cl}(\partial K_r^+)$  and  $\text{cl}(\partial K_r^-)$  respectively, and  $|\mathcal{B}|$  has less than  $C\tau_1^{-d}r^{-s} = C\tau_1^{-d}(\lambda/\ln\lambda)^{-s/d}$  balls. Indeed, consider  $\tau_1 r/2$ -packings of  $\partial K_r^+$  and  $\partial K_r^-$ , both optimal in the sense that no ball can be added without losing the packing property. Because of volume issues, the packings have less than  $C\tau_1^{-d}r^{-s}$  balls, and because of the optimality assumption doubling the radii of the balls gives the desired  $\tau_1 r$ -coverings.

The intersection of  $K$  with a ball  $B \in \mathcal{B}_{\tau_2}^-$  of center  $x$  has volume exactly  $\kappa_d(\tau_2 r)^d f_{\tau_2 r}(x)$ . Because  $f_r \geq \varepsilon$  for large enough  $\lambda$  and  $\tau_2\beta > \beta_K$  it follows that

$$\mathbf{P}(B \cap \chi \cap K = \emptyset) \leq \exp(-\lambda\tau_2^d \varepsilon \kappa_d r^d) = \lambda^{-s/d-\delta}$$

for some  $\delta > 0$ . The same bound is valid for  $\mathbf{P}(B \cap \chi \cap K^c = \emptyset), B \in \mathcal{B}_{\tau_2}^+$ .

Applying Lemma 5 twice with  $A = \partial K_r^+, \partial K_r^-$  successively gives

$$\mathbf{P}(\Omega_\lambda^c) \leq |\mathcal{B}| \lambda^{-s/d-\delta} \leq C\tau_1^{-d} \ln(\lambda)^{s/d} \lambda^{-\delta}$$

so that, since  $\tau_1$  is fixed,  $\mathbf{P}(\Omega_\lambda^c) \rightarrow 0$  as desired.

The proof for the lower bound is quite similar. Fix  $\delta > 0$ , and redefine  $\Omega_\lambda$  to be the event where the requirements described in point 4 of Lemma 4 are met for  $\chi = \chi'_\lambda$ ,  $r = r_\lambda = \alpha(\ln(\lambda)\lambda^{-1})^{1/d}$  with  $\alpha < \alpha_K$ . Again, we shall show  $\mathbf{P}(\Omega_\lambda^c) \rightarrow 0$ . Let  $\mathcal{B} = \mathcal{B}_\lambda$  be a  $3r$ -packing of  $\partial K$ . We can assume  $|\mathcal{B}| \geq cr^{-s}$ .

The probability of there being no points of  $K \cap \chi_\lambda$  in a ball  $B(x, 3r)$  of  $\mathcal{B}$  and at least one point of  $K^c \cap \chi$  in  $B(x, r)$  for a point  $x$  in the boundary is exactly

$$\exp(-\lambda\kappa_d(1 - g_{3r}(x))3^d r^d) (1 - \exp(-\lambda\kappa_d g_r(x)r^d))$$

because  $B(x, 3r) \cap K^c$  and  $B(x, r) \cap K$  are disjoint. So we have the following upper bound, for  $\lambda$  big enough

$$\mathbf{P}(\Omega_\lambda^c) \leq (1 - e^{-\lambda\kappa_d(1-\varepsilon)3^d r^d} (1 - e^{-\lambda\kappa_d r^d \varepsilon}))^{|\mathcal{B}|}.$$

We would like the right hand to go to 0 with  $\lambda$ . Taking logarithms this is equivalent to

$$|\mathcal{B}| \exp(-\lambda\kappa_d(1 - \varepsilon)3^d r^d) (1 - \exp(-\lambda\kappa_d r^d \varepsilon)) \xrightarrow{\lambda \rightarrow +\infty} +\infty.$$

Because  $\exp(-\lambda\kappa_d(1 - \varepsilon)3^d r^d) = \lambda^{\delta-s/d}$  with  $\delta > 0$ ,  $\exp(-\lambda\kappa_d r^d \varepsilon) \rightarrow 0$  and  $|\mathcal{B}| \geq c(\lambda/\ln(\lambda))^{s/d}$ , it is indeed the case. □

The proof and the result call for some comments. Most of them are minor variants on the result which were not included in the proof for clarity's sake.

