Voronoi set approximation

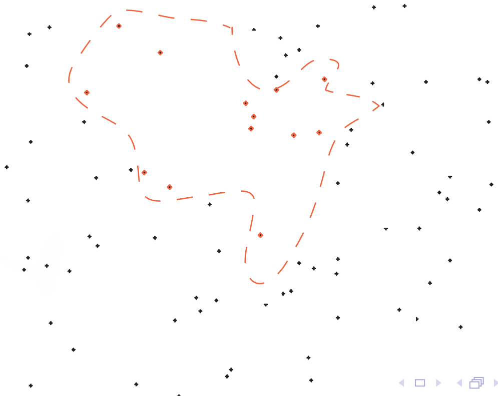
Raphaël Lachièze-Rey, Univ. Paris 5 René Descartes,

In collaboration with Giovanni Peccati, Université de Luxembourg,
and Sergio Vega, Université Paris Descartes

SGSIA 2015, Lingen, Germany

Set estimation

- K is an unknown set in $[0,1]^d$
- $\mathcal{X}$ is a random sample of points
- We have the information $\{1_{x \in K}, x \in \mathcal{X}\} \rightarrow$ Construction of $K^{\mathcal{X}} \sim K$
- How to get a good idea of K ? Measure the quality of the approximation ?



Distance

Several features of interest :

- Volume estimation : $\text{Vol}(K^{\mathcal{X}}) \to \text{Vol}(K)$?
- Shape estimation : $d_{\mathcal{H}}(K^{\mathcal{X}}, K) \to 0$?

$$d_{\mathcal{H}}(X, Y) = \sup\{d(y, X), d(x, Y) : y \in Y, x \in X\}$$

- Perimeter/Minkowski content : $a_{\mathcal{X}}\text{Vol}(K^{\mathcal{X}} \Delta K) \to \text{Per}(K)$ with the right renormalization $a_{\mathcal{X}}$?
- Perimeter shape : $d_{\mathcal{H}}(\partial K^{\mathcal{X}}, \partial K) \to 0$?

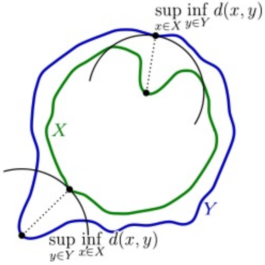


FIGURE: Hausdorff distance

Random input

The random input is in general

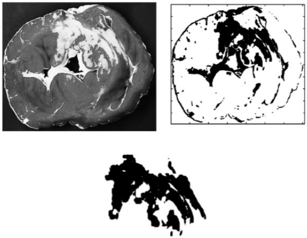
- A family of n IID points $\mathcal{X}_n$ in $[0,1]^d$. The points might sometimes be only assumed to have a density bounded from below.
- A Poisson process P_λ of intensity λ (restricted to $[0,1]^d$)

and $n, \lambda \to \infty$.

Potential applications

- If the shape is the epigraph of a function, estimation of the function values (The random variables are supported by K)
- Perimeter estimation $\rightarrow$ Test regularity hypotheses

$$\frac{\text{length}}{area} \text{ used in oncology/cardiology (Cuevas et. al 2007)}$$



- Jimenez/Yukich 2012 : Evolution of the Aral sea boundary (integral of cliff elevation along the boundary)
- Known image : Approximate volume, perimeter, image compression

Previous ideas

- Support estimation : K is the support of the random points
- K convex : Specific problem studied in a different litterature. The quality much depends on wether K is a polytope or a smooth convex body.
- Devroy and Wise (80's), study the union

$$K_n^{\text{balls}} = \cup_{X_i} B(X_i, \varepsilon_n)$$

where the X_i are IID uniform in K, $\varepsilon_n \to 0, n\varepsilon_n^d \to \infty$. Then

$$\text{Vol}(K_n^{\text{balls}} \Delta K) \to 0$$

in probability.

Regularity assumptions

are required if we want to assess the quality of the approximation.

- Korostelev and Tsybakov (1993),

$$\sup_K \mathbf{E}|\mathrm{Vol}(K_n^{\text{balls}}) - \mathrm{Vol}(K)| \leq C \left(\frac{\log(n)}{n} \right)^{1/d}$$

  for K belonging to a class of sets with Lipschitz boundary.

