



**HAL**  
open science

## 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. 2015. hal-01105205v2

**HAL Id: hal-01105205**

**<https://hal.science/hal-01105205v2>**

Preprint submitted on 21 Jan 2015 (v2), last revised 23 Nov 2015 (v3)

**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 †

January 21, 2015

## Contents

<b>1 Self-similar sets</b>	<b>4</b>
1.1 Self-similar set theory . . . . .	4
1.2 Minkowski contents of self-similar sets . . . . .	6
1.3 Boundary regularity . . . . .	8
<b>2 Voronoi approximation</b>	<b>11</b>
2.1 Asymptotic normality . . . . .	12
2.2 Convergence for the Hausdorff distance . . . . .	14
2.3 A counter-example . . . . .	18

## 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 theory, 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-Essen 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

---

\*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

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

**Notations** In all the following,  $d(., .)$  designates the Euclidean distance between points or subsets of  $\mathbb{R}^d$ . The closure, the interior, the topological boundary and the diameter of a set  $E \subset \mathbb{R}^d$  are designated by  $\text{cl}(E)$ ,  $\text{int}(E)$ ,  $\partial K$ ,  $\text{diam}(E)$  respectively.  $\text{Vol}$  is the  $d$ -dimensional Lebesgue measure and  $\kappa_d$  is the volume of the Euclidean unit ball.  $B(x, r)$  is the open ball with center  $x \in \mathbb{R}^d$  and radius  $r \geq 0$ .

## 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 [5, 15] or a Poisson point process [10, 11, 16]. 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 [12, 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 [16, 19],  $r$ -convex [6, 17], to have volume polynomial expansion [2], positive reach, or a  $(d-1)$ -rectifiable boundary [11]. Another class of regularity assumptions usually needed is that of *sliding ball* or *rolling ball* conditions. 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  $\chi \cap K^c$ . An advantage of this model is that  $K_\chi$  will be less likely to contain holes inside the set  $K$  due to the possible sparsity of points in that region, and the corresponding volume estimator has the remarkable property to be unbiased with Poisson input, see [16]. We also have the handy property  $(K^c)_\chi = (K_\chi)^c$ . A more formal presentation of Voronoi approximation along with a summary of existing results ([3, 10, 11, 16, 19]) is given in Section 2.

## Approach and main results

This work is motivated by the results of [13], in which it was proved that if  $\partial K$  has Minkowski dimension  $s$ , then under sufficient assumptions regarding  $K$ 's boundary (precised in Section

2.1) the following variance asymptotics hold

$$0 < \liminf_n \mathbf{Var}(\text{Vol}(K_{\chi_n}))n^{s/d-2} \leq \limsup_n \mathbf{Var}(\text{Vol}(K_{\chi_n}))n^{s/d-2} < \infty, \quad (1)$$

where  $\chi_n$  is a set of  $n$  IID variables uniformly distributed in a cube surrounding  $K$ . An estimation of the Kolmogorov distance between the renormalized approximation volume and the standard Gaussian law was also given. For  $\varepsilon > 0$ , there is  $C_\varepsilon > 0$  such that for  $n \geq 1$

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

where  $N$  is a standard normal variable. In order to study the extent to which this result applies to sets with a possible irregular boundary, we focus here on sets with a self-similar boundary. Interactions between fractal sets and Voronoi tessellation with random input have also been studied in some recent works, but in other contexts, see for instance [4] and [20]. Surprisingly, it turns out that second order results can still be obtained after dropping all convexity and smoothness assumptions, allowing sets with a fractal boundary, such as the Von Koch flake, and the rolling ball assumption has to be replaced by a much weaker assumption. We are able in this setting to obtain the power law decay of (2), as well as a.s. convergence results in the Hausdorff distance.

The central condition imposed on  $K$ , presented here as Assumption 2, is concerned with the density of  $K$  at radius  $r$  in  $x$ , defined by

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

Assumption 2 holds whenever  $f_r^K > \varepsilon$  or  $f_r^{K^c} > \varepsilon$  on  $\partial K$ , for some fixed  $\varepsilon > 0$  and small enough  $r$ . This is not the general setting, but it gives the gist of what Assumption 2 imposes on  $K$ , and the inequality  $f_r^K, f_r^{K^c} > \varepsilon$  has already appeared in set estimation ([5, Theorem 1], [3]). Assumption 2 can be seen as a drastically weakened form of rolling ball type conditions, in the sense that it requests  $B(x, r) \cap K$  or  $B(x, r) \cap K^c$  to contain a certain volume, whereas rolling ball conditions require them to contain a certain rigid shape to touch the boundary. When it exists, the density of  $K$  in  $x$  is the limit of  $f_r(x)$  as  $r$  goes to 0. The study of these densities on the boundary is also related with works in geometric measure theory. Points with density 0 and 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] proved that if  $K$  is a measurable set with finite measure-theoretic perimeter then most of the essential boundary's points have density 1/2, see [1, Th.3.60].

The question we address here regarding this issue is whether this assumption still holds if  $\partial K$  is irregular, focusing on sets with self-similar boundary. In general, such boundaries have a Hausdorff dimension  $s > d - 1$  and don't have finite perimeter. But, because of self-similarity, the densities  $f_r^K, f_r^{K^c}$  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  $f_r^K > \varepsilon$  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 (2). Remark that if this assertion holds for  $K$ , it does not automatically hold for  $K^c$ , even though they share the same boundary. We actually give conditions ensuring that both  $\inf_{r < 1} f_r^K > 0$  and  $\inf_{r < 1} f_r^{K^c} > 0$ .