**Remark 6.** It is possible to dispose of the hypothesis that  $\partial K$  has Minkowski upper and lower contents, by using instead the so-called upper and lower Minkowski dimension, which always exist, see [18]. In particular, we can always do the coverings in the proof with  $Cr^{-d}$  balls, so the upper bound still holds after replacing  $s$  by  $d$  in the expression of  $\beta_K$ . This compares with the result given by Calka and Chenavier in [5, Corollary 2]. One can also show, using the fact that  $K$  is bounded and has positive volume, that  $\mathcal{H}^{d-1}(\partial K) > 0$  so that  $s$  can be replaced by  $d - 1$  in the expression of  $\alpha_K$ . Hence a lower bound also holds with no assumption on  $\partial K$ 's geometry when  $d \geq 2$ .

For the results concerned with  $d_H(\partial K, \partial K_\chi)$ , this is a remarkable feature that to our knowledge no other estimators possess. For instance, in [9] a so-called expandability condition is required to obtain similar rates with the Devroye-Wise estimator.

**Remark 7.** If  $s = 0$  and  $\partial K$  has Minkowski contents then actually  $d = 1$ ,  $\partial K$  has a finite number of points, and  $d_H(K, K_\chi)$  has order  $\lambda^{-1}$  in the sense that for  $\lambda$  large enough

$$\mathbf{P}(d_H(K, K_{\chi'_\lambda})\lambda > t) \leq 2|\partial K| \exp(-2\epsilon t),$$

which is enough to guarantee the existence of moments of all orders for  $d_H(K, K_\chi)\lambda$ . This is not true of other shape estimators, and is due to the fact that Voronoi approximation only requires  $\chi$  to be dense near  $\partial K$  and not on all of  $K$ . If we don't have Minkowski contents the situation might be more delicate.

**Remark 8.** Better estimations of the  $\mathbf{P}(\Omega_\lambda^c)$  in the proof along with an application of the Borel-Cantelli lemma yield the almost sure convergence rates advertised in the introduction. Explicitly

$$\alpha_K \leq \liminf_{n \rightarrow +\infty} \frac{d_H(K, K_{\chi'_n})}{(n^{-1} \ln(n))^{1/d}} \leq \limsup_{n \rightarrow +\infty} \frac{d_H(K, K_{\chi'_n})}{(n^{-1} \ln(n))^{1/d}} \leq \beta'_K$$

and similarly for  $d_H(\partial K, \partial K_\chi)$ , with  $\alpha_K, \beta_K$  as in Theorem 4 and  $\beta'_K = (\beta_K^d + (1/\kappa_d \epsilon))^{1/d}$ .

**Remark 9.** For binomial input, some minor changes in the proof give the same upper bound. It can't be done for the lower bound since we use the fact that  $\chi \cap A, \chi \cap B$  are independent when  $A$  and  $B$  are disjoint and  $\chi$  is a Poisson point process.

**Remark 10.** Using similar techniques as in the proof above it is possible to show that

$$\frac{d_H(K, K \cap \chi'_\lambda)}{(\lambda^{-1} \ln(\lambda))^{1/d}} \xrightarrow{\mathbf{P}} \left( \frac{2(d-1)}{d\kappa_d} \right)^{1/d}$$

if  $K$  has no improper points and  $\partial K$  is a  $\mathcal{C}^2$  manifold. Theorem 4 shows that, under the same assumptions, the above limit can be used as an upper bound for  $d_H(K, K_{\chi_\lambda})(\lambda/\ln(\lambda))^{-1/d}$ . Hence, as a shape estimator,  $K_\chi$  is not worse than  $\chi \cap K$ . It would be interesting to know if it is better in some sense, a question related to the optimality of the bounds in Theorem 4.

**Remark 11.** Applying point 2 of Lemma 4 instead of point 3 in the proof of the theorem yields a better result for  $d_H(K, K_\chi)$ . Specifically if  $f_r \geq \epsilon_f$  on  $K$  then

$$\mathbf{P} \left( \frac{d_H(K, K_{\chi'_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta \right) \xrightarrow{\lambda \rightarrow \infty} 1$$

whenever

$$\beta > \left( \frac{s}{d\kappa_d\varepsilon_f} \right)^{1/d}.$$

Together with Remark 6 this shows that inner standardness is a sufficient assumption to have convergence rates for  $d_H(K, K_\chi)$ .

### 2.3 A counter-example

Here we construct a set  $K_{\text{cantor}}$  with self-similar boundary not satisfying the boundary permeability condition. This example shows that Theorem 1 cannot be generalised by dropping Assumption 1, even if the conclusion is weakened.

The example  $K$  below is uni-dimensional, but a counter-example in dimension 2 can be obtained by considering  $K \times [0, 1]$ .