- r-convexity ($r > 0$) : A ball with radius r can be slided outside the set. Additionally, assume that

$$\inf_{x \in K, 0 < r < 1} \frac{\mathrm{Vol}(K \cap B(x, r))}{\mathrm{Vol}(B(x, r))} > 0$$

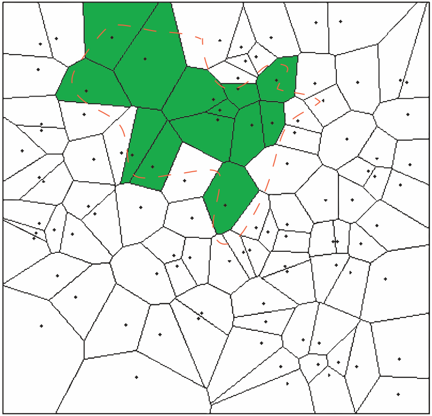
  Then (Cuevas et. al)

$$\mathbf{P}\left(d_{\mathcal{H}}(K_n^{\text{balls}}, K) \leq C \left(\frac{\log(n)}{n} \right)^{1/d} \right) \to 1$$

Voronoi approximation

Reconstruct K with

$$K^{\mathcal{X}} = \{x \in \mathbb{R}^d : x \text{ is closer from a point of } \mathcal{X} \text{ inside } K \text{ than outside } K\}.$$



Several nice features

Define

$$\varphi(\mathcal{X}) = \mathrm{Vol}(K^{\mathcal{X}})$$

- $(K^c)^{\mathcal{X}} = (K^{\mathcal{X}})^c$
- If by "chance" the points leave a large gap inside K, this gap is filled by the Voronoi approximation
- If $\mathcal{X}_\lambda$ is a homogeneous Poisson process on $\mathbb{R}^d$,

  $$\mathbf{E}\varphi(\mathcal{X}_\lambda) = \mathrm{Vol}(K)$$

  whatever is K, λ
- Computationally efficient 500.000 points/minute. in $3D$

Expectations with Poisson input

Theorem

Let K measurable with "finite perimeter", where

$$Per(K) = TV(\mathbb{1}_K) = \sup_{\varphi \in \mathcal{C}_c^1(\mathbb{R}^d), \|\varphi(x)\| \leq 1} \int_K div(\varphi)$$

Then,

$$|(\mathbf{E}\, Vol(K^{\mathcal{P}_\lambda})) - Vol(K)| \leq c_d \lambda^{-1}$$

$$\mathbf{E}\, Vol(K^{\mathcal{P}_\lambda} \Delta K) = c'_d \lambda^{-1/d} Per(K)(1 + O(\lambda^{-1/d}))$$

- Dimension 1 : Khlamadze and Toronjadze (2001) + Law of large numbers.
- K Convex : Heveling and Reitzner (2009)
- K with finite perimeter : Reitzner, Spodarev and Zaporozhets (2011)

Law of large numbers

Theorem (Penrose (2007))

For $K \subseteq (0,1)^d$ mesurable, with probability 1

$$Vol(K^{\mathcal{X}_n}) \to Vol(K)$$
$$Vol(K^{\mathcal{X}_n} \Delta K) \to 0$$

as $n \to \infty$, still holds if the X_i are IID with density $\geq c > 0$ on $[0,1]^d$.

Without any assumption on ∂K !

Variance : Heuristics for binomial input

Theorem (Efron-Stein inequality)

Let φ be a measurable symmetric mesurable functional. For $q \geq 1$

$$Var(\varphi(\mathcal{X}_n)) \leq \frac{n}{2}\mathbf{E}|\underbrace{\varphi(\mathcal{X}_{n+1}) - \varphi(\mathcal{X}_n)}_{\text{Add-one cost}}|^2.$$