However, 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 weak rolling ball condition ensuring (1) and (2). Simulations we ran suggest that this irregularity of  $K_{\text{cantor}}$ 's boundary indeed reflects on the behaviour of its Voronoi approximation and prevents it from satisfying (1).

It is remarkable that the densities  $f_r^K$  and  $f_r^{K^c}$  are also crucial when one studies the quality of the approximation with regard to the Hausdorff distance  $d_H$ . This quantity seems less stable than the volume of the approximation, where compensation mechanisms might occur around the boundary of  $K$ . The problem of assessing the Hausdorff distance between  $K$  and  $K^\chi$  had been raised in [10], and a first result was obtained by Calka and Chenavier [3]. We complete their findings by showing that, under positive inner and outer densities for  $K$ , there are explicit constants  $c_K, c'_K > 0$  such that almost surely

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

where  $\chi_n$  is a Poisson point process with intensity  $n \in \mathbb{N}$ . The upper bound also holds for a binomial point process.

## 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 self-similar boundaries have positive inner and outer densities. Voronoi approximation is formally introduced in Section 2. We then derive the volume normal approximation for sets with well-behaved self-similar boundaries and more general 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 (1).

## 1 Self-similar sets

### 1.1 Self-similar set theory

We start with some brief reminders of self-similar set theory. A precise treatment of the subject can be found in [8]. We recall the definition of the Hausdorff distance between two

sets  $A, B \subset \mathbb{R}^d$ ,

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

Let  $\{\phi_i, i \in I\}$  be an iterated function system, i.e a finite set of contracting similitudes. We define the following set transformation

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

$\psi$  is easily seen to be a contracting transformation 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 a unique set  $E \in \mathcal{K}^d$  satisfying  $\psi(E) = E$ , who 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 [18] that we can pick  $U$  so that  $U \cap E$  is not empty. We will always do so here.

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 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 (one can find an alternative proof for the lower content in [9, Paragraph 2.4], it can also be considered a consequence of  $\mathcal{H}^s(E) > 0$ , like suggested in [14]). We will need the following classical lemmae, that we prove for completeness.

**Lemma 1.** *Let  $(U_i)$  be a collection of disjoint open sets in  $\mathbb{R}^d$  such as 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  $E$  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 as*

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 as follows

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

## 1.2 Minkowski contents of self-similar sets

We 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. That leaves no ambiguity on the choice of  $s$ , since in that case  $s$  is necessarily the Minkowski dimension of  $E$ , i.e

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

**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, i.e*

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

*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  $\Phi_k$  we get

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

for some positive  $c', C$  independent from  $r$  (the exact value of the constants doesn't matter here).

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

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

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  $r c_2 \text{diam}(U) \geq \Lambda_k c_2$ . 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 set  $K^{\text{PROP}}$  of *proper points* of  $K$  as the support of  $\text{Vol}(K \cap \cdot)$ . Points of  $K \setminus K^{\text{PROP}}$  are said *unproper* to  $K$ . This notion must not to be mistaken with the *essential points* mentioned in the introduction. Further use of proper points will be made in Section 2.2. We can already note that  $K$  must have no unproper points if we want a positive lower bound for the  $f_r^K$  on  $K$ . Our main result holds for self-similar subsets  $E$  of  $\partial K$  satisfying the following assumption:

**Assumption 1.**  *$E$  satisfies the open set condition with some set  $U$  (with  $U \cap E \neq \emptyset$ ) 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 the space into infinitely many connected components, then because of self-similarity it also does so locally, and  $K$  and  $K^c$  are too disconnected for a ball to be rolled inside them. Example 2 will show that these concerns are legitimate.

**Theorem 1.** *Let  $K$  be a non-empty compact set with no unproper points and  $\text{Vol}(\partial K) = 0$ . Let  $E$  be a self-similar subset of  $\partial K$  for which Assumption 1 holds. Then  $K$  has a rolling ball along  $E$ , i.e there are constants  $\delta, \varepsilon > 0$  such that, for all  $x \in E, r < \delta$ ,  $B(x, r) \cap K^c$  and  $B(x, r) \cap K$  both 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. The  $V_j$  are the connected components of  $U \setminus E$ . Since there are finitely many of them, we can suppose they all contain a ball of radius  $\tau > 0$ . Suppose  $\text{diam}(U) = \text{diam}(\text{cl}(U)) = 1$ , pick any  $0 < r < 1$  and  $x \in E$ .

Lemma 2 shows that there is a similarity  $\Phi$  such that  $\min(\lambda_i)r \leq \text{diam}(\Phi(U)) < r$  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 as  $x' \in \Phi'(E)$  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 unproper 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  $\text{diam}(\Phi(U))\tau$ . Since  $\Phi(U) \geq \min(\lambda_i)r$ , the rolling ball condition holds with  $\varepsilon = \min(\lambda_i)\tau$ .  $\square$

**Remark 2.** As we pointed out in the introduction, this implies that  $K, K^c$  have lower density bounds on  $E$ . More precisely, for appropriate  $\delta, \varepsilon > 0$

$$\forall x \in E, r < \delta, \quad f_r^K(x), f_r^{K^c}(x) \geq \varepsilon. \quad (3)$$

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

We show below that the Von Koch flake satisfies 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, who satisfy the open set condition with  $U$  the interior of the triangle  $A_0A_2A_4$ . Such sets  $E$  are called Von Koch curves. Looking at the iterates  $\psi^{(n)}(S)$  in Figure 2 gives an idea of the general form of the Von Koch curve and of why it is said to be self-similar.

