Extrêmes, géométrie aléatoire, analyse topologique des données et modèle IDLA

Nicolas Chenavier

UNIVERSITÉ DU LITTORAL
CÔTE D'OPALE

Document de synthèse en vue de
L'HABILITATION A DIRIGER DES RECHERCHES

Discipline:

Mathématiques

Ecole doctorale Sciences, Technologie, Santé (ED585)

Présentée par

Nicolas CHENAVIER

Extrêmes, géométrie aléatoire, analyse topologique des données et modèle IDLA

date de soutenance: 08 décembre 2022

devant le jury composé de :

Mme. Hermine Biermé	Professeur (Université de Tours), *Examinatrice*
M. Pierre Calka	Professeur (Université de Rouen Normandie), *Examinateur*
M. Frédéric Chazal	Directeur de recherches (inria Saclay), *Examinateur*
M. David Coupier	Professeur (IMT Nord Europe), *Examinateur*
M. Clément Dombry	Professeur (Université de Franche-Comté), *Rapporteur*
M. Nathanaël Enriquez	Professeur (Université Paris-Saclay), *Rapporteur*
M. Zakhar Kabluchko	Professeur (Université de Münster), *Rapporteur*
M. Dominique Schneider	Professeur (Université du Littoral Côte d'Opale), *Examinateur*

Remerciements

Mes premiers remerciements s'adressent chaleureusement à Clément Dombry, Nathanaël Enriquez et Zakhar Kabluchko qui m'ont fait l'honneur d'être mes rapporteurs. Merci pour vos rapports très détaillés et pour vos encouragements. Merci encore, Nathanaël, pour les discussions que nous avons eues; j'ai éprouvé également beaucoup de plaisir.

Je remercie Hermine Biermé, Pierre Calka, Frédéric Chazal, David Coupier et Dominique Schneider d'avoir accepté d'être dans mon jury. Vos présences me touchent et m'honorent. Merci encore, Pierre, de m'avoir initié à la géométrie aléatoire. Merci, Dominique, pour les nombreux conseils que tu m'as donnés et les nombreuses expériences scientifiques que nous avons eues: co-encadrement de la thèse d'Ahmad, rédaction d'articles, co-organisation de conférences, etc.

Je remercie les directeurs ou ex-directeurs du LMPA: Shalom Eliahou, Carole Rosier et Hassane Sadok qui, chacun, ont offert ou offrent d'excellentes conditions pour la recherche, ainsi que Loïc Foissy pour la direction du département et pour m'avoir fait entrer au jury du capes; c'était une belle expérience. L'ambiance qui règne au LMPA est très agréable; c'est avec enthousiasme que je remercie tous ses membres pour leur bonne humeur et particulièrement les collègues que j'ai le plaisir de voir lors des pauses déjeuners et en dehors du LMPA: Antoine[1], Bruno, Christophe, Christian, Jean, Julie, Loïc (à nouveau!), Lucile, Pierre, Romuald, Thierry, Vincent et Xavier. Merci également à Denis Bitouzé pour son aide précieuse concernant LaTeX et à Isabelle Buchard pour son sérieux et sa grande efficacité au secrétariat. Petite mention pour Christophe: évidemment, je m'incline respectueusement devant toi pour avoir su t'entourer de nombreux disciples qui sont entrés dans l'univers de Magic et avec lesquels nous passons toujours des moments très agréables. Grâce à toi, j'ai appris que le monde est peuplé de nombreuses créatures, avec des rituels et des enchantements; et qu'il faut toujours avoir un deck par sécurité.

Je remercie tous les collègues avec lesquels j'ai travaillé pour mes travaux de recherche; particulièrement David Coupier (à nouveau!) et Arnaud Rousselle pour le beau travail que nous avons fait sur l'IDLA et le co-encadrement de thèse de Keenan; ainsi que Christian Y. Robert avec qui j'ai appris de nombreux concepts scientifiques. Merci, Ahmad et Keenan, pour la confiance que vous me témoignez ou m'avez témoignée concernant l'encadrement de vos thèses; je suis ravi et ai été ravi de travailler avec vous. Merci également à l'équipe de probabilités/statistique du LPP, notamment David Dereudre et Chi Tran, de m'avoir intégré dans son groupe de travail de géométrie aléatoire dès mon recrutement à l'ULCO.

Enfin, merci à mes amis et à ma famille pour leur soutien constant; et particulièrement à mon frère, Cyrille, avec qui je suis toujours très content de discuter de mathématiques et de jouer au tennis. Merci à Ania, qui est une épouse et une maman très attentionnée; et à notre enfant, Antoine[2], qui est l'être le plus merveilleux de la Terre (sans vouloir dénigrer qui que ce soit, et encore moins tous ceux que j'ai mentionnés ci-dessus).

[1] Le (grand) Antoine
[2] Le (petit) Antoine

Contents

Avant-propos

Le mémoire s'articule autour de quatre chapitres. Les trois premiers regroupent des travaux par ordre thématique et présentent, chacun, une introduction au domaine en question. Le dernier porte sur divers travaux.

Le premier chapitre, de loin le plus conséquent, porte sur des problèmes d'extrêmes en géométrie aléatoire. Les problèmes considérés s'inscrivent dans la continuité de ce que j'ai fait en thèse. Les modèles étudiés sont essentiellement les mosaïques aléatoires et les résultats sont principalement des théorèmes limites. Plusieurs d'entre eux concernent des approximations poissoniennes. Ces approximations permettent d'étudier les comportements limites des maxima et minima, et plus généralement des statistiques d'ordre, de diverses caractéristiques géométriques. D'autres résultats concernent la répartition spatiale d'extrêmes et notamment la taille d'un cluster typique d'excédents. Je présente également quelques résultats d'extrêmes sur Delaunay lorsque ce dernier est vu comme un graphe et non plus comme une mosaïque.

Le deuxième chapitre porte sur deux problèmes d'analyse topologique des données. Le premier concerne un théorème central limite fonctionnel et fournit des tests permettant de discriminer les processus ponctuels. Le second est un problème d'extrêmes sur les durées de vie de cycles persistants. Bien que je ne sois pas spécialiste d'analyse topologique des données, j'ai voulu y consacrer un chapitre car j'aimerais développer cette thématique.

Le troisième chapitre porte sur un modèle de croissance: l'IDLA. Ce dernier est construit récursivement à partir de marches aléatoires. Le protocole permettant de le définir permet également de construire un arbre aléatoire qui est délicat à étudier. Le résultat principal de ce chapitre est l'existence d'une forêt aléatoire, basée sur le protocole IDLA, qui a pour but d'approcher l'arbre. Même si je n'ai fait qu'un seul travail sur l'IDLA, j'y consacre un chapitre entier car il s'agit d'un axe que je souhaite développer prioritairement. En particulier, à partir d'octobre 2022, je co-encadrerai la thèse de Keenan Penner (officiellement seulement avec David Coupier mais, en pratique, également avec Arnaud Rousselle) sur cette thématique.

Le quatrième chapitre porte sur divers travaux, notamment le phénomène de premier chiffre (en collaboration avec deux collègues de mon laboratoire) et un problème d'extrêmes de marches aléatoires en environnement aléatoire avec mon ex-doctorant (Ahmad Darwiche) que j'ai co-encadré avec Dominique Schneider. Je présente également des travaux soumis récemment ou encore en cours. Le plus significatif, de loin, porte sur un problème d'inférence pour un champ max-stable dans une fenêtre fixée, en collaboration avec Christian Y. Robert. Le mémoire se termine par quelques perspectives.

Tout au long du mémoire, j'énonce en général le théorème principal de chaque publication et donne des esquisses de preuves.

Foreword

The manuscript is divided into four chapters. The first three one deal with three different topics, with introductions of the latter. The last one concerns various works.

The first chapter is the most significant and deals with extremes in Stochastic Geometry. The problems which are considered are in the continuity of my PhD thesis. The main models are the random tessellations and the results mainly consist of limit theorems. Many of them are Poisson approximations. These approximations are useful to investigate the asymptotic behaviours of maxima and minima, and more generally of order statistics, of various geometric characteristics. Other results concern the spatial repartition of extremes and in particular the size of a typical cluster of exceedances. I also give results on extremes for Delaunay when the latter is seen as a graph and not as a tessellation.

The second chapter deals with two problems in Topological Data Analysis. The first one concerns a functional central limit theorem and provides tests to discriminate point processes. The second one deals with extremes of lifetimes for persistent cycles. Although I am not a specialist of Topological Data Analysis, I have written an entire chapter on this topic because it is one of the domains that I would like to deal with in depth.

The third chapter concerns a random growth model: the IDLA. The latter is constructed recursively from simple random walks. The protocol defining this model can also define a random tree whose the study is delicate. The novelty of this chapter is the construction of a random forest which is based on the IDLA protocol and which should approximate the random tree. Although I have written only one paper on this topic, I devote an entire chapter because it is one of the topics that I would like to investigate firstly. In particular, with David Coupier and Arnaud Rousselle, I will supervise the PhD thesis of Keenan Penner from October 2022.

The fourth chapter concerns various works, including the first digit phenomenon (in collaboration with two colleagues of my laboratory) and extremes of random walks in random scenery with an ex-PhD student (Ahmad Darwiche) that I supervised with Dominique Schneider. I also present some recent works and works in progress. The most significant one, in collaboration with Christian Y. Robert, deals with inference for a max-stable random field in a fixed window. The manuscript ends with several perspectives.

In general, throughout the manuscript, I state the main theorem of each publication with sketch of proof.

Publications

1. The bi-dimensional directed IDLA forest, travail joint avec D. Coupier et A. Rousselle; à paraître dans *Annals of Applied Probability* (2022+).
2. Limit laws for large kth-nearest neighbor balls, travail joint avec N. Henze et M. Otto; *Journal of Applied Probability*: **59**(3), p. 880-894 (2022).
3. Extremal lifetimes of persistent loops and holes, travail joint avec C. Hirsch; *Extremes*: **25**(2), p. 299-330 (2022).
4. Testing goodness of fit for point processes via topological data analysis, travail joint avec C. Biscio, C. Hirsch et A.M. Svane; *Electronic Journal of Statistics*: **14**, No. 1, p. 1024-1074 (2020).
5. The maximal degree in a Poisson-Delaunay graph, travail joint avec G. Bonnet; *Bernoulli*: **26**(2), p. 948-979 (2020).
6. Extremes for transient random walks in random sceneries under weak independence conditions, travail joint avec A. Darwiche; *Statistics and Probability Letters*: **158** (2020).
7. The largest order statistics for the inradius in an isotropic STIT tessellation, travail joint avec W. Nagel; *Extremes*: **22**, issue 4, p. 571-598 (2019).
8. Cluster size distributions of extreme values for the Poisson-Voronoi tessellation, travail joint avec C. Y. Robert; *Annals of Applied Probability*: **28**(6), p. 3291-3323 (2018).
9. Products of random variables and the first digit phenomenon, travail joint avec B. Massé et D. Schneider; *Stochastic Processes and their Applications*: **128**, p. 1615-1634 (2018).
10. Stretch factor in a planar Poisson-Delaunay triangulation with a large intensity, travail joint avec O. Devillers; *Advances in Applied Probability*: **50**, p. 35-56 (2018).
11. On the discrepancy of powers of random variables, travail joint avec D. Schneider; *Statistics and Probability Letters*: **134**, p. 5-14 (2018).
12. Extremes for the inradius in the Poisson line tessellation, travail joint avec R. Hemsley; *Advances in Applied Probability*: **48**, p. 544-573 (2016).
13. The extremal index for a random tessellation; *Geometric Science of Information, LNCS, Springer*: **9389** (2015), p. 171-178
14. A general study of extremes of stationary tessellations with applications; *Stochastic Processes and their Applications*: **124**, p. 2917-2953 (2014).
15. Extreme values for characteristic radii of a Poisson-Voronoi Tessellation, travail joint avec P. Calka; *Extremes*: **17**, p. 359-385 (2014).

Chapter 1

Extremes in Stochastic Geometry

Sommaire

1.1 Introduction

1.1.1 Random tessellations

Random tessellations are one of the most classical objects in Stochastic Geometry. By a (convex) tessellation of the Euclidean space $\mathbf{R}^d$, $d \geq 1$, endowed with the Euclidean norm $|\cdot|$, we mean a locally finite collection $\{C_i\}_{i\geq 1}$ of polytopes (referred to as *cells*) such that C_i and C_j have disjoint interiors for any $i \neq j$ and $\bigcup_{i\geq 1} C_i = \mathbf{R}^d$. We endow the family of (convex) tessellations of $\mathbf{R}^d$ with the σ-field generated by sets of the form

$$\{\mathcal{T} = \{C_i\}_{i\geq 1} : (\cup_{i\geq 1} \partial C_i) \cap K \neq \emptyset\},$$

where K is a compact set of $\mathbf{R}^d$. A *random tessellation* is a random variable with values in the set of (convex) tessellations of $\mathbf{R}^d$.