**Example 2.** Let  $E \subset \mathbb{R}$  the self-similar set generated by the similarities  $\phi_1 : x \mapsto x/3$ ,  $\phi_2 : x \mapsto (2+x)/3$  who satisfy the open set condition with  $U = (0, 1)$ .  $E$  is in fact the Cantor set, and can be characterized as the set of points having a ternary expansion with no ones.

$K_{\text{cantor}}$  will be defined as the closure of open intervals of  $[0, 1] \setminus E$ . The trick is to choose few intervals with quickly decreasing length, so that  $f_r$  is small on most of  $K_{\text{cantor}}$ 's boundary, but to distribute them well so that  $\partial K_{\text{cantor}} = E$ .

To every positive integer  $n$  associate the sequence  $s^n$  of its digits in base 2 in reverse order and double the terms to get  $s^n$ . For example, since 6 is 110 in base 2,  $s^6 = (0, 2, 2)$ . This defines a bijection between  $\mathbb{N}$  and the set of finite sequences of zeroes and twos ending in 2, with the additional property that  $s^n$  always has length  $l_n \leq n$ . Now for all  $n$  define

$$\begin{aligned} a_n &= \frac{1}{3^{n+1}} + \sum_{k \geq 1} \frac{s_k^n}{3^k} \\ b_n &= \frac{2}{3^{n+1}} + \sum_{k \geq 1} \frac{s_k^n}{3^k} \\ A_n &= (a_n, b_n) \end{aligned}$$

We have the following ternary expansions

$$\begin{aligned}
a_n &= 0.s_1^n s_2^n \dots s_{l_n}^n 000\dots 01 \\
&= 0.s_1^n s_2^n \dots s_{l_n}^n 000\dots 0022222\dots \\
b_n &= 0.s_1^n s_2^n \dots s_{l_n}^n 000\dots 02
\end{aligned}$$

Now, set  $K = \text{cl}(\bigcup A_n)$ . We claim that  $K$  has no improper points,  $\partial K = E$  and that  $K$  does not satisfy the regularity condition of Theorem 3.

*Proof.* The first assertion is easy to prove. Being segments, the  $A_n$  have no improper points to themselves, so  $\bigcup A_n \subset K^{\text{PROP}}$  and  $K \subset K^{\text{PROP}}$  by taking closures.

For the second assertion we need to show that  $\partial K = K \setminus \bigcup A_n = \text{cl}(\bigcup \{a_n, b_n\})$ . We already have the obvious  $\partial K \subset K \setminus \bigcup A_n$ . Define

$$\begin{aligned}
a'_n &= \frac{1}{3^{n+1}} - \frac{2}{3^{l_n}} + \sum_{k \geq 1} \frac{s_k^n}{3^k} \\
b'_n &= \frac{2}{3^{n+1}} - \frac{2}{3^{l_n}} + \sum_{k \geq 1} \frac{s_k^n}{3^k} \\
A'_n &= (a'_n, b'_n)
\end{aligned}$$

Since for all  $n$ ,  $s_{l_n}^n = 2$ , the corresponding ternary expansions are

$$\begin{aligned}
a'_n &= 0.s_1^n s_2^n \dots s_{l_n-1}^n 000\dots 01 \\
&= 0.s_1^n s_2^n \dots s_{l_n-1}^n 000\dots 0022222\dots \\
b'_n &= 0.s_1^n s_2^n \dots s_{l_n-1}^n 000\dots 02
\end{aligned}$$

If  $x \in A_i \cap A'_j$  then every ternary expansion of  $x$  has the same digits as the finite ternary expansions of  $a_i, a'_j$  up to the first 1, which is impossible. So  $\bigcup A'_n$  is an open set disjoint from  $\bigcup A_n$  and hence from  $K$ . Furthermore,  $\bigcup A'_n$  is dense near the  $a_n$ , because for all  $k, N \in \mathbb{N}^*$ , we can find an  $a'_{k'}$  whose ternary expansion has the same  $N$  first digits as the non-terminating expansion of  $a_k$ , so that  $d(a_k, a'_{k'}) \leq 1/3^N$ . A similar argument works for the  $b_n$ , so that the  $a_n, b_n$  belong to  $\partial K$  and, since the latter is closed,  $\text{cl}(\bigcup \{a_n, b_n\}) \subset \partial K$ .

Finally, consider a point  $x \in K \setminus \bigcup A_n$ . For all  $r > 0$ ,  $B(x, r)$  contains a point from an  $A_k$ , and since  $x \notin A_k$ , one of the two points  $a_k, b_k$  must also be in  $B(x, r)$ . Consequently,  $x$  is also an accumulation point of  $\bigcup \{a_n, b_n\}$ . We just proved that  $K \setminus \bigcup A_n \subset \text{cl}(\bigcup \{a_n, b_n\})$ . Putting this together with the previous two inclusions we get the desired equality.