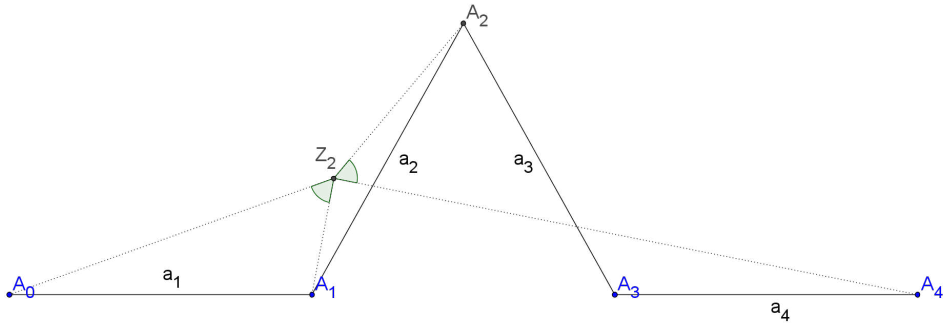


Figure 1: The generating similitudes of the Van Koch curve.  $Z_2$  is the center of the similarity  $\phi_2$ .

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  $\mathbb{R}^2$ . Then their limit  $\gamma$  is also a continuous mapping,  $\gamma([0, 1])$  is compact and has distance 0 with  $E$  in the Hausdorff metric, so  $\gamma([0, 1]) = E$ . This proves that the Van Koch curve



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

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

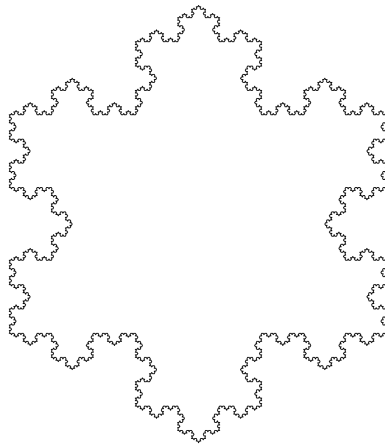


Figure 3: The boundary of the Von Koch flake  $K$ .

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 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.

Consequently, Theorem 1 can be applied three times to obtain a lower bound for  $f_r$  on  $\partial K$ .

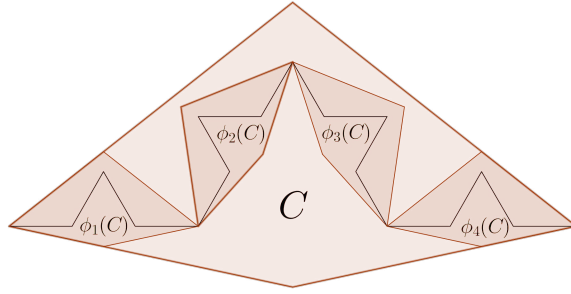


Figure 4: Assumption 1 is satisfied with the kite  $C$ .

## 2 Voronoi approximation

In this paper,  $\chi$  is a locally finite point process. 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  $\chi = \chi'_n$  is a homogenous Poisson point process of intensity  $n$  we speak of Poisson input. We also suppose from now on that  $K$  is a subset of  $[0, 1]^d$ .

On the unit cube, define the Voronoi cell  $v_\chi(x)$  of nucleus  $x$  with respect to  $\chi$  as the set of points closer to  $x$  than to  $\chi$

$$v_\chi(x) = \{y \in [0, 1]^d : \forall x' \in \chi, d(x, y) \leq d(x', y)\}$$

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

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

$\text{Vol}(K_{\chi_n})$  can be given as a consistent estimator for  $\text{Vol}(K)$ , it converges almost surely for binomial input as proved by Penrose in [15] as  $n \rightarrow \infty$ . The volume  $\varphi(\chi) = \text{Vol}(K^\chi)$  first arised in Einmahl and Khmaladze [7] as a discriminating statistic in the two-sample problem. These authors proved a strong law of large numbers in dimension 1 for the volume approximation. Reitzner and Heveling [10] proved that if  $K$  is convex and compact in  $\mathbb{R}^d$  and  $\chi = \chi_\lambda$  is a homogeneous Poisson process with intensity  $\lambda$ ,  $\mathbf{E}\varphi(\chi) = \text{Vol}(K)$ , and  $\mathbf{Var}(\varphi(\chi)) \leq c\lambda^{-1-1/d}S(K)$  where  $c$  is an explicit constant and  $S(K)$  is the surface area of  $K$ . They also established that  $\mathbf{E}\varphi_{\text{Per}}(\chi) = c'\lambda^{-1/d}S(K)(1 + O(\lambda^{-1/d}))$  and  $\mathbf{Var}(\varphi_{\text{Per}}(\chi)) \leq c'\lambda^{-1-1/d}S(K)$ , where  $\varphi_{\text{Per}} = \text{Vol}(K \Delta K^\chi)$  aims at estimating the perimeter of the set. This last statistic echoes the surface estimation problems arising in set estimation, or boundary estimation, see for instance [12, Section 11.2.1]. Reitzner, Spodarev and Zaporozhets [16] 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 [19] proved a similar lower bound for the variance, i.e.  $CS(K)\lambda^{-1-1/d} \leq \mathbf{Var}(\varphi(\chi))$  with  $K$  a convex body and  $C$  a universal constant, and the

corresponding CLT

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

For Binomial input, Penrose proved the remarkable fact that for  $\chi_n$  consisting in  $n$  iid variables with density  $\kappa(x) > 0$  on  $[0, 1]^d$ ,

$$\mathbf{E}\varphi(\chi_n) \rightarrow \text{Vol}(K),$$

independently of any assumption on  $K$ 's boundary.