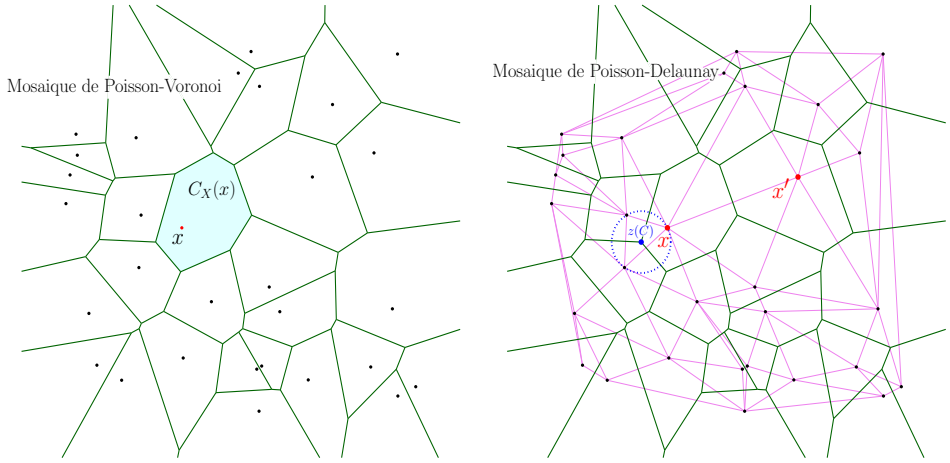


Figure 1.1: *(a)* Voronoi tessellation. *(b)* The same one (green) and its dual: the Delaunay tessellation (purple).

Examples of random tessellations

The most famous tessellations are the Voronoi, the Delaunay, the hyperplane and the STIT tessellations. We present below each of them.

Voronoi tessellation Given a point process Φ such that the convex hull $\text{conv}(\Phi)$ of Φ is $\mathbf{R}^d$ a.s., the *Voronoi tessellation* based on Φ is defined as the collection $\{C_\Phi(x) : x \in \Phi\}$, where

$$C_\Phi(x) = \{y \in \mathbf{R}^d, |y - x| \leq |y - x'|, x' \in \Phi\},$$

is called the *Voronoi cell* with nucleus x. When Φ is a Poisson point process, the family $\{C_\Phi(x) : x \in \Phi\}$ is called the Poisson-Voronoi tessellation; see Figure 1.1, *(a)* for a realization in $2D$ (observed in a window). The latter is stationary and isotropic if Φ is stationary.

Delaunay tessellation Let Φ be a point process in general position (i.e. such that there is no $d+2$ points of Φ contained in a sphere and no $d+1$ points on the same hyperplane) a.s. and such that $\text{conv}(\Phi) = \mathbf{R}^d$ a.s.. The Delaunay graph based on Φ is the unique triangulation with vertices in Φ such that the circumball of each simplex contains no point of Φ in its interior. The *Delaunay tessellation* is then defined as the family of these simplices. A Delaunay tessellation corresponds to the dual graph of Voronoi tessellation in the following way: there exists an edge between two points $x_1, x_2 \in \Phi$ in the Delaunay graph if and only if they are Voronoi neighbors, i.e. $C_\Phi(x_1) \cap C_\Phi(x_2) \neq \emptyset$. When Φ is a Poisson point process, the random tessellation is referred to as the Poisson-Delaunay tessellation; see Figure 1.1, *(b)* for a realization in $2D$ (observed in a window). The latter is stationary and isotropic when Φ is stationary.

Hyperplane tessellation Given a point process Φ in $\mathbf{R}^d$ which a.s. does not contain the origin, we denote by H_x, $x \in \Phi$, the hyperplane which is orthogonal to x and which contains x, i.e.

$$H_x = \{y \in \mathbf{R}^d : \langle y - x, x \rangle = 0\}.$$

The *Hyperplane tessellation* based on Φ consists of the set of closures of connected components of $\mathbf{R}^d \setminus \cup_{x\in\Phi} H_x$. When Φ is a Poisson point process whose the intensity measure has a density w.r.t. the Lebesgue measure which equals $|\cdot|^{-(d-1)}$, the hyperplane tessellation is stationary and isotropic; see Figure 1.2, *(a)* for a realization in $2D$ (observed in a window). This model is referred to as the Poisson hyperplane tessellation.

STIT tessellation To introduce the notion of STIT (STable under ITeration) tessellation, we proceed as follows. We start with the construction of a tessellation process $(\mathfrak{m}_{t,W},\, t \geq 0)$ in a window $W \subset \mathbf{R}^d$. Let $\tau_0, \tau_1, \tau_2, \ldots$ be independent and identically distributed (i.i.d.) random variables, all exponentially distributed with parameter 1. Let Λ be the measure on the set of hyperplanes $\mathcal{H}$ of $\mathbf{R}^2$, which is invariant under translations and rotations of $\mathbf{R}^d$ and let

$$[B] := \{H \in \mathcal{H} :\, H \cap B \neq \emptyset\}.$$

(i) The initial state of the process is $\mathfrak{m}_{0,W} = \{C_1\} := \{W\}$, and the random holding time in this state is $\tau_0/\Lambda([W])$, i.e. it is exponentially distributed with parameter $\Lambda([W])$.

(ii) At the end of the holding time, the window W is divided by a random hyperplane H_1 with law $(\Lambda([W]))^{-1}\Lambda(\cdot \cap [W])$. The new state of the STIT process is now $\{C_1, C_2\}$, where $C_1 := W \cap H_1^+$ and $C_2 := W \cap H_1^-$, and H_1^+ and H_1^- are the two closed half-planes generated by H_1. The random life times of C_1 and C_2 are $\tau_1/\Lambda([C_1])$ and $\tau_2/\Lambda([C_2])$, respectively.

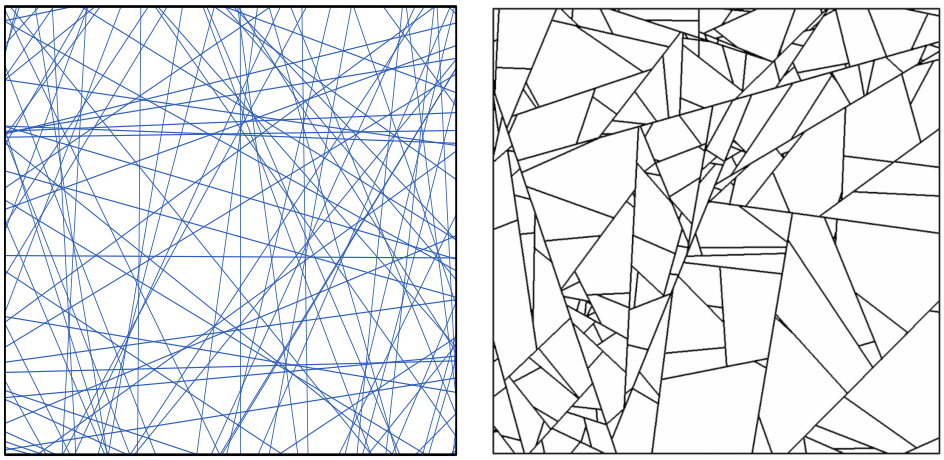
(iii) Now, inductively, for $t > 0$, assume that $\mathfrak{m}_{t,W} = \{C_{i_1}, \ldots, C_{i_n}\}$. The life times of the cells are $\tau_{i_1}/\Lambda([C_{i_1}]), \ldots, \tau_{i_n}/\Lambda([C_{i_n}])$, respectively. At the end of the life time of a cell C_{i_j}, this cell is divided by a random hyperplane H_{i_j} with the law $(\Lambda([C_{i_j}]))^{-1}\Lambda(\cdot \cap [C_{i_j}])$, which is a probability distribution on $[C_{i_j}]$. Given the state of the tessellation process at the time of division, this hyperplane is conditionally independent from all the other dividing lines used so far. The divided cell C_{i_j} is deleted from the tessellation and is replaced by the two "daughter" cells $C_{i_j} \cap H_{i_j}^+$ and $C_{i_j} \cap H_{i_j}^-$. These cells are endowed with new indexes from $\mathbf{N}$ which are not used before in this process.

An essential property of the construction is that the distribution of the tessellation generated in a window W is spatially consistent in the following sense. If W and W' are two convex polygons with $W \subset W'$ and $\mathfrak{m}_{t,W}$, $\mathfrak{m}_{t,W'}$ the respective random tessellations, then $\mathfrak{m}_{t,W} \stackrel{d}{=} \mathfrak{m}_{t,W'} \wedge W$ are identically distributed, where

$$\mathfrak{m}_{t,W'} \wedge W := \{C \cap W :\, C \in \mathfrak{m}_{t,W},\, C \cap W^\circ \neq \emptyset\}$$

is the restriction of $\mathfrak{m}_{t,W'}$ to W. This property yields the existence of a stationary random tessellation $\mathfrak{m}_t$ of $\mathbf{R}^2$, referred to as *STIT tessellation*, such that its restriction $\mathfrak{m}_t \wedge W$ to any window W has the same distribution as the constructed tessellation $\mathfrak{m}_{t,W}$. Since the measure Λ is invariant under rotation, the STIT tessellation $\mathfrak{m}_t$ is isotropic. See Figure 1.2, *(b)* for a realization in $2D$ (observed in a window).

The first three tessellations that we discussed above, when they are based on suitable Poisson point processes, and the STIT tessellation have in common the fact that they have mixing properties. As we will see in the next subsections, such properties are important to deal with extremes.



(a) *(b)*

Figure 1.2: *(a)* Line tessellation. *(b)* STIT tessellation.

Typical cell

Given a fixed realization of a stationary random tessellation $\mathfrak{m}$, we associate with each cell $C \in \mathfrak{m}$ in a deterministic way a point $z(C)$, which is called the *nucleus* of the cell, such that $z(C+x) = z(C) + x$ for all $x \in \mathbf{R}^d$. For instance, when $\mathfrak{m}$ is a Voronoi tessellation based on a point process Φ (resp. a Delaunay tessellation, or an hyperplane/STIT tessellation), it is usual to take $z(C_\Phi(x)) = x$ for any $x \in \Phi$ (resp. it is usual to associate with each Delaunay cell the circumcenter of the cell, or to associate with each cell of the hyperplane/STIT tessellation its incenter). To describe the mean behaviour of a stationary random tessellation, the notions of intensity and typical cell are introduced as follows. Let B be a Borel subset of $\mathbf{R}^d$ such that $\lambda_d(B) \in (0,\infty)$, where λ_d is the d-dimensional Lebesgue measure. The *intensity* γ of the tessellation is defined as $\gamma = \frac{1}{\lambda_d(B)} \cdot \mathbb{E}\left[\#\{C \in \mathfrak{m}, z(C) \in B\}\right]$ and we assume that $\gamma \in (0,\infty)$. Since $\mathfrak{m}$ is stationary, γ is independent of B. The *typical cell* $\mathscr{C}$ is a random polytope whose the distribution is given by

$$\mathbb{E}[f(\mathscr{C})] = \frac{1}{\gamma\lambda_d(B)} \cdot \mathbb{E}\left[\sum_{\substack{C\in\mathfrak{m},\\ z(C)\in B}} f(C - z(C))\right]$$

for all $f : \mathcal{K}_d \to \mathbf{R}$ bounded measurable function on the set of convex bodies $\mathcal{K}_d$, i.e. convex compact sets with non-empty interior.

Thanks to the Slivnyak-Mecke formula (see e.g. Theorem 3.2.5 in [90]), it can be proved that the typical cell of a Voronoi tessellation, based on a stationary Poisson point process Φ, is equal in distribution to $C_{\Phi\cup\{0\}}(0)$. Explicit representations of the distribution of the typical cell (or of the interior of the typical cell) for other classical random tessellations have also been established (see e.g. Theorem 10.4.4 in [90] for the Poisson-Delaunay tessellation, and Theorem 10.4.6 in [90] for the Poisson hyperplane tessellation).

A lot of works has been done on random tessellations, including ergodic theorems [35], central limit theorems [52, 91], computations or estimates of laws or tails for various geometric characteristics [17, 84, 101], shape theorems for large cells [21, 57, 58] and models in non-Euclidean

geometry, see e.g. [19]. For a complete account of random tessellations and their applications, we refer to the books [79, 90].

1.1.2 Concepts of Extreme Value Theory

Univariate case

Extreme Value Theory (EVT) deals with rare events and has many applications in various domains such as hydrology, finance and climatology. EVT was first introduced in a univariate setting.

Given a stationary sequence of real random variables (X_i), the main question is to investigate the limit behaviour of the maximum $M_n = \max_{i\leq n} X_i$. When n goes to infinity, the random variable M_n converges in probability to a constant $x_* \in \mathbf{R}\cup\{+\infty\}$. One way to be more precise is to find a threshold u_n, of the form $u_n = u_n(t) = a_n t + b_n$, where $a_n > 0$, $b_n \in \mathbf{R}$ and where $t \in \mathbf{R}$ is a parameter, in such a way that $\mathbb{P}(M_n \leq u_n(t))$ converges to a non-degenerate limit, i.e. such that

$$a_n^{-1}(M_n - b_n) \xrightarrow[n\to\infty]{\mathcal{D}} Y, \tag{1.1.1}$$

where Y is a random variable whose the distribution function is not degenerate. When the sequence (X_i) is i.i.d. and when Equation (1.1.1) holds, the random variable Y necessarily belongs to the class of extreme value distributions. This class consists of three types of laws, namely Fréchet, Gumbel and Weibull (see e.g. Theorem 1.1.3 in [37]).