Since for all  $x \in E$ ,  $N \in \mathbb{N}^*$  we can find an  $a_k$  with the same first  $N$  digits as  $x$  in base 3, the  $a_n$  are dense in  $E$  and  $E \subset \partial K$ . Conversely,  $\partial K \subset E$ , since the  $a_n, b_n$  belong to  $E$ , who is closed.

For the last assertion, pick any  $r > 0$  and set  $N = 2\lceil -\log_3(r) \rceil$ . Let  $X$  be the union of the balls of radius  $r$  centered on the endpoints of the  $N$  first  $A_n$ .  $X$  has area at most  $-4r \log_3(r)$  and for any  $x \in \partial K_r \setminus X$ ,  $B(x, r)$  does not intersect the  $A_k, k \leq N$ . Since  $\text{Vol}(\partial K) = 0$

$$\text{Vol}(K \setminus (A_1 \cup A_2 \dots \cup A_N)) = \text{Vol}\left(\bigcup_{n>N} A_n\right) = \frac{1}{2 \cdot 3^{N+1}} \leq r^2.$$

But  $\text{Vol}(\partial K_r)$  has order  $r^{1-\ln(2)/\ln(3)}$  and

$$\text{Vol}(\partial K_r^-) \leq \text{Vol}(X) + \text{Vol}(K \setminus (A_1 \cup A_2 \dots \cup A_N)) \leq -4r \log_3(r) + r^2$$

so that  $\text{Vol}(\partial K_r^-) \ll \text{Vol}(\partial K_r)$ . According to Remark 2, this prevents (7) from holding.  $\square$

Simulations were made for the quality of the Voronoi volume approximation with this set  $K$ . The magnitude order of the empirical variance of  $\text{Vol}(K_{\chi_n})$  seems to be  $n^\tau$  with  $\tau \approx -1.8$ , as shown in Figure 2.2. Looking at Theorem 3, the approximation behaves as if the set had a “nice” fractal boundary of dimension  $\approx 0.2$ , whereas its real fractal dimension is  $1 - \ln(2)/\ln(3) \approx 0.37$ .

Simulations also suggest that a central limit theorem still holds. Such a fact indicates that though the results of Lachieze-Rey and Peccati [17] seem to be generalisable, the variance of  $\text{Vol}(K_{\chi_n})$  is indeed related to the behaviour of  $f_r$  and  $g_r$  near  $\partial K$ .

**Example 3.** It is possible to construct other sets not satisfying the regularity condition of Assumption 2. If we don’t require  $\partial K$  to be a self-similar set, a much simpler example is given by

$$K = \text{cl}\left(\bigcup_{n \in \mathbb{N}^*} \left(\frac{1}{n} - \frac{1}{3^n}, \frac{1}{n}\right)\right).$$



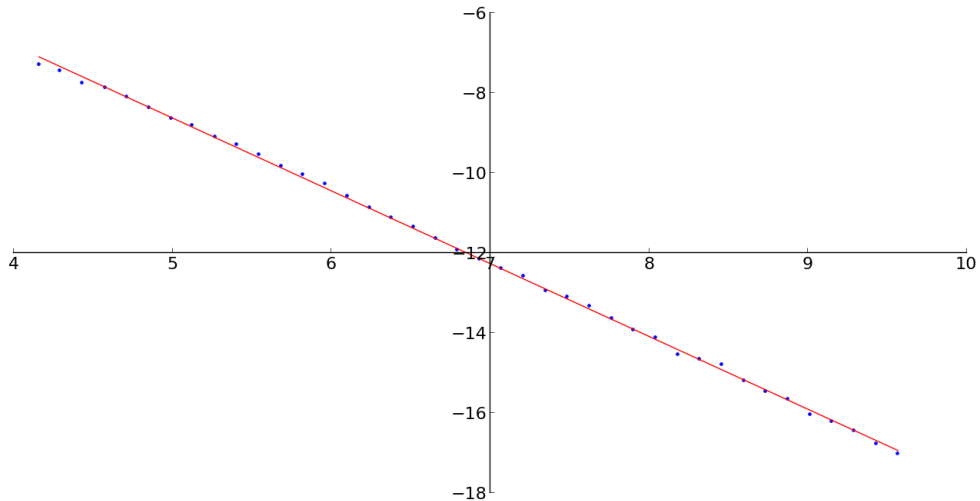


Figure 2.2: In blue  $\ln(\mathbf{Var}(K_{\chi_n}))$  as a function of  $\ln(n)$ , in red the associated linear regression. For each  $n$ , the variance was estimated with 1000 realisations of  $\text{Vol}(K_{\chi_n})$ .