In this Section, we give an upper bound on Kolmogorov distance between the renormalized approximation volume and the normal law, and an a.s. convergence result for the Hausdorff distance between  $K^\chi$  and  $K$ , which also holds for Poisson input.

## 2.1 Asymptotic normality

We recall below the results of [13], conditions on  $K$  that ensure that with binomial input, the volume approximation is asymptotically normal when the number of points tends to  $\infty$ . We furthermore give the variance magnitude and an upper bound on the speed of convergence for the Kolmogorov distance. The regularity of the boundary is essential to have a matching lower bound on the variance and a good rate of convergence, but it is still possible to have a bound in the case where the set does not satisfy the conditions below. For all  $r > 0$  define

$$\begin{aligned} \partial K_r &= \partial K + B(0, r) = \{x \in \mathbb{R}^d : d(x, K) \leq r\}, \\ \partial K_r^+ &= \partial K_r \cap K^c, \\ \partial K_r^- &= \partial K_r \cap K. \end{aligned}$$

We now explicitly state the boundary regularity assumption made on  $K$ . As explained in the introduction, it can be seen as a weakened form of the rolling ball condition.

**Assumption 2** (Weak rolling ball condition). *A set  $K$  with no unproper points satisfies the weak rolling ball condition whenever*

$$\liminf_{r>0} \frac{1}{\text{Vol}(\partial K_r)} \left( \int_{\partial K_r^+} (f_r^K(x))^2 dx + \int_{\partial K_r^-} (f_r^{K^c}(x))^2 dx \right) > 0. \quad (4)$$

*If  $\partial K$  has upper and lower Minkowski contents, this last assertion is equivalent to the apparently weaker one*

$$\liminf_{r>0} \frac{1}{\text{Vol}(\partial K_r)} \left( \int_{\partial K_r^+} (f_{C_r}^K(x))^2 dx + \int_{\partial K_r^-} (f_{C_r}^{K^c}(x))^2 dx \right) > 0 \text{ for some positive } C. \quad (5)$$

*Proof.* The first condition obviously implies the second. Now suppose the second condition is satisfied for some  $C > 0$ . If  $C < 1$  then the inequality  $f_r^K \geq f_{Cr}^K C^{-d}$  and its counterpart for  $f_r^{K^c}$  show that (4) holds. If  $C > 1$  we can replace  $\partial K_r$  by  $\partial K_{Cr}$  in the domains of the integral, then divide by  $\text{Vol}(\partial K_{Cr})$  instead of  $\text{Vol}(\partial K_r)$  and put  $r' = Cr$ . We're back to (4) and the  $\liminf$  is still be positive, since the first operation only made the integrals bigger, and  $\partial K$ 's Minkowski contents put a lower bound on  $\frac{\text{Vol}(\partial K_r)}{\text{Vol}(\partial K_{Cr})}$ .  $\square$

If the lower density bounds of (3) hold, then (4) holds as well with the left hand being greater than  $\varepsilon^2$ . We can now reproduce below the result derived in [13, Th. 6.1] for Voronoi approximation.

**Theorem 2.** *Let  $K$  be a subset of  $[0, 1]^d$ . Assume that for some  $s \geq 0$*

$$0 < \liminf_{r>0} r^{s-d} \text{Vol}(\partial K^r) \leq \limsup_{r>0} r^{s-d} \text{Vol}(\partial K^r) < \infty, \quad (6)$$

*and that  $K$  satisfies the weak rolling ball condition (Assumption 2), then*

$$0 < \liminf_{r>0} \frac{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}{n^{-2+s/d}} \leq \limsup_{r>0} \frac{\mathbf{Var}(\text{Vol}(K_{\chi_n}))}{n^{-2+s/d}} < \infty, \quad (7)$$

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

$$\sup_{t \in \mathbb{R}} \left| \mathbf{P} \left( \frac{\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})}{\mathbf{Var}(\text{Vol}(K_{\chi_n}))} \leq t \right) - \mathbf{P}(N \leq t) \right| \leq C_\varepsilon n^{-s/2d+\varepsilon}. \quad (8)$$

The consequences of Theorems 1 and 2 for sets  $K$  with self-similar boundary are immediate, condition (6) automatically holds by Proposition 1.

**Remark 1.** If  $K$  satisfies all the hypotheses of Theorem 2 except the weak rolling ball condition, then we have

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left| \mathbf{P} \left( \frac{\text{Vol}(K_{\chi_n}) - \mathbf{E}\text{Vol}(K_{\chi_n})}{\mathbf{Var}(\text{Vol}(K_{\chi_n}))} \leq t \right) - \mathbf{P}(N \leq t) \right| \\ \leq C_\varepsilon n^\varepsilon (\sigma^{-2} n^{-2+s/2d} + \sigma^{-3} n^{-3+s/d} + \sigma^{-4} n^{-4+s/d}) \end{aligned} \quad (9)$$

where  $\sigma^2$  is the variance of  $\text{Vol}(K_{\chi_n})$ . See [13, Th. 6.2] for more details.