Extremes under the $D(u_n)$ and $D'(u_n)$ conditions It is straightforward that if (X_i) is a sequence of i.i.d. (real) random variables then the following property holds: for any sequence of real numbers (u_n), and for $\tau > 0$,

$$n\mathbb{P}(X_1 > u_n) \underset{n\to\infty}{\longrightarrow} \tau \Longrightarrow \mathbb{P}(M_n \leq u_n) \underset{n\to\infty}{\longrightarrow} e^{-\tau}. \tag{1.1.2}$$

The above property has been extended for sequences of dependent random variables satisfying two conditions due to Leabdetter [71]. We recall them since we will use conditions of this type many times in our manuscript.

Definition 1.1.1. *Let (X_i) be a stationary sequence of real random variables and let (u_n) be a deterministic sequence of real numbers. Let $F_{i_1,\dots,i_n}(u) = \mathbb{P}(X_{i_1} \leq u, \dots, X_{i_n} \leq u)$. We say that $(X_n)_{n\geq 1}$ satisfies the $D(u_n)$ condition if, for any n, ℓ and for any intergers $i_1, \dots, i_p, j_1, \dots, j_p$ such that $1 \leq i_1 < i_2 < \cdots < i_p < j_1 < \cdots < j_{p'} \leq n$ with $j_1 - i_p \geq \ell$, we have*

$$\left|F_{i_1,\dots,i_p,j_1,\dots,j_{p'}}(u_n) - F_{i_1,\dots,i_p}(u_n)F_{j_1,\dots,j'_p}(u_n)\right| \leq \alpha_{n,\ell},$$

where $\alpha_{n,\ell_n} \to 0$ as $n \to \infty$ for some sequence $(\ell_n)_{n\geq 1}$ with $\ell_n = o(n)$.

Roughly, the $D(u_n)$ condition is a condition ensures a mixing-type behaviour for the tails of the joint distributions of a stationary sequence of random variables. In particular, sequences satisfying a strong mixing property also satisfy the $D(u_n)$ condition. The second condition of Leadbetter is stated below.

Definition 1.1.2. *In conjunction to the $D(u_n)$ condition, we say that the sequence (X_i) satisfies the $D'(u_n)$ condition if*

$$\limsup_{n\to\infty} n \sum_{j=2}^{\lfloor n/k\rfloor} \mathbb{P}(X_1 > u_n, X_j > u_n) \longrightarrow 0, \quad \text{with } k \to \infty.$$

The above condition is of local type. It ensures that, with high probability, it is not possible to have a pair of exceedances, i.e. a pair of random variables exceeding the threshold u_n, in the same neighborhood. Under the $D(u_n)$ and $D'(u_n)$ conditions, the random variable $M_n = \max_{i\leq n} X_i$ has the same behaviour as if the random variables X_i are i.i.d.. In particular, Equation (1.1.1) holds, and the limit law of the normalized maximum (provided that the limit exists) belongs to the class of extreme value distributions.

Assume from now on that, for any $\tau > 0$, there exists a threshold $u_n = u_n^{(\tau)}$ such that

$$n\mathbb{P}\left(X_1 > u_n\right) \underset{n\to\infty}{\longrightarrow} \tau.$$

One of the classical objects in EVT is the so-called *point process of exceedances*, defined as

$$\Phi_n = \{i/n : X_i > u_n,\ i \leq n\}.$$

Under the assumptions that the $D(u_n)$ and $D'(u_n)$ conditions hold for any $u_n = u_n^{(\tau)}$, the point process Φ_n converges to a homogeneous Poisson process Φ in $[0,1]$ of intensity τ. As a consequence, asymptotic results on the order statistics, i.e. on the random variables $M_n^{(r)}$, $r \geq 1$, where $M_n^{(r)}$ denotes the r-th largest value of the X_i's, can be derived. Indeed, since $M_n^{(r)} \leq u_n$ if and only if $\#\Phi_n \leq r-1$ and since $\Phi_n \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \Phi$,

$$\mathbb{P}\left(M_n^{(r)} \leq u_n\right) \underset{n\to\infty}{\longrightarrow} \sum_{k=0}^{r-1} e^{-\tau}\cdot\frac{\tau^k}{k!}.$$

Poisson approximation with rate of convergence Under stronger assumptions than the $D(u_n)$ and $D'(u_n)$ conditions, rates of convergence for the Poisson approximation of the number of exceedances can be provided. A general result, due to Arratia *et al* [2] and based on the Chen-Stein method, gives explicit rates. We recall their result since we will use many times in our works.

For an arbitrary index set I, and for $i \in I$, let X_i be a Bernoulli random variable with $p_i = \mathbb{P}\left(X_i = 1\right) = 1 - \mathbb{P}\left(X_i = 0\right)$. For each $i, j \in I$, we write $p_{ij} = \mathbb{E}\left[X_i X_j\right]$. Further, we let

$$X := \sum_{i\in I} X_i \quad \text{and} \quad \lambda := \mathbb{E}\left[X\right] = \sum_{i\in I} p_i, \quad \text{and assume that} \quad 0 < \lambda < \infty.$$

For each $i \in I$, fix a "neighborhood" $B_i \subset I$ with $i \in B_i$, and define

$$b_1 := \sum_{i\in I}\sum_{j\in B_i} p_i p_j, \quad b_2 := \sum_{i\in I}\sum_{i\neq j\in B_i} p_{ij}, \quad b_3 := \sum_{i\in I}\mathbb{E}\left[\left|\mathbb{E}\left[X_i - p_i \,\middle|\, \sum_{j\in I\setminus B_i} X_j\right]\right|\right]. \quad (1.1.3)$$

Roughly, b_1 measures the neighborhood size, b_2 measures the expected number of neighbors of a given occurrence and b_3 measures the dependence between an event and the number of occurrences outside its neighborhood. We are now prepared to state the main result of Arratia *et al.* (see Theorem 1 of [2]).

Proposition 1.1.3. *(Arratia, Goldstein, Gordon) Let Z be a Poisson random variable with mean $\lambda \in (0,\infty)$. With the above notation and the assumptions, we have*

$$d_{TV}(X,Z) \leq 2\left((b_1+b_2)\cdot\frac{1-e^{-\lambda}}{\lambda} + b_3\cdot\min\{1,\, 1.4\,\lambda^{-1/2}\}\right),$$

where $d_{TV}(\cdot,\cdot)$ denotes the Total Variation distance.

Clusters of exceedances When the stationary sequence (X_i) only satisfies the $D(u_n)$ condition (for any $u_n = u_n^{(\tau)}$, $\tau > 0$) but not necessarily the $D'(u_n)$ condition, clusters of exceedances can appear. In this case, there exists a number $\theta \in [0,1]$ such that, for any $\tau > 0$,

$$\mathbb{P}\left(M_n \leq u_n\right) \underset{n\to\infty}{\longrightarrow} e^{-\theta\tau}, \tag{1.1.4}$$

provided that the limit exists. When the latter holds, we say that θ is the *extremal index*. Such a quantity can be interpreted as the reciprocal of the mean size of a cluster of exceedances. Under suitable assumptions, including a slight modification of the $D(u_n)$ condition, the point process of exceedances converges to a compound Poisson point process (Theorem 4.2 in [56]) of intensity $\theta\tau$ and cluster size distribution $\pi = (\pi_k)$, where for any $k \geq 1$,

$$\pi_k = \lim_{n\to\infty} \mathbb{P}\left(\#\Phi_{B_n} = k \,|\#\Phi_{B_n} > 0\right), \tag{1.1.5}$$

with $B_n = \{0, \ldots, q_n\}$, $\Phi_{B_n} = \{i/n : X_i > u_n,\ i \in B_n\}$, $q_n \underset{n\to\infty}{\longrightarrow} \infty$ and $q_n = o(n)$. The probability π_k can be interpreted as the probability that we have k exceedances given that we observe a block of exceedances. Equation (1.1.5) is called the *blocks* characterization of the cluster size distribution. Under additional mild conditions, the extremal index is equal to the reciprocal of the mean of π. As an example, when the $D'(u_n)$ condition holds, exceedances are isolated and clusters are of size 1, i.e. $\pi = \delta_1$ and $\theta = 1$.

Another characterization is proposed in Theorem 4.1 in [88] and is given by

$$\lim_{n\to\infty} \mathbb{P}\left(\#\Phi_{B_n} = k | \, X_0 > u_n\left(\tau\right)\right) := p'_k = \theta \sum_{m=k}^{\infty} \pi_m, \qquad k \geq 1. \tag{1.1.6}$$

In particular, the extremal index is $\theta = p'_1$. This second characterization is useful to compute the values of the extremal index and the cluster size probabilities when the conditional distributions of the exceedances may be derived from the dynamics of (X_n), e.g. for the regularly varying multivariate time series [6] or the Markov sequences [83]. Equation (1.1.6) may be called the *runs* characterization of the cluster size distribution. This characterization is natural for a random object as a time series where the direction of time is used to design the dynamics of the series. Estimators of the extremal index and the cluster size distribution, based on the blocks and runs characterizations, are extensively investigated, see e.g. [86, 95].

Max-stable random fields

In the previous sub-section, we mentioned the class of extreme value distributions. A distribution $\mathcal{D}$ belonging to this class is *max-stable* in the sense that, if (X_i) is a sequence of i.i.d. random variables with distribution $\mathcal{D}$, then there exist two sequences (a_n) and (b_n), with $a_n > 0$, such that for any integer n,

$$a_n^{-1}(M_n - b_n) \overset{\mathcal{D}}{=} X_1,$$

where $M_n = \max_{i\leq n} X_i$. Reciprocally, any max-stable distribution belongs to the class of extreme value distributions.

The notion of max-stable distributions can be extended in a continuous framework in the following way. A stationary random field $\eta = (\eta(x))_{x\in\chi}$ in $\chi \subset \mathbf{R}^d$ with non-degenerate marginals and continuous paths is called *max-stable* if there exist two sequences of continuous functions $(a_n(x))_{x\in\chi}$, $(b_n(x))_{x\in\chi}$, with $a_n(x) > 0$, such that if (η_i) are i.i.d. copies of η, then

$$\left(a_n(x)^{-1}(M_n(x) - b_n(x))\right)_{x\in\chi} \overset{\mathcal{D}}{=} \eta.$$

In particular, the marginals of a max-stable stationary random field are (univariate) max-stable distributions. Being interested in the dependence structure, the attention can be reduced to max-stable random fields with standard unit Fréchet marginals, i.e. satisfying

$$\mathbb{P}\left(\eta(x) \leq z\right) = \exp(-z^{-1}),$$

for any $x \in \chi$ and $z > 0$. The max-stability property has then the simple form

$$n^{-1} \max_{i \leq n} \eta_i \stackrel{\mathcal{D}}{=} \eta.$$

A fundamental tool in the study of max-stable process is their spectral representation [36] due to de Haan: any stochastically continuous max-stable process η (with standard unit Fréchet marginals) can be written in the form

$$\eta(x) = \max_{i \geq 1} U_i Y_i(x),\ x \in \chi,$$

where (U_i) is the decreasing enumeration of the points of a Poisson point process on $(0, \infty)$ with intensity $u^{-2}\mathrm{d}u$, (Y_i) are i.i.d. copies of a non-negative stochastic random field Y on χ such that $\mathbb{E}\left[Y(x)\right] = 1$ for all $x \in \chi$, the sequences (U_i) and (Y_i) are independent. As a consequence of the spectral representation, the finite dimensional distributions can be made explicit in the following way: for any $(x_1, \ldots, x_k) \in \chi^k$, $k \geq 1$, and $(z_1, \ldots, z_k) \in \mathbf{R}_+^k$,

$$\mathbb{P}\left(\eta(x_1) \leq z_1, \ldots, \eta(x_k) \leq z_k\right) = \exp\left(-V_{x_1,\ldots,x_k}(z_1, \ldots, z_k)\right), \tag{1.1.7}$$

where

$$V_{x_1,\ldots,x_k}(z_1, \ldots, z_k) = \mathbb{E}\left[\max_{i=1,\ldots,k} \frac{Y(x_i)}{z_i}\right]$$

is the so-called *exponent measure*.

Various models of max-stable random fields have been introduced such as the Smith process [94], the stationary Gaussian extremal process originally introduced by Schlather [89] and the Brown-Resnick process introduced by Kabluchko, Schlather and de Haan [65]. A lot of results concerning max-stable random fields has been established such as ergodicity and mixing [64, 96], links with extremes of Gaussian processes [63], computations of conditional laws [42], conditional simulations [43] and estimates of parameters, see e.g. [41]. For a complete account on Extreme Value Theory and its applications, we refer to the books [37, 49].

1.1.3 Main problems

We give below a short description of our works on extremes in Stochastic Geometry. Each of them uses concepts which have been introduced in Sections 1.1.1 and 1.1.2.