- Construct $K^{\mathcal{X}_n}$, then draw X_{n+1}.
- If X_{n+1} is far from ∂K, no variation of $K^{\mathcal{X}_n}$
- The typical diameter of a Voronoi cell is $n^{-1/d}$, and its typical volume is in n^{-1}.
- If ∂K is smooth, there is a probability $n^{-1/d}$ that X_{n+1} is close from the boundary. In this case

$$|\varphi(\chi_{n+1}) - \varphi(\mathcal{X}_n)| \sim \text{ the volume of a Voronoi cell} \sim n^{-1}.$$

Finally

$$\text{Var}(\varphi(\mathcal{X}_n)) \leq \frac{n}{2}n^{-2-1/d} = \frac{1}{2}n^{-1-1/d}$$

Variance estimates (in the Poisson framework)

Theorem (Upper bounds : Heveling et Reitzner '09 - Reitzner, Spodarev et Zaporozhets '11)

$$\mathbf{Var}(Vol(K^{\mathcal{P}_\lambda})) \leq c_d S(K)\lambda^{-1-1/d}$$

$$\mathbf{Var}(Vol(K^{\mathcal{P}_\lambda}\Delta K)) \leq c'_d S(K)\lambda^{-1-1/d}$$

Theorem (Lower bounds : Schulte 2012)

For K convex

$$\mathbf{Var}(Vol(K^{\mathcal{P}_\lambda})) \geq c''_d S(K)n^{-1-1/d}$$

$+CLT$

Kolmogorov Berry-Esseen bounds

Theorem (Yukich '15)

In the case where K is a closed set which boundary is a $\mathcal{C}^2$ orientable sub manifold, or if K is compact and convex,

$$Vol(K^{\mathcal{P}_\lambda}) = \mathbf{E}Vol(K^{\mathcal{P}_\lambda}) + \sigma\lambda^{-\frac{1}{2}-\frac{1}{2d}}\mathcal{N} + O\left(\log(\lambda)^{3d+1}\lambda^{-\frac{1}{2}+\frac{1}{2d}}\right)$$

More formally, with $\Phi(t) = \mathbf{P}(\mathcal{N} \leq t)$ and

$$F_\lambda(t) = \mathbf{P}\left(\frac{Vol(K^{\mathcal{P}_\lambda} - \mathbf{E}Vol(K^{\mathcal{P}_\lambda}))}{\sqrt{\mathbf{Var}(Vol(K^{\mathcal{P}_\lambda}))}} \leq t\right),$$

then

$$\sup_{t\in\mathbb{R}} |F_\lambda(t) - \Phi(t)| \leq \log(\lambda)^{3d+1}\lambda^{-\frac{1}{2}+\frac{1}{2d}}$$

Irregular sets : Minkowski dimension

Assume that

$$x \in \partial K \Leftrightarrow \text{Vol}(B(x,\varepsilon) \cap K^c) > 0, \text{Vol}(B(x,\varepsilon) \cap K) > 0 \text{ for all} \varepsilon > 0$$

(no isolated points) Define

$$\partial K^r = \{x : d(x, \partial K) \leq r\}.$$

∂K has Minkowski dimension $s > 0$ if

$$d - s = \sup\{t : \liminf_r \frac{\text{Vol}(\partial K^r)}{r^t} < \infty\}$$

Variance Heuristics with fractal boundary

Theorem (Efron-Stein inequality)

Let φ be a measurable symmetric mesurable functional. For $q \geq 1$

$$Var(\varphi(\mathcal{X}_n)) \leq \frac{n}{2}\mathbf{E}|\underbrace{\varphi(\mathcal{X}_{n+1}) - \varphi(\mathcal{X}_n)}_{\text{Add-one cost}}|^2.$$

- Construct $K^{\mathcal{X}_n}$ and draw X_{n+1}.
- If X_{n+1} is far from ∂K, no variation of $K^{\mathcal{X}_n}$
- Probability $\sim \mathrm{Vol}(\partial K^{n^{-1/d}})$ that X_{n+1} is at distance $\leq n^{-1/d}$ from the boundary. In this case

$$|\varphi(\mathcal{X}_{n+1}) - \varphi(\mathcal{X}_n)| \leq \text{ Volume of a typical Voronoi cell} \sim n^{-1}.$$