**Remark 2.** Set-estimation literature is also concerned with perimeter approximation [12, Sec. 11.2.1]. In the context of Voronoi approximation, the functional  $\varphi(\chi_n) = \text{Vol}(K_{\chi_n} \Delta K)$  is studied in [10, 16]. Although the result is not formally stated, a bound of the form (9) for that functional is available using the exact same method. One has to work separately to obtain a variance lower bound.

**Corollary 1.** *Let  $K$  be a compact set such that  $\partial K$  is a finite union of self-similar sets satisfying Assumption 1. Then (7) and (8) hold.*

This corollary applies to the Von Koch flake with  $s = \ln(4)/\ln(3)$  (Example 1). The conclusions of Theorem 2 also apply for instance to the Von Koch anti flake, where three Von Koch curves are stuck together like for building the flake, but here the curves are pointing inwards, and not outwards (Figure 5). 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 (3) on a self-similar  $E$  with same Minkowski dimension as  $\partial K$  is actually enough for the weak rolling ball condition to hold.

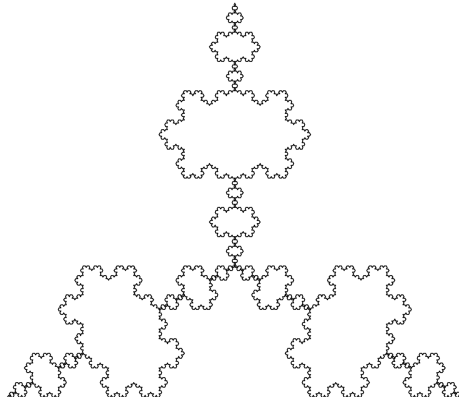


Figure 5: The Von Koch anti-flake

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 also run simulations suggesting that (7) 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 section we will make use of  $r$ -coverings and  $r$ -packings. Let  $\mathcal{B}$  be a collection of balls having radius  $r$  and centers belonging to some set  $E$ .  $\mathcal{B}$  is said to be an  $r$ -packing of  $E$  if the balls are disjoint. It is an  $r$ -covering if the balls cover  $E$ .

Coverings and packings are closely related to the Minkowski dimension of  $E$ . In particular, if  $E$  has upper and lower Minkowski contents, then for  $r$  small enough we can find  $r$ -coverings of  $E$  and  $E + B(0, r)$  with less than  $Cr^{-s}$  balls, and  $r$ -packings of the same sets with more than  $cr^{-s}$  balls, where  $c$  and  $C$  are positive constants not depending on  $r$ . More precise results can be found in [14]. The following geometrical lemma shows a connection between coverings, packings, and the Hausdorff distance with the Voronoi approximation.

**Lemma 3.** *Let  $K$  be a non-empty compact set and  $\partial K$  its topological boundary.*

1. Consider a  $r/2$ -covering  $\mathcal{B}$  of  $\partial K_r$  containing an  $r/2$ -covering of  $\partial K_r^-$ . If the interior of every ball from  $\mathcal{B}$  contains a point from  $\chi$ , and every ball from  $\mathcal{B}$  centered on  $K$  contains a point from  $\chi \cap K$ , then  $d_H(K, K_\chi) \leq r$ .
2. Consider a  $3r$ -packing  $\mathcal{B}$  of  $\partial K$ . If for some ball  $B(x, 3r)$  of  $\mathcal{B}$  centered on  $x \in \partial K$  we have  $\chi \cap B(x, 3r) \cap K = \emptyset$  and  $\chi \cap B(x, r) \cap K^c \neq \emptyset$ , then  $d_H(K, K_\chi) \geq r$ .

*Proof.* We begin with the first point. Let us prove that for all  $x \in K \setminus \partial K_r$  we have  $x \in K_\chi$ . Indeed, if this were not the case, there would be a point  $c_x \in \chi \cap K^c$  such that  $x \in v_\chi(c_x)$ . The segment joining  $c_x$  and  $x$  must contain a point from  $\partial K$ . Let  $x_0$  be the point of  $\partial K_r$  closest to  $x$  on that segment. We must have  $d(x_0, \partial K) = r$  and  $x_0 \in K$  since otherwise there would be another point of  $\partial K_r$  closer to  $x$ . As a consequence  $d(x_0, c_x) > r$ . But then by hypothesis there is a ball of  $\mathcal{B}$  who contains  $x_0$  along with a point of  $\chi$ . So  $c_x$  isn't the point of  $\chi$  closest to  $x$ , and we have a contradiction. Similarly, using  $(K_\chi)^c = (K^c)_\chi$ , we can show that  $K^c \setminus \partial K_r \subset K_\chi^c$ , which reformulates as  $K_\chi \subset K + B(0, r)$ .

To have  $d_H(K, K_\chi) \leq r$  it is enough to show that  $K \subset K_\chi + B(0, r)$ . Let  $x$  be a point of  $K$ . We just showed that if  $x \notin \partial K_r$  then  $x \in K_\chi$ . And if  $x \in \partial K_r$ , then by hypothesis there is a ball of  $\mathcal{B}$  centered on  $\partial K_r^- \subset K$  with a point of  $K \cap \chi$  inside that contains  $x$ .

In all cases  $x \in K_\chi + B(0, r)$ .

Now we prove the second point. 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  $x \notin K_\chi + r'$  for all  $r' \leq r$ , so that  $d_H(K, K_\chi) \geq r$ .  $\square$

To formulate results regarding the Hausdorff distance between  $K$  and  $K_\chi$ , the concept of proper points (beginning of Section 1.3) proves to be useful. Unproper points are 'forgotten' by the Voronoi approximation  $K_\chi$  of  $K$ . Though that has no incidence when measuring volumes, it becomes a nuisance when measuring Hausdorff distances.