In Section 1.2, we consider the following problem. Given a geometric characteristic, such as the volume or the inradius, and given a stationary random tessellation, we investigate the order statistics of the geometric characteristic over all cells centered in a window, say $W_\rho = \rho^{1/d}[-\frac{1}{2}, \frac{1}{2}]^d$, as ρ goes to infinity. All the tessellations that we consider, namely Poisson-Voronoi, Poisson-Delaunay, Poisson line and STIT tessellations, have mixing properties and satisfy in particular an analog of the $D(u_n)$ condition. Under an adaptation of the $D'(u_n)$ condition, we establish Poisson approximations for the point processes of exceedances. A similar result is also established in the context of large k-th nearest neighbor balls.

In Section 1.3, we consider a similar problem but this time without assuming the analog of the $D'(u_n)$ condition. In this case, clusters of exceedances can appear. In the context of Poisson-Voronoi and Poisson-Delaunay tessellations, we provide a new characterization of the extremal index, namely the reciprocal of the mean size of a cluster of exceedances, and show that the point process of exceedances converges to a compound Poisson point process. We illustrate our results through various numerical examples.

In Section 1.4, we consider two problems of extremes for a Poisson-Delaunay graph. The first one deals with the maximum of the degrees over all nodes in the window $W_\rho = \rho^{1/d}[-\frac{1}{2}, \frac{1}{2}]^d$ as ρ goes to infinity. Given two fixed nodes in the graph, the second one concerns the length of the shortest path in the graph between these two nodes.

1.2 Poisson approximation for the point process of exceedances

1.2.1 Voronoi and Delaunay tessellations

We present below the papers [20, 23]. The latter were written during our PhD thesis. Let us consider the following quantities:

- $\mathfrak{m}$: stationary random tessellation in $\mathbf{R}^d$, where each cell C of $\mathfrak{m}$ is associated with a point $z(C)$ of $\mathbf{R}^d$, referred to as the nucleus of the cell (see Section 1.1.1);
- $g : \mathcal{K}_d \to \mathbf{R}$: geometric characteristic (for instance the volume, the diameter), which is a translation invariant function defined on the set $\mathcal{K}_d$ of convex bodies of $\mathbf{R}^d$ (convex compact sets with non-empty interior);
- $W_\rho = \rho^{1/d} W$: window with volume ρ, where W is a convex body with unit volume.

The main question is: what can we say about the maximum of the geometric characteristic g over all cells with nucleus in W_ρ, i.e.

$$M_{W_\rho} = \max_{C \in \mathfrak{m} : z(C) \in W_\rho} g(C), \tag{1.2.1}$$

as ρ goes to infinity? Investigating this question is interesting for various reasons. First, the study of extremes could describe the regularity of the tessellation (e.g. presence of elongated cells). For instance, in the finite element method, the quality of the approximation depends on some consistency measurements over the partition. Another potential application field is statistics of point processes. The key idea would be to identify a point process from the extremes of a tessellation induced by the point process.

Extremes for characteristic radii

The pioneering work on extremes of random tessellations is [20]. In this paper, limit theorems on extremes for characteristic radii of a Poisson-Voronoi tessellation are given, namely (Theorem 1 in [20]):

- the maximum/minimum of inradii over all cells with nucleus in W_ρ;
- the maximum/minimum of circumradii over all cells with nucleus in W_ρ.

The inradius (resp. circumradius) of a Voronoi cell is defined as the radius of the largest ball included in the cell (resp. the smallest ball containing the cell) and centered at the nucleus of the cell. Theorem 1 in [20] is classical in the sense that the limit distributions are Fréchet or Weibull. As an application, the result on the maximum of circumradii gives an upper-bound for the Hausdorff distance between the convex body W and its so-called Poisson-Voronoi approximation. More precisely, given a Poisson point process η of intensity γ in $\mathbf{R}^d$, the Poisson-Voronoi approximation is defined as

$$\mathcal{V}_\eta(W) = \bigcup_{x \in \eta \cap W} C_\eta(x).$$

Then, under suitable assumptions on W, Theorem 1 in [20] (Equation (2c)) implies, with probability tending to 1 as γ goes to infinity,

$$d_H(W, \mathcal{V}_\eta(W)) \leq \left(c_1 \gamma^{-1} \log\left(c_2 \gamma (\log \gamma)^{d-1}\right)\right)^{1/d},$$

where c_1, c_2 are two constants which can be made explicit. Since then, the above result has been improved by Lachièze-Rey and Vega [67]. Furthermore, the paper [20] deals with the shape of the cell minimizing the circumradius. It is proved that, with high probability, such a cell is a simplex. The proofs of Theorem 1 in [20] use geometric interpretations. For the circumscribed radii, we write the distributions as covering probabilities of spheres and apply results of [18, 60]. The inscribed radii can be interpreted as interpoint distances. A study of the extremes of these distances was already done in several works, see e.g. [54].

A general result under a finite range condition

The paper [23] provides a more general method to investigate extremes of random tessellations. Some assumptions are required. First, similarly to (1.1.2), we assume that for any $\tau > 0$ there exists a family of thresholds $v_\rho = v_\rho(\tau)$ such that

$$\rho\, \mathbb{P}\left(g(\mathcal{C}) > v_\rho\right) \underset{\rho \to \infty}{\longrightarrow} \tau. \tag{1.2.2}$$

Secondly, we assume a finite range condition (the latter is restrictive compared to the $D(u_n)$ condition), referred to as *condition (FRC)*. The latter states that, conditional on a suitable event which occurs with high probability, the values of the function g in regions of the window which are distant enough are independent (see p2919 in [23] for a precise statement). Condition (FRC) is satisfied when $\mathfrak{m}$ is a Poisson-Voronoi or a Poisson-Delaunay tessellation. Thirdly, in order to avoid clusters of exceedances, and similarly to the $D'(u_n)$ condition, we assume that the following local correlation condition, referred to as *condition (LCC)*, holds:

$$N_\rho \mathbb{E}\left[\sum_{(C_1, C_2)_{\neq} \in \mathfrak{m}^2} \mathbb{1}_{z(C_1), z(C_2) \in \mathfrak{C}_\rho} \mathbb{1}_{g(C_1) > v_\rho, g(C_2) > v_\rho}\right] \underset{\rho \to \infty}{\longrightarrow} 0, \tag{1.2.3}$$

where $\mathfrak{C}_\rho = \left(\frac{\rho}{N_\rho}\right)^{1/d} [0,1]^d$ is a cube with volume ρ/N_ρ, with $N_\rho \underset{\rho \to \infty}{\longrightarrow} \infty$ and $N_\rho = o(\rho)$, and where $(C_1, C_2)_{\neq} \in \mathfrak{m}^2$ means that (C_1, C_2) is a pair of distinct cells of $\mathfrak{m}$. Equation (1.2.3) means that, with high probability, it is not possible to have a pair of exceedances in a small region of the window (in the sense that it is a cube with volume ρ/N_ρ, which is negligeable compared to the volume of W_ρ).

Similarly to Section 1.1.2, we define the so-called (normalized) point process of exceedances as

$$\Phi_{W_\rho} = \rho^{-1/d}\{z(C) \in W_\rho :\ C \in \mathfrak{m} \text{ and } g(C) > v_\rho\}.$$

The latter is a point process in the window $W_1 = W$ whose points consist of the nuclei of cells with geometric characteristic exceeding v_ρ. The main result of [23] is a Poisson approximation of the point process of exceedances and can be stated as follows:

Theorem 1.2.1. *Let $\mathfrak{m}$ be a stationary tessellation in $\mathbf{R}^d$. Let $\tau > 0$ and $v_\rho = v_\rho(\tau)$ be such that* (1.2.2) *holds. Under the (FRC) and (LCC) conditions, we have*

$$\Phi_{W_\rho} \xrightarrow[n\to\infty]{\mathcal{D}} \Phi,$$

where Φ is a homogeneous Poisson point process of intensity τ in W.

As a consequence of the above result, the asymptotic behaviour of the order statistics can be derived. Indeed, if we denote by $M_{W_\rho}^{(r)}$ the r-th largest values of g over all cells with nucleus in W_ρ (assuming that the $g(C)$'s are a.s. all different), $r \geq 1$, Theorem 1.2.1 implies

$$\mathbb{P}\left(M_{W_\rho}^{(r)} \leq v_\rho\right) = \mathbb{P}\left(\#\Phi_{W_\rho} \leq r-1\right) \underset{\rho\to\infty}{\longrightarrow} \sum_{k=0}^{r-1} e^{-\tau} \cdot \frac{\tau^k}{k!}. \tag{1.2.4}$$

In particular, $\mathbb{P}\left(M_{W_\rho} \leq v_\rho\right) \underset{\rho\to\infty}{\longrightarrow} e^{-\tau}$. In fact, Theorem 2 in [23] is slightly more precise than Theorem 1.2.1 in the sense that it does not only provide a spatial repartition of the exceedances but also the joint distribution of the order statistics. Moreover, a rate of convergence for (1.2.4) can be provided (Theorem 1 in [23]).

Numerous applications of Theorem 1.2.1 are given in [23] such as the limit behaviours of:

- the minimum of circumradii of a Poisson-Delaunay tessellation in any dimension and the maximum and minimum of the areas in the planar case;
- the minimum of distances to the farthest neighboring nucleus and the minimum of the volume of flowers for a Poisson-Voronoi tessellation;
- the maximum of inradii for a Voronoi tessellation induced by a Gauss-Poisson process.

The main difficulty to deal with these examples is to check the (LCC) condition since it requires delicate geometric estimates.

Although the (FRC) and (LCC) conditions are related to the $D(u_n)$ and $D'(u_n)$ conditions, the approach of the classical EVT cannot be applied to get Theorem 1.2.1. We describe below the main ideas to prove our theorem. For the sake of simplicity, we assume from now on that $W_\rho = \rho^{1/d}W$ with $W = [-\frac{1}{2}, \frac{1}{2}]^d$, and we only sketch the proof of the fact that $\#\Phi_{W_\rho}$ converges to a Poisson random variable with parameter τ. We subdivide the window W_ρ into a set V of sub-squares of equal size, such that $N_\rho^{1/d}$ is an integer and $N_\rho \underset{\rho\to\infty}{\longrightarrow} \infty$. These sub-cubes are indexed by the set of $\mathbf{i} = (i_1, \ldots, i_d) \in \left[1, N_\rho^{1/d}\right]^d$ and have the same volume as $\mathfrak{C}_\rho$, i.e. ρ/N_ρ. With a slight abuse of notation, we identify a cube with its index. For each $\mathbf{i} \in V$, we denote by

$$M_{\mathbf{i}} = \max_{C\in\mathfrak{m}: z(C)\in W_\rho} g(C)$$

and, when $\{C \in \mathfrak{m}, z(C) \in \mathbf{i} \cap \mathbf{W}_\rho\}$ is empty, we take $M_{\mathbf{i}} = -\infty$. Then we begin by observing that

$$\#\Phi_{W_\rho} = \sum_{C \in \mathfrak{m}} \mathbb{1}_{z(C) \in W_\rho} \mathbb{1}_{g(C) > v_\rho} \simeq \sum_{\mathbf{i} \in V} \mathbb{1}_{M_{\mathbf{i}} > v_\rho},$$

where the approximation $\simeq$ comes from the (LCC) condition. Indeed, thanks to this condition, we know that for each sub-cube $\mathbf{i}$, and with high probability, it is not possible to have a pair of exceedances in $\mathbf{i}$; thus the number of exceedances is close to the number of sub-cubes in which there are at least one exceedance. Now, notice that the number of exceeding sub-cubes, i.e. $\sum_{\mathbf{i} \in V} \mathbb{1}_{M_{\mathbf{i}} > v_\rho}$, is a sum of non-independent Bernoulli random variables, but which tend to be independent thanks to the (FRC) condition. Using this, and again the (LCC) condition, we have

$$\sum_{\mathbf{i} \in V} \mathbb{1}_{M_{\mathbf{i}} > v_\rho} \simeq \mathrm{Po}(\tau_\rho),$$

where τ_ρ is the expectation of the number of exceeding sub-cubes. This last approximation is a consequence of Proposition 1.1.3 by taking $X_i = \mathbb{1}_{M_{\mathbf{i}} > v_\rho}$ (the terms b_1, b_2 are easily controlled thanks to (1.2.2) and the (LCC) condition, and the term b_3 equals 0 thanks to the (FRC) condition; the neighborhood B_i appearing in Equation (1.1.3) is defined as the sub-cube $\mathbf{i}$ up to multiplicative constant). Using again the (LCC) condition, we can easily prove that τ_ρ is close to τ, which gives

$$\#\Phi_{W_\rho} \simeq \mathrm{Po}(\tau).$$

1.2.2 STIT and Poisson line tessellations

This section deals with extremes for stationary STIT and Poisson line tessellations in $\mathbf{R}^2$. We consider a problem which is similar to the one presented in Section 1.2.1. This time, we only investigate the case where the geometric characteristic is the inradius, i.e. the radius of the largest ball included in the cell, since it is one of the rare characteristics for which the distribution concerning the typical cell $\mathcal{C}$ can be made explicit. Indeed, according to Lemma 3 in [77] (resp. Theorem 10.4.6 in [90]), the random variable $R(\mathcal{C})$ has an exponential distribution with explicit parameter, where $R(\mathcal{C})$ denotes the inradius of the typical cell of a STIT (resp. Poisson line) tessellation. The fact that the typical cells have the same inradius in distribution (when the intensities of the STIT and Poisson line tessellations are equal) is not surprising since their interiors are the same in distribution. The results which are presented below come from the papers [28, 32].