If ∂K has dimension s,

$$\mathrm{Var}(\varphi(\mathcal{X}_n)) \leq \frac{n^{-1}}{2}\mathrm{Vol}(\partial K^{n^{-1/d}}) \leq \frac{1}{2}n^{-2+s/d}$$

Results

If for $s > 0$, $r > 0$,

$$\mathrm{Vol}(\partial K^r) \leq C_K^+ r^{d-s}$$

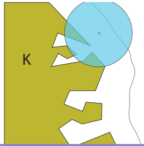
then

$$\mathbf{Var}(\varphi(\mathcal{X}_n)) \leq C_{d,s} C_K^+ n^{-2+s/d}$$

Lower bounds : To have a matching lower bound, we need $c > 0$ such that for $x \in \partial K^{cr}$

$$\text{for } x \notin K,\ \frac{\mathrm{Vol}(B(x,r) \cap K)}{\mathrm{Vol}(B(x,r))} \text{ is large}$$

$$\text{or for } x \in K,\ \frac{\mathrm{Vol}(B(x,r) \cap K^c)}{\mathrm{Vol}(B(x,r))} \text{ is large}$$



Variance magnitude

Define

$$\gamma(K) := \inf_{r>0} \frac{1}{\mathrm{Vol}(\partial K^{cr})} \left(\int_{\partial K^r \cap K^c} \frac{\mathrm{Vol}(B(x,r) \cap K)}{\mathrm{Vol}(B(x,r))} dx + \int_{\partial K^{cr} \cap K} \frac{\mathrm{Vol}(B(x,r) \cap K^c)}{\mathrm{Vol}(B(x,r))} dx \right)$$

Theorem (LR, Peccati)

Let K measurable such that

$$0 < C_K^- \leq \frac{Vol(\partial K^r)}{r^{d-s}} \leq C_K^+ < \infty, r > 0$$

for some $s > 0$. If $\gamma(K) > 0$ for some $c > 0$, then for n sufficiently large

$$c_{d,s} C_K^- \leq \frac{\mathbf{Var}(\varphi(\mathcal{X}_n))}{n^{-2+s/d}} \leq c'_{d,s} C_K^+$$

Strong regularity assumption

The regularity condition is satisfied if for instance for some $c > 0$

$$\inf_{x, 0<r<1} \frac{\mathrm{Vol}(B(x,r) \cap K)}{\mathrm{Vol}(B(x,r))} > 0 \text{ for } x \in K^c, d(x, \partial K) \leq cr$$

$$\inf_{x, 0<r<1} \frac{\mathrm{Vol}(B(x,r) \cap K^c)}{\mathrm{Vol}(B(x,r))} > 0 \text{ for } x \in K, d(x, \partial K) \leq cr$$

Gaussian fluctuations ($\mathcal{N}$: Normal law)

Theorem (LR, Peccati)

Let $\Phi(t)$ the Gaussian distribution function

$$F_n(t) = \mathbf{P}\left(\frac{\varphi(\mathcal{X}_n) - \mathbf{E}\varphi(\mathcal{X}_n)}{\sqrt{\mathbf{Var}(\varphi(\mathcal{X}_n))}} \leq t\right).$$

If K satisfies the conditions in the slide above then for all $\varepsilon > 0$,

$$\sup_{t\in\mathbb{R}} |F_n(t) - \Phi(t)| \leq C_\varepsilon n^{-s/2d} \log(n)^{4-s/d+\varepsilon}$$

if $s = d - 1$, it gives $n^{-(d-1)/2d} \log(n)^{3+1/d+\varepsilon}$compares to Yukich, where the bound is in $\lambda^{-(d-1)/2d} \log(\lambda)^{3d+1}$. The more the set is irregular, the more the convergence seems to be fast (but the variance is larger).

Self-similar set

A set E is self-similar if there are contracting similitudes Φ_i such that

$$E = \psi(E) := \bigcup_i \Phi_i(E).$$

In this case s is the solution of

$$\sum_i n_i^s = 1.$$

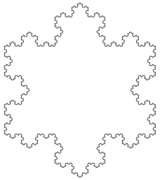


FIGURE: Von Koch flake : $s = \ln(4)/\ln(3)$

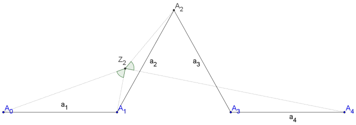


FIGURE: The generating similitudes of the Van Koch curve. Z_2 is the center of the similarity ϕ_2.