Intentionally,  $\partial K$  looks like the set  $\{n^{-1}, n \in \mathbb{N}^*\}$ , who is often given as an example of a countable set with positive Minkowski dimension.  $K$  has no improper points, its boundary has Minkowski dimension  $1/2$  with upper and lower contents, but  $K$  does not satisfy (6) or (7). This can be proved using the same methods as in Example 2. Again, simulations tend to show that the variance of  $\text{Vol}(K_{\chi_n})$  is about  $n^\tau$  with  $\tau \approx -1,8$  and that a central limit theorem still holds.

## References

- [1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Science Publications, 2000.
- [2] J. R. Berrendero, A. Cholaquidis, A. Cuevas, and R. Fraiman. A geometrically motivated parametric model in manifold estimation. *Statistics*, 48(5):983–1004, 2014.

- [3] G. Biau, B. Cadre, D.M. Mason, and B. Pelletier. Asymptotic normality in density support estimation. *Electronic Journal of Probability*, 14:2617–2635, 2009.
- [4] G. Biau, B. Cadre, and B. Pelletier. Exact rates in density support estimation. *J. Mult. Anal.*, 99:2185–2207, 2008.
- [5] P. Calka and N. Chenavier. Extreme values for characteristic radii of a Poisson-Voronoi tessellation. *Extremes*, 17(3):359–385, 2014.
- [6] A. Cuevas. On pattern analysis in the non-convex case. *Kybernetes*, 19:26–33, 1990.
- [7] A. Cuevas, R. Fraiman, and B. Pateiro-Lopez. On statistical properties of sets fulfilling rolling-type conditions. *Adv. Appl. Prob.*, 44(2):311–329, 2012.
- [8] A. Cuevas, R. Fraiman, and A. Rodriguez-Casal. A nonparametric approach to the estimation of lengths and surface areas. *The Annals of Statistics*, 35(3):1031–1051, 2007.
- [9] A. Cuevas and A. Rodriguez-Casal. On boundary estimation. *Adv. Appl. Prob.*, 36:340–354, 2004.
- [10] L. Devroye and G. Wise. Detection of abnormal behaviour via non-parametric estimation of the support. *SIAM J. Appl. Math.*, 3:480–488, 1980.
- [11] K. J. Falconer. *The Geometry of Fractal Sets*. Cambridge University Press, 1985.
- [12] D. Gatzouras. Lacunarity of self-similar and stochastically self-similar sets. *Trans. AMS*, 352(5):1953–1983, 2000.
- [13] M. Heveling and M. Reitzner. Poisson-Voronoi approximation. *The Annals of Applied Probability*, 19(2):719–736, 2009.
- [14] R. Jimenez and J. E. Yukich. Nonparametric estimation of surface integrals. *The Annals of Statistics*, 39(1):232–260, 2011.
- [15] W. S. Kendall and I. Molchanov. *New perspectives in stochastic geometry*. Oxford University Press, 2010.

- [16] E. Khmaladze and N. Toronjadze. On the almost sure coverage property of Voronoi tessellation. *Advances in Applied Probability*, 33(4):756–764, 2001.
- [17] R. Lachièze-Rey and G. Peccati. New Kolmogorov bounds for geometric functionals of binomial point processes. arXiv:1505.04640.
- [18] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces*. Cambridge University Press, 1995.
- [19] M. D. Penrose. Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli*, 13(4):1124–1150, 2007.
- [20] M. Reitzner, Y. Spodarev, and D. Zaporozhets. Set reconstruction by Voronoi cells. *Advances in Applied Probability*, 44(4):938–953, 2012.
- [21] A. Rodriguez-Casal. Set estimation under convexity-type assumptions. *Ann. Inst. H. Poincaré Prob. Stat.*, 43:763–774, 2007.
- [22] A. Schief. Separation properties for self-similar sets. *Proc. Amer. Math. Soc.*, 122(1):111–115, 1994.
- [23] M. Schulte. A central limit theorem for the Poisson-Voronoi approximation. *Advances in Applied Mathematics*, 49(3-5):285–306, 2012.
- [24] G. Walther. Granulometric smoothing. *Ann. Statist.*, 25:2273–2299, 1997.
- [25] G. Walther. On a generalisation of Blaschke’s Rolling Theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.*, 22:301–316, 1999.
- [26] J. E. Yukich. Surface order scaling in stochastic geometry. *Ann. Appl. Probab.*, 25(1):177–210, 2015.