Large inradii for STIT tessellations

To present the main result of [32], we first give some notation. Let $\mathfrak{m}_t$ be a stationary STIT tessellation at time $t > 0$, with intensity $\gamma_t = \frac{t^2}{\pi}$. For a threshold $v \geq 0$, let $N_{W_\rho}(v)$ be the number exceedances, i.e.

$$N_{W_\rho}(v) := \sum_{C \in \mathfrak{m}_t : z(C) \in W_\rho} \mathbb{1}_{R(C) > v},$$

where $W_\rho = t^{-1}\sqrt{\pi\,\rho} \cdot [-\frac{1}{2}, \frac{1}{2}]^2$ and where $z(C)$ denotes the incenter of the cell C. Given $\tau > 0$, and similarly to (1.2.2), we define a threshold $v_\rho = v_\rho(\tau)$ in such a way that the mean number of exceedances is equal to τ, i.e.

$$\mathbb{E}\left[N_{W_\rho}(v_\rho)\right] = \gamma_t\, t^{-2}\, \pi\, \rho\, \mathbb{P}\left(R(\mathcal{C}) > v_\rho\right) = \tau.$$

Because $R(\mathcal{C})$ has an exponential distribution (with parameter $2t$), the threshold can be made explicit and is equal to

$$v_\rho := v_\rho(\tau) = \frac{1}{2t}(\log \rho - \log \tau).$$

The main result of [32] is a Poisson approximation of the number of exceedances and can be stated as follows:

Theorem 1.2.2. *Let $Po(\tau)$ be a Poisson random variable with parameter τ. Then*

$$N_{W_\rho}(v_\rho) \xrightarrow[n\to\infty]{\mathcal{D}} Po(\tau). \tag{1.2.5}$$

In particular, the above result gives the asymptotic distributions of the largest order statistics. For instance, if we denote by M_{W_ρ} the maximum of inradii, we have $\mathbb{P}\left(M_{W_\rho} \leq v_\rho\right) \underset{\rho\to\infty}{\longrightarrow} e^{-\tau}$ and therefore (by taking $\tau = e^{-x}$),

$$\mathbb{P}\left(M_{W_\rho} \leq \frac{1}{2t}\log\rho + \frac{1}{2t}x\right) \underset{\rho\to\infty}{\longrightarrow} e^{-e^{-x}},$$

for any $x \in \mathbf{R}$. This last property is classical since it shows that the maximum of inradii belongs to the domain of attraction of a Gumbel distribution. Theorem 1.2.2 can be expressed in terms of total variation distance and a rate of convergence can be made explicit.

Although Theorem 1.2.1 is a general result, Theorem 1.2.2 is not a consequence of the latter since the (FRC) condition is not satisfied. However, the methods which are used to prove these theorems are similar: each one is based on a discretization of the window (see p. 23) and on a Poisson approximation result due to Arratia *et al.* (Proposition 1.1.3). The main difficulty to prove Theorem 1.2.2 is to deal with the term b_3 (as defined in (1.1.3)). Indeed, while this term is equal to 0 when the random tessellation satisfies the (FRC) condition, this property fails in the case of a STIT tessellation. The reason is that the STIT tessellation only satisfies a β-mixing [76], which is weaker than a finite range condition.

We describe below the main ideas to deal with b_3. First, we construct a family of neighborhoods $B_{\mathbf{i}}$, appearing in (1.1.3), with a suitable size (not too big if not b_2 is too large, and not too small if not b_3 is too large). Then, as a key argument, we use the fact that the STIT tessellation has a mixing property. However, the general upper bound for the β-mixing coefficient provided in [76] is not sufficient for our purposes. A more specific treatment of rare events has to be developed. To do it, we use the concept of *encapsulation*, which was introduced by Martinez and Nagel [75]. Given two convex polygons K, K' such that $0 \in K' \subset K$, it means that there is a state of the STIT process $\mathbf{m} = (\mathfrak{m}_t, t > 0)$ such that all facets of K' are separated from the facets of K by facets of the tessellation before the interior of K' is divided by a facet of the tessellation. Formally, denoting the 0-cell by C_t^0, i.e. the cell of $\mathfrak{m}_t$ that contains the origin, we define the encapsulation time as

$$S(K, K') := \inf\{t > 0 : K' \subset C_t^0 \subset K^\circ\},$$

with the convention $\inf \emptyset = \infty$. Roughly, it means that if we control the encapsulation time, then extreme events appearing in K' (approximately the latter is the sub-cube $\mathbf{i}$) tend to be independent of extreme events appearing outside K (approximately it is the neighborhood $B_{\mathbf{i}}$); which implies that b_3 is small.

Smallest and largest inradii for a Poisson line tessellation

In [28], we investigate the smallest and largest order statistics for the inradius in a planar Poisson line tessellation. In particular, we establish Poisson approximations for the number of cells with inradii larger than a large threshold (largest order statistics) and for the number of cells with inradii smaller than a small threshold (smallest order statistics). Here again, we cannot apply Theorem 1.2.1 because the Poisson line tessellation does not satisfy the (FRC) condition. The main difficulty is that cells can have a line in common even if they are far from each others. Besides, it seems that the proof which we used to derive Theorems 1.2.1 and 1.2.2 cannot be adapted since it requires to deal with the term b_3 (see Equation (1.1.3)).

In Theorem 1.1. *(ii)* in [28], we also establish a Poisson approximation for the largest inradii. The result is similar to Theorem 1.2.2 but is based on a different method, namely the method of moments. The computation of the moments of the number of exceedances is highly technical since we discuss the number of lines which are in common between cells exceeding the threshold. As opposed to [20, 23, 32], we cannot express our results in terms of total variation distance with explicit rate of convergence. In Theorem 1.1. *(i)* in [28], we establish a Poisson approximation for the smallest order statistics. The main idea is to apply a result due to Schulte and Thäle on U-statistics (Theorem 1.1 in [92]). In complement, we also show that cells with a small inradius are triangles with high probability. More precisely, we get the following theorem (with $W_\rho = \rho^{1/d}[-\frac{1}{2}, \frac{1}{2}]^d$):

Theorem 1.2.3. *Let $r \geq 1$ be fixed. Let $C_{W_\rho}[r]$ be the cell, with incenter in W_ρ, such that $R(C_{W_\rho}[r])$ is the r-th smallest inradius for cells with incenters in W_ρ. Let $n(C_{W_\rho}[r])$ be its number of vertices. Then*

$$\mathbb{P}\left(\bigcap_{k=1}^{r}\left\{n\left(C_{W_\rho}[k]\right)=3\right\}\right) \underset{\rho\to\infty}{\longrightarrow} 1.$$

1.2.3 Large k-th nearest neighbor balls

In [29], we consider the following problem. Let (S,ρ) be a metric space and let $X_1,\dots,X_n$ be a sequence of i.i.d. random variables with distribution μ. Given (fixed) $k \geq 1$, and under suitable assumptions on the metric ρ, there is a.s. a unique k-th nearest neighbor of X_i among $\{X_1,\dots,X_n\}\setminus\{X_i\}$, for each $i \leq n$, and we denote by $R_{i,n,k}$ the distance of this point to X_i. The main topic of [29] is to deal with the largest values of $\mu(B(X_i,R_{i,n,k}))$ as n goes to infinity, where $B(x,r)$ denotes the closed ball centered at x with radius r. To do it, we introduce for fixed $t \in \mathbf{R}$, the following threshold:

$$v_{n,k} := v_{n,k}(t) := \frac{t+\log n+(k-1)\log\log n-\log(k-1)!}{n}, \tag{1.2.6}$$

Roughly, $v_{n,k}$ is the order of the maximum of the measures of the balls. In some sense, it is universal since it does not depend neither on the metric space (S,ρ) nor on the measure μ. In the same spirit as Sections 1.2.1 and 1.2.2, we investigate the asymptotic behaviour of the number of exceedances, namely

$$N_{n,k} := \sum_{i=1}^{n} \mathbb{1}_{\mu(B(X_i,R_{i,n,k}))>v_{n,k}}.$$

The study of extremes of k-th nearest neighbor balls is classical in stochastic geometry, and it has various applications, see e.g. [82]. As a first result, we prove that the mean number of exceedances, i.e. $\mathbb{E}[N_{n,k}]$, converges to e^{-t} as n goes to infinity, with explicit rate of convergence.

When the metric space (S, ρ) is the Euclidean space $\mathbf{R}^d$ endowed with its usual norm, we can derive a Poisson approximation of $N_{n,k}$. More precisely, under suitable assumptions on μ (in particular, μ has to admit a density w.r.t. the Lebesgue measure in $\mathbf{R}^d$ with compact support, say $[0,1]^d$), we get the following result:

$$d_{TV}\left(N_{n,k}, \mathrm{Po}(e^{-t})\right) = O\left(\frac{\log\log n}{\log n}\right), \tag{1.2.7}$$

where d_{TV} denotes the total variation distance.

Equation (1.2.7) is a a generalization of a result due to Györfi *el al.* (see Theorem 2.2 in [51]) since it holds for any integer $k \geq 1$ whereas [51] only deals with the case $k = 1$. Moreover (1.2.7) is more precise in the sense that, as opposed to [51], we make explicit the rate of convergence for the Poisson approximation.

1.3 Extremal index for random tessellations

1.3.1 A new characterization of the extremal index

Similarly to Section 1.2, we consider a stationary random tessellation $\mathfrak{m}$ in $\mathbf{R}^2$, a geometric characteristic g and a window $W_\rho = \rho^{1/d}[-\frac{1}{2}, \frac{1}{2}]^d$. As stated in Theorem 1.2.1, under a finite range condition (FRC) and a local correlation condition (LCC), the point process of exceedances, i.e.

$$\Phi_{W_\rho}(\tau) = \rho^{-1/d}\{z(C) \in W_\rho :\ C \in \mathfrak{m} \text{ and } g(C) > v_\rho(\tau)\},$$

converges to a homogeneous Poisson point process, where $v_\rho(\tau)$ is a threshold satisfying (1.2.2). The (LCC) condition ensures that the exceedances are isolated. When the latter does not hold, clusters of exceedances can appear in the same spirit as they can appear for sequences of real random variables when the $D'(u_n)$ is not satisfied (see Section 1.1.2).

A natural question, in the context of random tessellations, is to investigate the mean size of a typical cluster of exceedances. To this end, we introduce in the same spirit as (1.1.4), the concept of extremal index as follows. We say that θ is the *extremal index* if, in conjunction to (1.2.2), we have

$$\mathbb{P}\left(M_{W_\rho} \leq v_\rho(\tau)\right) \underset{\rho\to\infty}{\longrightarrow} e^{-\theta\tau},$$

where M_{W_ρ} denotes the maximum of the characteristic g over all cells with nucleus in W_ρ (see Equation (1.2.1)). Here again θ is interpreted as the reciprocal of the mean size of a cluster of exceedances.

In [33], we provide a new characterization of the extremal index and of the asymptotic cluster size distribution. This characterization is based on the Palm version of the point process of exceedances. Roughly, given any Borel subset $B \subset W_\rho$, the Palm version Φ_B^0 of the point process of exceedances observed in B is defined as the (normalized) set of nuclei of cells exceeding the threshold $v_\rho(\tau)$ *conditional on* the fact that the origin is the nucleus of a cell and that this cell (which, in distribution, is the typical cell) is an exceedance. Then we introduce a distribution $(p_{k,B}(\tau))$ as follows:

$$p_{k,B}(\tau) := \mathbb{P}\left(\#\Phi_B^0(\tau) = k\right), \quad k \geq 1.$$

The quantity $p_{k,B}(\tau)$ can be interpreted as the probability that there are k exceedances in B conditional on the fact that the origin is a nucleus and that the cell with nucleus at the origin is an

exceedance. In the particular case where $\mathfrak{m}$ is a Voronoi tessellation based on a stationary Poisson point process η, and thanks to the Mecke-Slivnyak formula, the probability can be expressed as:

$$p_{k,B}(\tau) = \mathbb{P}\left(\#\Phi_B^{\eta\cup\{0\}}(\tau) = k | g(C_{\eta\cup\{0\}}(0)) > v_\rho(\tau)\right),$$

where

$$\Phi_B^{\eta\cup\{0\}}(\tau) = \rho^{-1/d}\{x \in (\eta\cup\{0\}) \cap B : g(C_{\eta\cup\{0\}}(x)) > v_\rho(\tau)\}.$$

The main result of [33] (Theorem 4) claims that the point process of exceedances converges to a homogeneous compound Poisson point process and can be stated as follows.