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

**Proposition 2.**  $K_{\chi_n} \xrightarrow[n \rightarrow +\infty]{} K^{\text{prop}}$  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 unproper points, this is equivalent to the fact that  $K_{\chi_n} \rightarrow K$  almost surely when  $K$  has no unproper points. By the Borel-Cantelli lemma it is enough to show the series  $\sum_{n \geq 1} \mathbf{P}(d_H(K_{\chi_n}, K) > r)$  is convergent for any positive  $r$ .

Cover  $\partial K_r$  as in point 1 of the previous lemma. Since  $K$  is bounded, this can be done with finitely many balls. Let  $V$  be the minimum of  $\text{Vol}(K \cap B)$  over the  $B \in \mathcal{B}$  centered on



$K$ . Because  $K$  has no unproper points,  $V > 0$ . The probability of having  $d_H(K_{\chi_n}, K) > r$  is at most that of the requirements of point 1 not being satisfied. The latter 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  $\mathbf{P}(d_H(K_{\chi_n}, K) > r)$  converges, 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  $\partial K$ ,  $f_r^K$  and  $f_r^{K^c}$  resembling those of Theorem 2. Before we state such a result we need a lemma linking the behaviour of  $f_r^K$  on  $\partial K$  to its behaviour on all of  $K$ .

For ease of notation, we put  $f_r = f_r^K$  and  $g_r = f_r^{K^c}$ .

**Lemma 4.** *If for all  $r < \delta$  we have  $f_r \geq \varepsilon$  on  $\partial K$ , then for all  $r < \delta$  we have  $f_r \geq \frac{\varepsilon}{2^d}$  on  $K$ .*

*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 not then the ball  $B(x, r/2)$  is contained in  $K$  so that  $f_r \geq 2^{-d} \geq \varepsilon 2^{-d}$ .  $\square$

**Theorem 3.** *Suppose that  $\partial K$  has Minkowski dimension  $s > 0$  with upper and lower contents, and that for all  $r$  small enough and  $x \in \partial K$ ,*

$$\begin{aligned} f_r(x) &\geq \varepsilon_f, \\ g_r(x) &\geq \varepsilon_g, \end{aligned}$$

then

$$\alpha \leq \liminf_{\lambda \rightarrow +\infty} \frac{d_H(K, K_{\chi_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \limsup_{\lambda \rightarrow +\infty} \frac{d_H(K, K_{\chi_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta$$

where the  $\chi_\lambda, \lambda \in \mathbb{N}$  are Poisson point processes of intensity  $\lambda$  and

$$\begin{aligned} \alpha &= \frac{1}{3} \left( \frac{s}{d\kappa_d(1 - \varepsilon_g)} \right)^{1/d} \\ \beta &= 4 \left( \frac{s + d}{d\kappa_d\varepsilon_f} \right)^{1/d}. \end{aligned}$$

*Proof.* We start with the upper bound. Fix  $\delta > 0$  and for all  $\lambda$  let  $A_\lambda$  be the event where all the requirements from point 1 of Lemma 3 are met with  $\chi = \chi_\lambda$ ,  $r = r_\lambda = (\beta + \delta)(\lambda^{-1} \ln(\lambda))^{1/d}$ ,  $\mathcal{B} = \mathcal{B}_\lambda$  a deterministic covering having  $\mathcal{N}_\lambda = O(r_\lambda^{-s})$  balls of radius  $r_\lambda/2$ .

We show below that  $\lim_{\lambda \rightarrow +\infty} \mathbf{P}(A_\lambda^c) \ll \lambda^{-1-\delta'}$  for some  $\delta' > 0$ . Since  $A_\lambda \subset \{d_H(K, K_\chi) \leq r_\lambda\}$ , it follows from the Borel-Cantelli lemma that

$$\mathbf{P}(\limsup \frac{d_H(K, K_{\chi_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta + \delta) = 1.$$

The desired conclusion comes from the arbitrariness of  $\delta$  and the equality

$$\mathbf{P}(\limsup \frac{d_H(K, K_{\chi_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta) = \lim_{n \rightarrow +\infty} \mathbf{P}(\limsup \frac{d_H(K, K_{\chi_\lambda})}{(\lambda^{-1} \ln(\lambda))^{1/d}} \leq \beta + n^{-1}).$$

The intersection of  $K$  with a ball  $B \in \mathcal{B}_\lambda$  of center  $x$  has volume  $\kappa_d(r_\lambda/2)^d(f_{r_\lambda/2}(x))$ . If  $x$  belongs to  $K$ , it follows from Lemma 4 that for  $\lambda$  large enough the probability of  $B$  not having a point from  $\chi_\lambda \cap K$  is at most

$$\exp(-\lambda \varepsilon_f \kappa_d 4^{-d} r_\lambda^d) = \lambda^{-s/d-1-\delta'}$$

for some  $\delta' > 0$ .

This bound also works for the probability that a ball of  $\mathcal{B}_\lambda$  does not contain a point from  $\chi_\lambda$ , no matter the location of its center, so we have

$$\mathbf{P}(A_\lambda^c) \leq \mathcal{N}_\lambda \lambda^{-s/d-1-\delta'}$$

with  $\delta' > 0$  The right hand term has order  $\ln(\lambda)^{-s/d} \lambda^{-1-\delta'}$  as desired.