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

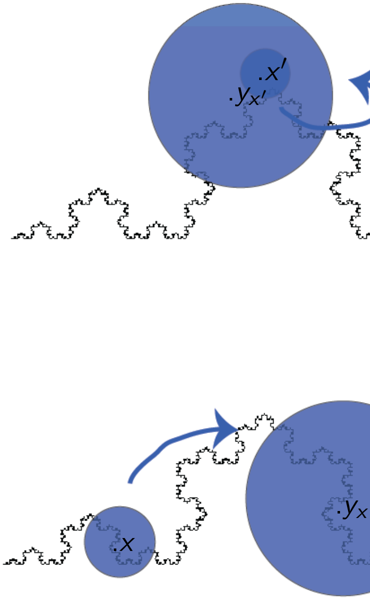
Heuristic proof

The proportion of K (or K^c) contained in a ball at a microscopic scale is the same than at a macroscopic scale, therefore

$$\sup_{x \in \partial K^r} \frac{\mathrm{Vol}(B(x,r)) \cap K}{\mathrm{Vol}(B(x,r))}$$

has periodical fluctuations in r, and

$$\gamma(K) \geq \inf_{r>0, x \in \partial K^{cr}} \frac{\mathrm{Vol}(B(x,r) \cap K)}{B(x,r)} > 0$$

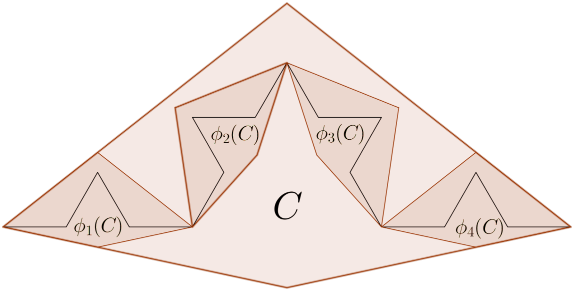


Example : Von Koch flake

Open set condition : there is an open set U such that

$$\Psi(U) = \bigcup_i \Phi_i(U) \subset U$$

and the union is disjoint.



Density of self-similar boundaries

Theorem (LR, Vega)

Let K compact such that ∂K is a finite union of self-similar sets $E_1, \dots, E_q$ such that

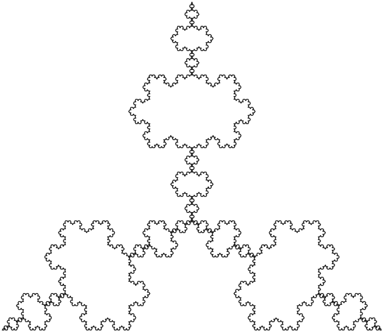
- *E_i satisfies the open set condition with some open set U_i such that $U_i \cap E_i \neq \emptyset$*
- *$U_i \cap \partial K \subset E_i$*
- *$U \setminus \partial K$ has a finite number of convex components*

Then there is $c > 0$ such that for some x on K's boundary there is $y \in K$ such that $B(y, c\varepsilon) \subset B(x, \varepsilon)$

Therefore the previous results apply

Antiflake

- Limit theorems apply to the Von Koch anti flake



- We exhibit a counter-example with a Cantor-like (disconnected) self-similar boundary

Hausdorff distance

We have (Cuevas et al.) under some assumptions

$$d_H(K, K_n^{balls}) \leq C \left(\frac{\log(n)}{n}\right)^{1/d}$$

Can be seen as the maximum of variables of magnitude $n^{-1/d}$. Compares to

Theorem (Calka, Chenavier 2013)