Theorem 1.3.1. *Let B_ρ be a cube with volume $(\log\log\rho)^{\log\log\rho}$. Assume that the following limit exist:*

$$p_k := \lim_{\rho\to\infty} p_{k,B_\rho}(\tau_0),$$

for any $k \geq 1$ and for some τ_0. Then, under mild additive assumptions (including the existence of an extremal index $\theta \in (0,1]$), for any $\tau > 0$,

$$\Phi_{W_\rho}(\tau) \xrightarrow[n\to\infty]{\mathcal{D}} \Phi(\tau),$$

where $\Phi(\tau)$ is a homogeneous compound Poisson point process in $[-\frac{1}{2},\frac{1}{2}]^d$ of intensity $\theta\tau$ with cluster size distribution $\pi = (\pi_k)$, where

$$\theta = \sum_{k=1}^{\infty} k^{-1} p_k \text{ and } \pi_k = \frac{p_k}{k\theta}.$$

In the same spirit as (1.1.5), we can show that, for any $k \geq 1$,

$$\pi_k = \lim_{\rho\to\infty} \mathbb{P}\left(\#\Phi_{B_\rho}(\tau_0) = k \,\middle|\, \#\Phi_{B_\rho}(\tau_0) > 0\right),$$

where B_ρ is as in Theorem 1.3.1. Similarly to Section 1.1.2, the above expression can be seen as a *blocks characterization* of the cluster size distribution, whereas the expression $\pi_k = \frac{p_k}{k\theta}$ can be seen as a *Palm characterization*.

Our theorem provides a new expression of the extremal index: this index was previously interpreted as the reciprocal of the mean of the cluster size distribution π. From now on, it can be viewed as the mean of the reciprocal of the Palm version of the cluster size.

Theorem 1.3.1 has also a practical interest. Indeed, in general the distribution π and the value of θ cannot be made explicit. It is necessary to use simulations to compute approximate values of these quantities. In Section 1.1.2, in the context of sequences of real random variables, we have seen that estimators of these quantities can be based on the blocks (Equation (1.1.5)) and runs (Equation (1.1.6)) characterizations. The runs'one cannot be adapted in our context since there is no natural order in $\mathbf{R}^d$. The blocks method competes with the Palm approach. The idea of the Palm approach is to consider clusters close to the origin given that the cell whose nucleus is the origin has an exceedance. Our approach (Palm characterization) provides better approximations of the extremal index and the cluster size distribution and requires less simulations. Indeed, it is sufficient to simulate the random tessellation only on blocks that contain at least one exceedance (the one with nucleus the origin), while with the blocks approach, it is necessary to simulate a very large number of blocks (including those without any extreme value). More precisely, in [33], we give numerical illustrations in $\mathbf{R}^2$ by simulating tessellations only observed in the square $[-173,173]^2$ to approximate θ and $p = (p_k)$ thanks to our Palm approach. A blocks approach would have required to simulate tessellations in the square $[-5.18\cdot10^{21}, 5.18\cdot10^{21}]$ for the same accuracy, which is practically impossible.

1.3.2 Numerical illustrations

In [33], we illustrate Theorem 1.3.1 throughout simulations for three geometric characteristics for which the value of the extremal index is known or can be conjectured (see [24] for examples of computations of extremal indices). For sake of simplicity, the simulations are done in the particular setting $d = 2$. We provide approximations of $p_1, \dots, p_9$ and of the extremal index by using the fact that $\theta = \sum_{k=1}^{\infty} k^{-1} p_k$ and we compare this approximation to the theoretical value of θ.

For each geometric characteristic g, we proceed as follows. We take $\tau = 1$ and $\rho = \exp(100)$. In particular, the cube B_ρ, as considered in Theorem 1.3.1, is approximatively

$$B_\rho \simeq [-173, 173]^2.$$

Then, we compute theoretically $v_\rho(1)$ so that $\rho \cdot \mathbb{P}\left(g(\mathscr{C}) > v_\rho(1)\right) \underset{\rho\to\infty}{\longrightarrow} 1$. We simulate 10000 realizations of independent Poisson-Voronoi tessellations given that the typical cell is an exceedance, i.e. $g(C_{\eta\cup\{0\}}(0)) > v_\rho(1)$. This sample of size 10000 is divided into 100 sub-samples of size 100. For each $1 \leq i \leq 100$ and for each $1 \leq k \leq 9$, we consider the empirical mean $\hat{p}_k^{(i)}$ of p_k, i.e. the mean number of realizations in which there exist exactly k Voronoi cells with nucleus in $B_\rho \simeq [-173, 173]$ and such that the geometric characteristic is larger than $v_\rho(1)$. We summarize our empirical results by box plots associated with the empirical values $(\hat{p}_k^{(i)})_{1\leq i\leq 100}$. The three examples that we consider are the inradius, the reciprocal of the inradius and the circumradius for a Poisson-Voronoi tessellation (another numerical example in [33] concerns the large circumradii of a Poisson-Delaunay triangulation but is not described below).

Inradius Recall that the inradius of a Voronoi cell with nucleus x is defined as the radius of the largest ball included in the cell and centered at x. It equals the half distance of x to its nearest neighbor. When we consider the largest values of the inradii, the extremal index is $\theta = 1$ and the cluster size distribution is $\pi = p = \delta_1$ (this can be seen as a consequence of Theorem 1.2.1). The left part of Figure 1.3 is a simulation of a Poisson-Voronoi tessellation given that the inradius of the typical cell $C_{\eta\cup\{0\}}(0)$ is larger than $v_{\exp(100)}(1) \simeq 2.82$. As observed in this figure, there is no cell with a large inradius, excepted the typical cell. This confirms that the cluster of exceedances are of size 1, i.e. $p_1 = 1$ and $\theta = 1$. The right part of Figure 1.3 provides the box plots of the empirical distributions. In particular, for all simulations, there is always exactly one cell with a large inradius.

Reciprocal of the inradius A second example deals with the large values of the reciprocal of the inradii for a Poisson-Voronoi tessellation in $\mathbf{R}^2$. Equivalently, this consists of the small values of the inradii. As observed in [24], the extremal is $\theta = 2$ and the cluster size distribution is $\pi = p = \delta_2$. This fact can be explained by a trivial heuristic argument: if a cell minimizes the inradius, one of its neighbors has to do the same (see also the left part of Figure 1.4). Moreover, we can easily prove that the probability that there is more than one such a cell is negligible. The left part of Figure 1.4 provides a realization of a Poisson-Voronoi tessellation when the inradius of the typical cell is lower than $v_{\exp(4)}(1) \simeq 0.0381$ (we have taken the threshold $v_{\exp(4)}(1)$ instead of $v_{\exp(100)}(1)$ for convenience). The right part of Figure 1.4 provides the box plots of the empirical distributions. In particular, for all simulations, there are always exactly two cells with a small inradius.

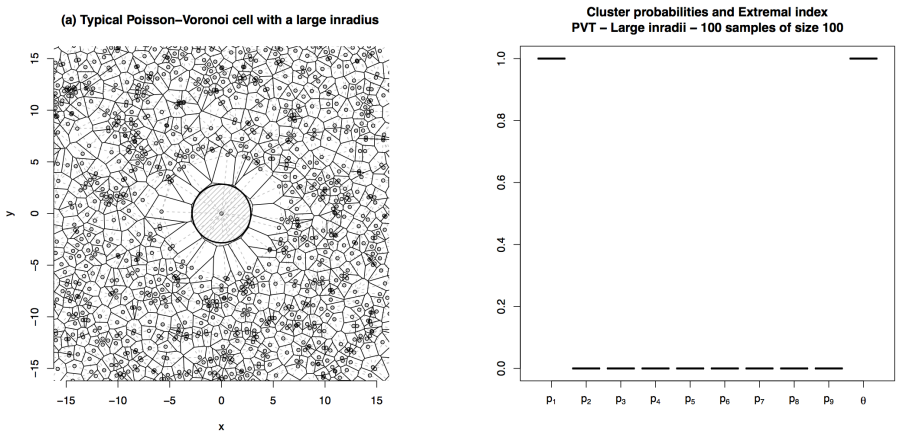


Figure 1.3: Large inradius for a Poisson-Voronoi tessellation

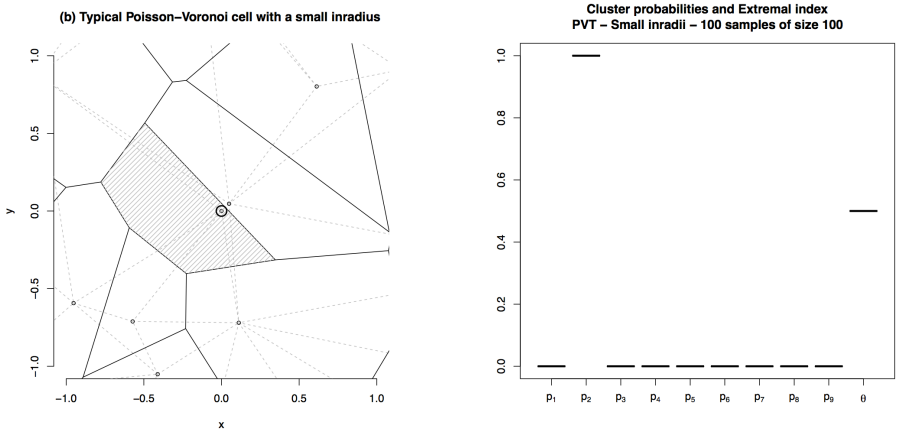


Figure 1.4: Small inradius for a Poisson-Voronoi tessellation

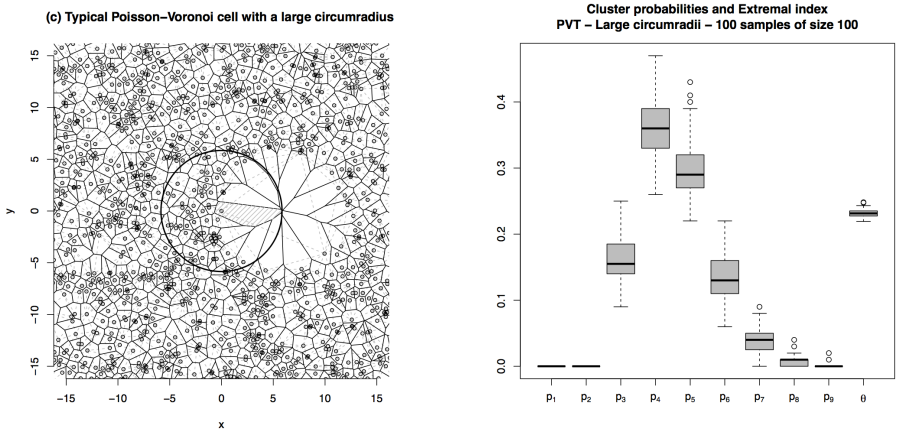


Figure 1.5: Large circumradius for a Poisson-Voronoi tessellation

Circumradius Recall that the circumradius of a Voronoi cell with nucleus x is defined as the radius of the smallest ball containing the cell and centered at x. Using a conjecture on the tail of the circumradius of the typical cell [17] and our result on the asymptotic behaviour of the maximum of circumradii (Equation (2.c) in [20]), we conjecture in [33] that the extremal index is $\theta = 1/4$. On the left part of Figure 1.5, we provide a simulation of the Palm version of the Poisson-Voronoi tessellation, given that the circumradius of the typical cell is larger than $v_{\exp(100)}(1) \simeq 5.81$. The size of a cluster of exceedances is random. On the right part of Figure 1.5, we provide the box plots of the empirical distributions. This time, the empirical distributions of the cluster size probabilities are not degenerated for $k = 3, \ldots, 9$, and their interquartile ranges are quite large for $k = 3, 4, 5$. We also notice that the empirical value of the extremal index is very concentrated around a value close to $1/4$.

1.4 Extremes on the Delaunay graph

1.4.1 The maximal degree

In this section, we present the paper [14]. Let η be a stationary Poisson point process of intensity 1 in $\mathbf{R}^d$ and let $W_\rho = [-\frac{1}{2}, \frac{1}{2}]^d$. With each node $x \in \eta$, we associate the degree of x, say $d_\eta(x)$, in the Delaunay graph induced by η. The main problem is to investigate the asymptotic behaviour of the maximum of the degrees over all nodes in W_ρ, namely

$$\Delta_{W_\rho} = \max_{x \in \eta \cap W_\rho} d_\eta(x),$$

as ρ goes to infinity. This question is related to the one considered in the previous sections. However, the main difference is that the maximum that we deal with concerns *discrete* random variables. The results which are obtained are radically different since it is not possible to find a threshold in such a way that (1.2.2) holds.

The maximal degree of random combinatorial graphs has been extensively investigated, but much less has been done when the vertices are given by a point process and the edges are built according to geometric constraints. One of the first results on the maximal degree in a

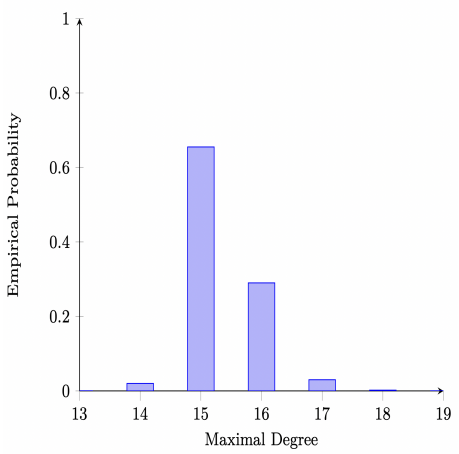


Figure 1.6: Empirical distribution of Δ_{W_ρ}, based on 75000 simulations, of the maximal degree in a planar Poisson-Delaunay graph observed in the window $W_{10^6} = 10^3[0,1]^2$.