The proof for the lower bound is quite similar. Fix  $\delta > 0$ , and redefine  $A_\lambda$  to be the event where the requirements described in point 2 of Lemma 3 are met for  $\chi = \chi_\lambda$ ,  $r = r_\lambda = (\alpha - \delta)(\ln(\lambda)\lambda^{-1})^{1/d}$  and  $\mathcal{B} = \mathcal{B}_\lambda$  a collection of  $\mathcal{N}_\lambda \geq cr_\lambda^{-s}$  balls having radius  $3r_\lambda$ .

We have  $A_\lambda \subset \{d_H(K, K_\chi) \geq r_\lambda\}$ . Now we show that  $\mathbf{P}(A_\lambda^c) \ll \lambda^{-2}$  (in fact  $\mathbf{P}(A_\lambda^c)$  has exponential decay). This gives the desired conclusion by applying again the Borel-Cantelli lemma and letting  $\delta$  go to 0.

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

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

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

$$\mathbf{P}(A_\lambda^c) \leq (1 - e^{-\lambda \kappa_d (1 - \varepsilon_g) 3^d r_\lambda^d} (1 - e^{-\lambda \kappa_d r_\lambda^d \varepsilon_g}))^{\mathcal{N}_\lambda}$$

We would like the right hand times  $\lambda^2$  to go to 0 with  $\lambda$ . Taking logarithms this is equivalent to

$$-2 \ln(\lambda) + \mathcal{N}_\lambda \exp(-\lambda \kappa_d (1 - \varepsilon_g) 3^d r_\lambda^d) (1 - \exp(-\lambda \kappa_d r_\lambda^d \varepsilon_g)) \xrightarrow{\lambda \rightarrow +\infty} +\infty$$

Because  $\exp(-\lambda \kappa_d (1 - \varepsilon_g) 3^d r_\lambda^d) = \lambda^{\delta'-s/d}$  with  $\delta' > 0$ , it is indeed the case. □

The proof and the result call for some comments.

**Remark 3.** The Minkowski dimension of  $\partial K$  has no impact on the order of magnitude of  $d_H(K, K_\chi)$ . In fact, if  $\partial K$  doesn't have a Minkowski dimension, we can do the coverings with  $O(r^{-d})$  balls, so the upper bound still holds after replacing  $s$  by  $d$  in the expression of  $\beta$ . This compares with the result given by Calka and Chenavier in [3, Corollary 2]. If  $\partial K$ 's so-called lower Minkowski dimension is positive, given any  $\epsilon > 0$ , we can do the packing with  $r^{-s+\epsilon}$  balls for  $\lambda$  big enough, and also get a lower bound. This is automatically the case when  $d \geq 2$  and  $K$  has positive volume.

**Remark 4.** If  $s = 0$  and  $\partial K$  has Minkowski contents then actually  $\partial K$  has a finite number of points and  $d_H(K, K_{\chi_\lambda})$  has order  $\lambda^{-1/d}$  in the sense that

$$\mathbf{P}(d_H(K, K_{\chi_\lambda})\lambda^{1/d} > t) \leq \text{card}(\partial K)C \exp(-\varepsilon_f(t/4)^d),$$

for some  $C > 0$  depending only on  $d$ , which is enough to guarantee the existence of moments of all orders for  $d_H(K, K_{\chi_\lambda})\lambda^{1/d}$ . If we don't have Minkowski contents the situation might be more delicate.

**Remark 5.** 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 6.** The order of magnitude  $\lambda^{-1/d} \ln(\lambda)^{1/d}$  also appears in [5], for a different kind of set estimation. Specifically, they approximate the length of  $\partial K$  by considering points at distance  $r_\lambda$  or less from  $K \cap \chi$  and  $K^c \cap \chi$ , and their estimator becomes consistent for  $1 \gg r_\lambda \gg \lambda^{-1/d} \ln(\lambda)^{1/d}$ . Seemingly,  $\lambda^{-1/d} \ln(\lambda)^{1/d}$  is a threshold scale under which geometric features of  $\partial K$  can no longer be correctly assessed with a random point process of intensity  $\lambda$ .

**Remark 7.** Bounds for  $d_H(K, K_\chi)(\lambda \ln(\lambda))^{-1/d}$  can be expressed very simply from  $s, \varepsilon_f$  and  $\varepsilon_g$ . For example, the following holds

$$\mathbf{P}\left(\alpha' \leq d\kappa_d \frac{d_H(K, K_{\chi_\lambda})^d}{\lambda^{-1} \ln(\lambda)} \leq \beta'\right) \xrightarrow{\lambda \rightarrow +\infty} 1$$

for  $\alpha' < s(1 - \varepsilon_g)3^{-d}$ ,  $\beta' > s\varepsilon_f 4^d$ . If those bounds are optimal, it would be evidence that the quantities studied in this work, Minkowski contents and boundary densities, affect directly the asymptotic behaviour of  $d_H(K, K_\chi)$ .

## 2.3 A counter-example

Here we construct a set  $K_{\text{cantor}}$  with self-similar boundary not satisfying the weak rolling ball 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 most of points of  $K_{\text{cantor}}$ 's boundary have density 0, 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 unproper points,  $\partial K = E$  and that  $K$  does not satisfy the regularity condition of Theorem 2.