Assume that

$$\inf_{x \in K, 0<r<1} \frac{Vol(B(x,r) \cap K)}{Vol(B(x,r))} > 0$$

Then, as $\lambda \to \infty$

$$\mathbf{P}\left(d_H(K, K^{\mathcal{P}_\lambda}) \quad \leq \quad c\left(\frac{\log(c'\lambda \log(\lambda)^{d-1}))}{\lambda}\right)^{1/d} \quad\right) \quad \to \quad 1$$

Hausdorff distance

Theorem (LR, Vega)

Assume that K satisfies the strong regularity assumption. Then, independently of s, almost surely, for large enough n

$$0 < c(\alpha^-, s, d) \leq \liminf_n \frac{d_{\mathcal{H}}(K, K^{\mathcal{X}_n})}{(n^{-1}\ln(n))^{1/d}} \leq \limsup_n \cdots \leq C(\alpha^+, s, d) < \infty$$

Therefore this rate is optimal.

- Let K be a r-regular set (a ball of radius r can be rolled inside AND outside the boundary).
  Pateiro-Lopez 2007 introduces an other estimator $K_r(\mathcal{X}_n)$ based on erosion which yields

$$d_{\mathcal{H}}(K_r(\mathcal{X}_n), K) \leq C \left(\frac{\log(n)}{n}\right)^{\frac{2}{d+1}}$$

  More promising direction for regular sets ?

Abstract Berry-Esseen bounds (LR,Peccati, in preparation)

We have the general bound, for any centred symmetric functional φ in n IID variables $X = (X_1, \dots, X_n)$,

$$\sup_{t \in \mathbb{R}} |\varphi(X_1, \dots, X_n) - \Phi(t)| \leq C \left[\sigma^{-2} n^{\frac{1}{2}} \sum_{j,k} \sqrt{\alpha_{j,k}} + \sigma^{-3} n \sqrt{\mathbf{E} \left| D_n \varphi(X) \right|^6} \right]$$

where $\sigma^{-2} = \mathbf{Var}(\varphi(X))$ and

$$D_i \varphi(X) = \varphi(X) - \varphi(X_1, \dots, \widehat{X_i}, \dots, X_n),$$

and

$$\alpha_{j,k} = \sup_{(Y, Y', Z)} \mathbf{E} \left[\mathbf{1}_{\{D_1(D_j \varphi(Y)) \neq 0, D_1(D_k \varphi(Y')) \neq 0\}} D_j \varphi(Z)^4 \right]$$

where $Y, Y', Z \overset{(d)}{=} X$ are such that Y_i, Y_i', Z_i are independent of everything but X_i.

→ Application to random grain boolean model ($n^{-1/2}$ bounds)

Variance tools

Theorem (Efron-Stein)

For φ measurable symetric :

$$\mathbf{E}\left(\varphi(\mathcal{X}_n) - \mathbf{E}\varphi(\mathcal{X}_n)\right)^2 \leq \frac{n}{2}\mathbf{E}\left(\varphi(\mathcal{X}_{n+1}) - \mathbf{E}\varphi(\mathcal{X}_n)\right)^2$$

Has the lower bound the same order of magnitude ?

Binomial lower bound (First Hoeffding term)

Theorem (LR, Peccati)

For φ symmetric square integrable

$$\mathbf{Var}(\varphi(\mathcal{X}_n)) \geq n \int_{[0,1]^d} g_1(x)^2 dx$$

where

$$\begin{aligned} g_1(x) &= \mathbf{E}\varphi(\mathcal{X}_n \cup \{x\}) - \varphi(\mathcal{X}_{n+1}) \\ &= \underbrace{\mathbf{E}\varphi(\mathcal{X}_n \cup \{x\}) - \mathbf{E}\varphi(\mathcal{X}_n)}_{u(x)} \quad + \quad \underbrace{\mathbf{E}\varphi(\mathcal{X}_n) - \mathbf{E}\varphi(\mathcal{X}_{n+1})}_{v} \end{aligned}$$

We have

$$\int_{[0,1]^d} u(x)dx = v \leq \sqrt{\int_{[0,1]^d} u(x)^2 dx} \text{ by Cauchy-Schwarz,}$$

whence $\|u(x)\|_{L^2} \sim v$ when the problem is homogeneous. In our case, the problem is concentrated around K's boundary, whence $v << \|u\|_2$.