Poisson-Delaunay graph was due to Bern *et al.* (see Theorem 7 in [8]) who showed that

$$\mathbb{E}\left[\Delta_{W_\rho}\right] = \Theta\left(\frac{\log \rho}{\log\log \rho}\right) \tag{1.4.1}$$

in any dimension $d \geq 2$. More recently, Broutin *et al.* [16] provided a new bound for Δ_{W_ρ} in the following sense: when $d = 2$, with probability tending to 1, the maximal degree Δ_{W_ρ} is less than $(\log \rho)^{2+\xi}$, for any fixed $\xi > 0$. The main result of [14] significantly improves these two results in dimension two and is stated below.

Theorem 1.4.1. *There exists a deterministic function $\rho \mapsto I_\rho$, $\rho > 0$, with values in $\mathbf{N} = \{1, 2, \ldots\}$, such that*

(i) $\mathbb{P}\left(\Delta_{W_\rho} \in \{I_\rho, I_\rho + 1\}\right) \underset{\rho\to\infty}{\longrightarrow} 1$;

(ii) $I_\rho \underset{\rho\to\infty}{\sim} \frac{1}{2} \cdot \frac{\log \rho}{\log\log \rho}$.

Our result provides the exact order of the maximal degree and claims that, with high probability, the maximal degree is concentrated on two consecutive values. As observed in Figure 1.6, the concentration is already visible for $\rho = 10^6$.

A similar result has been established in the context of i.i.d. discrete real random variables. More precisely, Anderson [1] was the first one to prove that the maximum of the first n terms is concentrated, with high probability as n goes to infinity, on two consecutive integers for a wide class of discrete random variables. Many results of this type have also been established for maxima of degrees in the context of random combinatorial graphs. For random geometric graphs, it seems that only one has been stated in this way (see Theorem 6.6. in [81]). Two difficulties are added in the context of Poisson-Delaunay graphs. The first one is that the distribution of the typical degree cannot be made explicit. The second one, which constitutes the main difficulty, comes from the dependence between the degrees of the nodes and the geometric constraints in the Poisson-Delaunay graph.

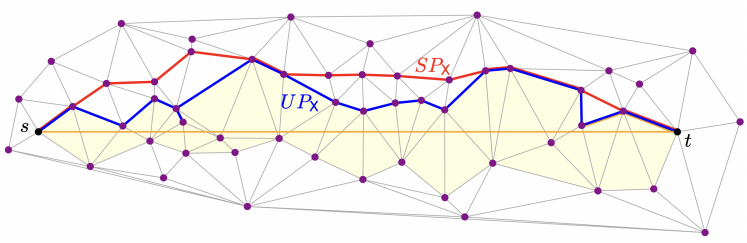


Figure 1.7: The shortest path $S_{s,t}(\chi)$ (red) and the upper path $U_{s,t}(\chi)$ (blue).

To prove Theorem 1.4.1, we first provide logarithmic estimates of the probability that the typical degree equals a large value. The latter are very similar to two results in [13] (Theorem 1.1 and Theorem 1.2) in which estimates for the distribution of the typical number of facets in a Poisson hyperplane tessellation are given. Then to deal with the dependence structure of the Poisson-Delaunay graph, we introduce a dependency graph in the same spirit as Avram and Bertsimas [4] did for deriving central limit theorems. The proof strongly uses the fact that the Delaunay graph in $\mathbf{R}^2$ is planar. As an intermediate result to derive Theorem 1.4.1, we prove that, with high probability, there is no family of five nodes in the Poisson-Delaunay graph which are close to each other, such that their degrees simultaneously exceed I_ρ. Such a result is essential in our proof and is specific to the two dimensional case. Surprisingly, the proof of Theorem 1.4.1 also shows that we can find arbitrary large windows for which the maximal degree is concentrated on only one integer with high probability.

An extension of Theorem 1.4.1 can be done in higher dimension (see Theorem 3 in [14]) in the following sense: with high probability, the maximum of the degrees in $\mathbf{R}^d$ is concentrated on a finite deterministic number (only depending on d) of consecutive values. However, as opposed to Theorem 1.4.1, we think that this extension is not optimal since we conjecture that it should also be concentrated only on *two* consecutive integers.

1.4.2 The stretch factor

In [27], we investigate the length of the smallest between two fixed nodes in a planar Delaunay graph induced by some set of points χ. By a path $P = P(s,t)$ between two nodes $s, t \in \chi$, we mean a sequence of edges $[Z_0, Z_1], [Z_1, Z_2], \ldots, [Z_{k-1}, Z_k]$ in the Delaunay graph, such that $Z_0 = s$, $Z_k = t$ (see Figure 1.7).

The investigation of paths is related to walking strategies which are commonly used to find the triangle containing a query point in a planar triangulation [38] or routing in geometric networks [15]. A classical problem is the study of the stretch factor associated with two nodes s, t. This quantity is defined as the length of the smallest path, say $\ell(S_{s,t}(\chi))$, divided by the Euclidean distance $|s - t|$. Many upper bounds were established for the stretch factor in the context of finite sets χ, see e.g. [40]. The best upper bound established until now for deterministic finite sets χ is due to Xia [98] who proves that the stretch factor is lower than 1.998. For the lower bound, Xia and Zhang [99] find a configuration of points χ such that the stretch factor is greater than 1.593.

In [27], we focus on a probabilistic version of the problem by taking a slight modification of the underlying point process. More precisely, we fix two points, say $s = (0, 0)$ and $t = (1, 0)$ and we consider a homogeneous Poisson point process η_n of intensity n in $\mathbf{R}^2$. We investigate the stretch factor between s and t, i.e. the length of the smallest path, when the underlying set of

nodes is $\chi = \eta_n \cup \{s,t\}$. Our main result consists of bounds, in expectation, of $\ell(S_{s,t}(\chi))$ as n goes to infinity and can be stated as follows.

Theorem 1.4.2. *When $\chi = \eta_n \cup \{s,t\}$, with $s = (0,0)$ and $t = (1,0)$, we have*

$$1 + 7 \cdot 10^{-9} \leq \lim_{n\to\infty} \mathbb{E}\left[\ell(S_{s,t}(\chi)\right] \leq \frac{35}{3\pi^2} \simeq 1.182.$$

The existence of the limit comes from subadditivity arguments. Our bounds are from optimal since simulations suggest that $\lim_{n\to\infty} \mathbb{E}\left[\ell(S_{s,t}(\chi)\right] \simeq 1.04$. However, our result provides the first (and seemingly the only one) non-trivial lower bound, i.e. strictly larger than 1, for the stretch factor when the intensity of the underlying Poisson point process goes to infinity. In parallel to our work, Hirsch *et al.* (Theorem 26 in [55]) also prove that the stretch factor is non-trivial, for a larger class of random graphs, but their technique cannot provide an explicit lower bound for the stretch factor.

The difficulties for obtaining the two bounds in Theorem 1.4.2 are radically different. The upper bound is easy: it consists in building an auxiliary path and in estimating its length. This path, referred to as the *upper path* $U_{s,t}(\chi)$, is defined as the sequence of all edges in $\mathbf{R}\times\mathbf{R}_+$ which belong to Delaunay triangles that intersects $[s,t]$ (see Figure 1.7). Because $U_{s,t}(\chi)$ is defined locally, it is not hard to compute the expectation of its length by applying classical formulas of Stochastic and Integral Geometry, namely the Mecke-Slivnyak formula (see e.g. Theorem 3.3.5 in [90]) and the Blaschke-Petkantschin type change of variables (see e.g. Theorem 7.3.1 in [90]). The lower bound in Theorem 1.4.2 is much more delicate since it requires to deal with directly the smallest path. Investigating the latter is difficult because it is not defined locally and long range dependence properties occur: the length strongly depends on the points s and t. The main idea to deal with is to discretize the plan into squares, called pixels, with a suitable size. Then, on each pixel, we consider a so-called *horizontal property* which roughly claims that, in the pixel, paths are almost horizontal. We prove that there is a non-negligeable proportion of pixels which intersect the smallest path and which do not have such a property. Then, on these pixels, we provide a lower bound for the smallest path. However, the main difficulty is that the properties of pixels intersecting the smallest path are not independent, although when the latter are far.

Chapter 2

Two problems in Topological Data Analysis

Sommaire

2.1 Introduction

2.1.1 Concepts of Topological Data Analysis

Topological Data Analysis (TDA) is a recent field of research which borned with the pioneering works of Edelsbrunner et al. [47] and Zomorodian and Carlsson [100] in Persistent Homology. The simple and effective idea is to leverage invariants from Algebraic Topology to extract insights from datasets. The latter are often represented as point clouds in Euclidean or more general metric spaces. TDA has a wide variety of applications such as astronomy, biology, finance and materials science. We recall below some notions of TDA. The description which we give is mainly inspired from a survey of Chazal and Michel [22].

Simplicial complexes A classical object in Algebraic Topology is the simplicial complex. The latter is a generalization of the notion of graph and can be associated with some topological space. Given a set $X = \{x_0, \dots, x_k\}$ of $k+1$ points of $\mathbf{R}^d$ which are affinely independent, the k-dimensional *simplex* $\sigma = [x_0, \dots, x_k]$ spanned by X is defined as the convex hull of X. A *geometric simplicial complex* K in $\mathbf{R}^d$ is a collection of simplices such that

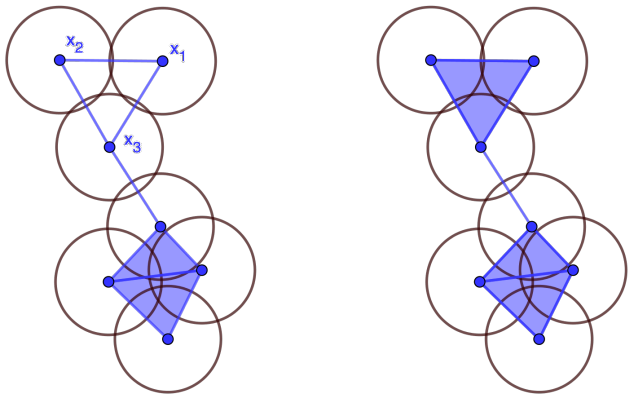


Figure 2.1: *(a)* Čech complex; *(b)* Vietoris-Rips complexes.

(i) any face of a simplex of K is a simplex of K;

(ii) the intersection of any two simplices of K is either empty or a common face of both.

Given a set V, an *abstract simplicial complex* with vertex set V is a set K of finite subsets of V such that the elements of V belongs to K and for any $\sigma \in K$ any subset of σ belongs to K. The elements of K are called the *faces* or the *simplices* of K. The dimension of an (abstract) simplex is defined as its cardinality minus 1 and the dimension of K is the largest dimension of its simplices. In particular, a simplicial complex of dimension 1 is a graph. Any geometric simplicial complex is an abstract simplicial complex. The reciprocal is in some sense true since an abstract simplicial complex can be associated with a geometric complex whose combinatorial description is the same.

Among the most usual (abstract) simplicial complexes are the Vietoris-Rips and the Čech complexes. To introduce them, we consider a set of points in $\mathbf{R}^d$, say X, and a real number $r \geq 0$.

- The *Vietoris-Rips complex* $\mathrm{Rips}_r(X)$ is the set of simplices $[x_0, \ldots, x_k]$ such that $B(x_i, r) \cap B(x_j, r) \neq \emptyset$ for all $0 \leq i, j \leq k$.

- The *Čech complex* $\check{\mathrm{C}}\mathrm{ech}_r(X)$ is the set of simplices $[x_0, \ldots, x_k]$ such that $\bigcap_{i \leq k} B(x_i, r) \neq \emptyset$.

In particular, any element in the Čech complex is an element of the Vietoris-Rips complex but the reciprocal is not true (see Figure 2.1).

An important result of Algebraic Topology is the so-called nerve theorem (see e.g. Theorem 1 in [22]). The latter implies that, in terms of homotopic equivalence, the Čech complex is the same object as the union of the balls $\bigcup_{x \in X} B(x, r)$. In particular, the topological invariants as defined below are the same.

Homology Among the most classical topological invariants are the Betti numbers. Loosely speaking, they capture the number of k-dimensional holes of the investigated structure. To define them, we first recall some notion of Homology. Roughly, given a simplicial complex, the k-dimensional holes, $k \geq 0$, are represented by a vector space H_k whose dimension is intuitively the number of independent k-holes. For example the 0 (resp. 1)-dimensional homology group H_0 (resp. H_1) represents the connected components (resp. the loops) of the simplicial complex.