*Proof.* The first assertion is easy to prove. Being segments, the  $A_n$  have no unproper 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  $\text{Vol}(\partial K) = 0$

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

As a consequence  $f_r \leq r$  on  $\partial K_r \setminus X$ , so that

$$\begin{aligned} \int_X \max(f_r^2, g_r^2) &\leq -4r \log_3(r) \\ \int_{\partial K_r^+ \setminus X} f_r^2(x) dx &\leq \text{Vol}(\partial K_r) r^2 \\ \int_{\partial K_r^- \setminus X} g_r^2(x) dx &\leq r^2. \end{aligned}$$

After dividing by  $\text{Vol}(\partial K_r)$ , who has order  $r^{1-\ln(2)/\ln(3)}$ , the sum of the left terms is bigger than the expression inside the limit of (4), and the sum of the right terms goes to 0 with  $r$ .  $\square$

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 6. Looking at Theorem 2, 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 [13] 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  $\partial K$ .

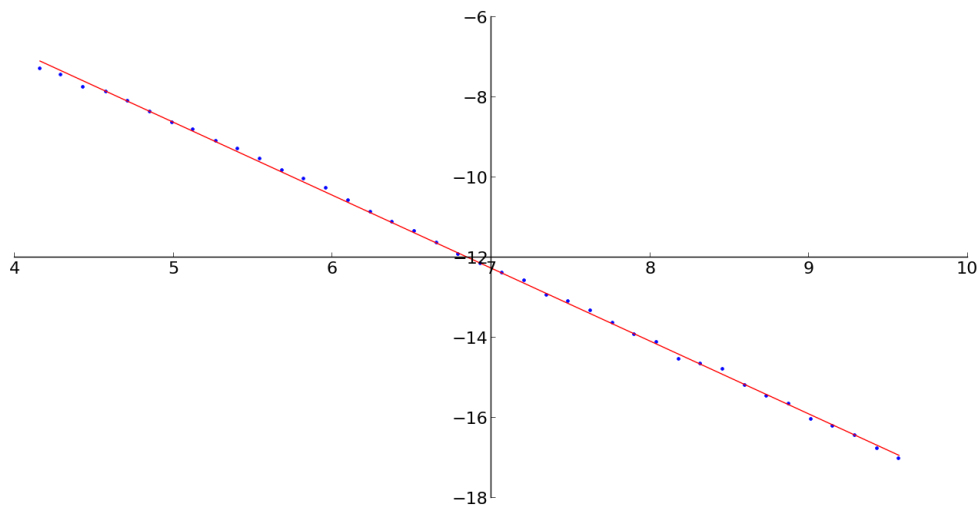


Figure 6: In blue  $\ln(\mathbf{Var}(K_{\chi_n}))$  as a function of  $\ln(n)$ , in red the associated linear regression. For each  $n$ , the variance was estimated with 1000 realisations of  $\text{Vol}(K_{\chi_n})$ .

**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

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

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 Minkowski dimension  $1/2$  with upper and lower contents, no unproper points, and does not satisfy (4) or (5). 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. arXiv:411.3145, 2014.
- [3] P. Calka and N. Chenavier. Extreme values for characteristic radii of a Poisson-Voronoi tessellation. arXiv:1304.0170, 2013.

- [4] P. Calka and Y. Demichel. Fractal random series generated by poisson-voronoi tessellation. to appear in *Trans. AMS*, 2014.
- [5] 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.
- [6] A. Cuevas and A. Rodriguez-Casal. On boundary estimation. *Adv. Appl. Prob.*, 36:340–354, 2004.
- [7] J. H. J Einmahl and E. V. Khmaladze. The two-sample problem in and measure-valued martingales. *Lecture Notes-Monograph Series*, pages 434–463, 2001.
- [8] K. J. Falconer. *The geometry of fractal sets*. Cambridge University Press, 1985.
- [9] D. Gatzouras. Lacunarity of self-similar and stochastically self-similar sets. *Trans. AMS*, 352(5):1953–1983, 2000.
- [10] M. Heveling and M. Reitzner. Poisson-Voronoi approximation. *The Annals of Applied Probability*, pages 719–736, 2009.
- [11] R. Jimenez and J. E. Yukich. Nonparametric estimation of surface integrals. *The Annals of Statistics*, 39(1):232–260, 2011.
- [12] W. S. Kendall and I. Molchanov. *New perspectives in stochastic geometry*. Oxford University Press, 2010.
- [13] R. Lachièze-Rey and G. Peccati. New Kolmogorov bounds for geometric functionals of binomial point processes. preprint, <http://w3.mi.parisdescartes.fr/~rlachiez/recherche/articles/LRPKol.pdf>.
- [14] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces*. Cambridge University Press, 1995.
- [15] M. D. Penrose. Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli*, 13(4):1124–1150, 2007.
- [16] M. Reitzner, Y. Spodarev, and D. Zaporozhets. Set reconstruction by voronoi cells. *Advances in Applied Probability*, 44(4):938–953, 2012.
- [17] A. Rodriguez-Casal. Set estimation under convexity-type assumptions. *Ann. Inst. H. Poincaré Prob. Stat.*, 43:763–774, 2007.
- [18] A. Schief. Separation properties for self-similar sets. *Proc. Amer. Math. Soc.*, 122(1):111–115, 1994.
- [19] Matthias Schulte. A central limit theorem for the poisson-voronoi approximation. *Advances in Applied Mathematics*, 49(3-5):285–306, 2012.

- [20] K. Tchoumatchenko and S. Zuyev. Aggregate and fractal tessellations. *Prob. Th. Rel. Fields*, 121(2):198–218, 2001.