To introduce the homology groups, we proceed as follows. Let K be a simplicial complex and let $\{\sigma_1, \ldots, \sigma_p\}$ be the set of k-simplices in K, where $k \geq 0$ is fixed. Define the space of *k-chains*, say C_k, as the collection of families of k-simplices. This space can be seen as the set of formal linear combinations of k-simplices with coefficients in $\mathbf{Z}/2\mathbf{Z}$, i.e.

$$C_k = \left\{ \sum_{i=1}^{p} \lambda_i \sigma_i,\ \lambda_i \in \mathbf{Z}/2\mathbf{Z} \right\}.$$

Endowed with an internal sum and an external product, the set C_k is a $\mathbf{Z}/2\mathbf{Z}$-vector space with dimension p. Now, define an operator $\partial_k : C_k \to C_{k-1}$ as follows. First, for any k-simplex $\sigma = [x_0, \ldots, x_k]$, we let

$$\partial_k(\sigma) = \sum_{i=0}^{k} [x_0, \ldots, \hat{x}_i, \ldots, x_k],$$

where $[x_0, \ldots, \hat{x}_i, \ldots, x_k] = [x_0, \ldots, x_{i-1}, x_{i+1}, x_k]$ and $[\hat{x}_0, x_1, \ldots, x_k] = [x_1, \ldots, x_k]$. In particular, the quantity $\partial_k(\sigma)$ is a $(k-1)$-chain. By linearity, the function ∂_k can be extended as a linear operator defined on the space C_k and with values in C_{k-1}. The kernel of the operator $\partial_k : C_k \to C_{k-1}$, say $Z_k = \mathrm{Ker}(\partial_k)$, is referred to as the space of *k-cycles* whereas the image of the operator $\partial_{k+1} : C_{k+1} \to C_k$, say $B_k = \mathrm{Im}(\partial_{k+1})$, is referred to as the space of *k-boundaries*. An important property is:

$$\partial_k \circ \partial_{k+1} = 0.$$

As a consequence, B_k is included in Z_k. Roughly, a k-cycle can be seen a the boundary of a $(k+1)$-dimensional solid whereas a k-boundary can be seen as the boundary of a $(k+1)$-dimensional solid which is a $(k+1)$-chain. The *k-th homology group* of the simplicial complex K is defined as the quotient space

$$H_k = \frac{Z_k}{B_k}.$$

The above set is a $\mathbf{Z}/2\mathbf{Z}$-vector space. Its dimension, denoted by

$$\beta_k = \dim H_k = \dim Z_k - \dim B_k,$$

is called the *k-th Betti number*. As an example, in Figure 2.1, *(a)*, we have $Z_1 \simeq (\mathbf{Z}/2\mathbf{Z})^3$, $B_1 \simeq (\mathbf{Z}/2\mathbf{Z})^2$ and $\beta_1 = 1$. This last equality has a clear intuition: the Betti number β_1 equals 1 because there is only one 1-dimensional hole in the simplicial complex (the one whose the boundary is given by the edges $[x_1, x_2]$, $[x_2, x_3]$ and $[x_3, x_1]$). Besides, in Figure 2.1, *(b)*, we have $Z_1 \simeq (\mathbf{Z}/2\mathbf{Z})^3$, $B_1 \simeq (\mathbf{Z}/2\mathbf{Z})^3$ and $\beta_1 = 0$.

Persistent Homology Persistent Homology is a powerful tool for computing topological features of a space at different spatial resolutions. In particular, it encodes the evolution of the

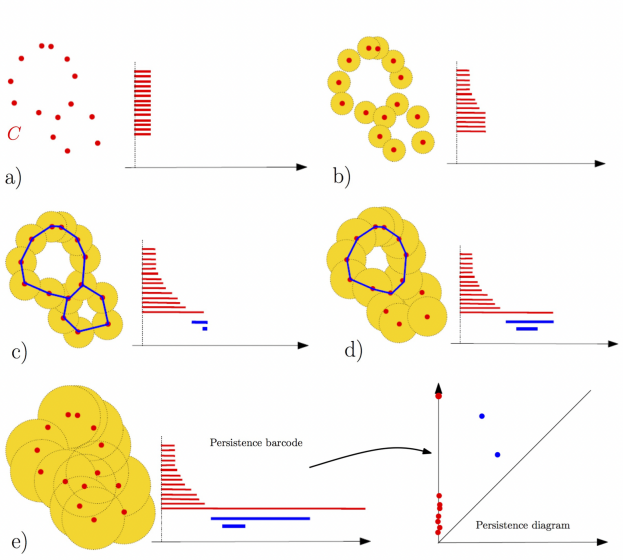


Figure 2.2: Lifetimes for connected components (red) and loops (blue); the figure comes from a survey of Chazal and Michel [22].

homology groups of the nested complexes across the scales. The main motivation is to detect the true features of the underlying space and to recognize artifacts of sampling or noise.

Figure 2.2 comes from a survey of Chazal and Michel [22] and illustrates this concept. Let us describe in few words the mechanism. Part a) depicts a point cloud (a dataset). Roughly, it seems that there are two loops. To detect them more rigorously, the main idea is to draw balls, centered at each point of the dataset with the same radius, during a time interval. When the radius grows, features such as loops, appear and disappear. For instance, Part c) gives a time at which two loops are alive whereas Part d) gives another one at which one loop is still alive and the other one is dead. The birth and death times of such features can be represented through a so-called *barcode*. For instance, in Part e), the blue bars stand for the lifetimes of the two loops appearing in the process. Birth and death time can also be represented through another diagram, called the *persistence diagram*. Roughly, a feature is represented through a point whose the x-coordinate is the birth time of the feature and the y-coordinate is its deathtime. Observing features during a time interval (which is the concept of Persistent Homology), and not at a fixed time (which only concerns Homology), is important. Indeed, an observation based on a time interval allows us to detect the features which are the most persistent, i.e. which live for a long time, and to detect on the opposite the one which are not relevant (for example, the one which come from some noise in the dataset).

To describe more rigorously the above concept, we introduce the notion of persistent homology groups as follows. Let K be a simplicial complex and let $(K_r)_{r\in T}$, $T \subset \mathbf{R}_+$ be an increasing sequence of simplicial complexes such that $K = \bigcup_{r\in T} K_r$. The sequence $(K_r)_{r\in T}$ is referred to as a *filtration* of K. For instance, if X is a point cloud in $\mathbf{R}^d$, the sequences $(\mathrm{Rips}_r(X))_{r\in T}$ and $(\check{\mathrm{C}}\mathrm{ech}_r(X))_{r\in T}$ are filtrations, which are called the Vietoris-Rips and the Čech filtrations, respectively. Now, for any $r \in T$, denote by $Z_k(K_r)$ (resp. $B_k(K_r)$) the space of k-cycles (resp. the space of k-boundaries) of the simplicial K_r. Notice that the sequences $Z_k(K_r)$ and $B_k(K_r)$

are increasing w.r.t. r. Now, let $r \leq s$ be fixed. Similarly to the definition of the homology groups, we define the k-th *persistent homology group* at times r, s as the quotient space

$$H_k^{(r,s)} = \frac{Z_k(K_r)}{Z_k(K_r) \cap B_k(K_s)}.$$

Such a quantity is a $\mathbf{Z}/2\mathbf{Z}$-vector space. Its dimension, denoted by $\beta_k^{(r,s)} = \dim H_k^{(r,s)}$, is called the k-th *persistent Betti number* at times r, s. A (non-null) element of $H_k^{(r,s)}$ is the class of a k-persistent cycle which is alive at time r and still alive at time s. Thus, the quantity $\beta_k^{(r,s)}$ counts the number of independent k-persistent cycles with birth time in $[0, r]$ and death time in (s, ∞). The family of persistent Betti numbers can be used to define rigorously the persistent diagram. Indeed, it can be shown that there exists a unique set of points $\mathrm{PD}_k = \{(b_k^{(i)}, d_k^{(i)}) : i \geq 1\} \subset \mathbf{R}^2$ such that, for any $r \leq s$,

$$\#\mathrm{PD}_k \cap ([0, r] \times (s, \infty)) = \beta_k^{(r,s)}.$$

The set PD_k is the so-called *persistence diagram*. As mentioned previously, each element of PD_k can be interpreted as a k-dimensional feature, where the x-coordinate (resp. the y-coordinate) represents the birth (rep. death) time.

2.1.2 Main problems

We give below a short description of our works in TDA. Each of them uses concepts which have been introduced in the previous section.

In Section 2.2, we introduce tests for the goodness of fit of point patterns via methods from TDA. More precisely, the persistent Betti numbers give rise to a bivariate functional summary statistic for observed point patterns that is asymptotically Gaussian in large observation windows. We analyze the power of tests derived from this statistic on simulated point patterns. As the main methodological contribution, we derive sufficient conditions for a functional central limit theorem on bounded persistent Betti numbers of point processes with exponential decay of correlations.

In Section 2.3, we consider the Čech and the Vietoris-Rips filtrations based on a stationary Poisson point process in $\mathbf{R}^d$. We study extreme values for the lifetimes of features dying in bounded components and with birth (resp. death) time bounded away from the threshold for continuum percolation and the coexistence region. We describe the scaling of the minimal lifetimes for general feature dimensions, and of the maximal lifetimes for cavities in the Čech filtration. We establish Poisson approximation for large lifetimes of cavities and for small lifetimes of loops. We also study the scaling of minimal lifetimes in the Vietoris-Rips setting and point to a surprising difference to the Čech filtration.

2.2 Testing goodness of fit for point processes via Topological Data Analysis

2.2.1 Functional central limit theorem for persistent Betti numbers

In [9], we introduce tests for the goodness of fit of point patterns via methods from TDA. Our main theoretical result is a functional central limit theorem. To state it, we consider a stationary point process $\mathcal{P}$ in $\mathbf{R}^2$ and we let $\mathcal{P}_\rho = \mathcal{P} \cap W_\rho$, with $W_\rho = \rho^{1/2}[-\frac{1}{2}, \frac{1}{2}]^2$.

As described in Section 2.1.1, when we grow balls centered at each point of $\mathcal{P}_\rho$ with the same radius during a time interval, k-features (in the Čech filtration) appear and disappear. Since we

deal with the planar case, only the cases $k = 0$ and $k = 1$ are of interest. Here a 0-feature living at time r can be identified to a connected component of the open space $\mathcal{O}_r(\mathcal{P}_\rho) = \bigcup_{x\in\mathcal{P}_\rho} B(x, r)$ and a 1-feature (i.e. a loop) living at time r can be seen as a bounded connected component of the vacant space $V_r(\mathcal{P}_\rho) = \mathbf{R}^2 \setminus \mathcal{O}_r(\mathcal{P}_\rho)$. Each 0-feature borns at time 0 and dies when it merges with another connected component (of the open space), with the convention that the connected component C_i is killed by the connected component C_j when the leftmost point of C_i is lower than the one of C_j for the lexicographic order. For $k = 1$, when the growing radii create a new loop at a radius r, the quantity r is the birth time of the new loop; the death time of a loop is the smallest radius $r > 0$ when it is covered completely by $O_r(\mathcal{P}_\rho)$.

For technical reasons, we only investigate k-features, $k \in \{0, 1\}$, which are M-bounded (i.e. with diameters lower than some fixed value M) and with death times lower than some fixed deterministic radius $r_f > 0$. Then, for any $0 \leq r \leq s \leq r_f$, we denote by $\beta_{k,M}^{(r,s)}(\mathcal{P}_\rho)$ the M-bounded k-th persistent Betti number at times r, s. Similarly to Section 2.1.1, such a quantity counts the number of k-features which are M-bounded during all their lives and with birth (resp. death) time in $[0, r]$ (resp. in $(s, r_f]$). Since all 0-features born at time 0, only death times are relevant for $k = 0$. Hence, only the quantity $\beta_{0,M}^{(s)}(\mathcal{P}_\rho) = \beta_{0,M}^{(0,s)}(\mathcal{P}_\rho)$ is of interest.

Some assumptions on the point process $\mathcal{P}$ are required to state a functional central limit theorem for the persistent Betti numbers. First, we suppose that the factorial moment measures exist and are absolutely continuous. The p-th factorial moment density $\rho^{(p)}$ is determined via the following identity:

$$\mathbb{E}\left[\prod_{i\leq p} \#(\mathcal{P} \cap A_i)\right] = \int_{A_1\times\cdots\times A_p} \rho^{(p)}(x_1, \ldots, x_p)\mathrm{d}x_1 \ldots \mathrm{d}x_p$$

for any pairwise disjoint bounded Borel sets $A_1, \ldots, A_p \subset \mathbf{R}^2$. We require that $\mathcal{P}$ exhibits exponential decay of correlations. Roughly, this expresses an approximate factorization of the factorial moment densities (see Definition 3.1 in [9] for a precise statement). We also assume that a moment condition holds under the reduced Palm version and that the point process is conditionally m-dependent for some $m > 0$ and satisfies an absolutely continuity-type condition (see Section 3 in [9] for more details). Then, under these assumptions, we get the following functional central limit theorem:

Theorem 2.2.1. *The processes*

$$\left\{\rho^{-1/2}\left(\beta_{0,M}^{(s)}(\mathcal{P}_\rho) - \mathbb{E}\left[\beta_{0,M}^{(s)}(\mathcal{P}_\rho)\right]\right)\right\}_{s\leq r_f}$$

and

$$\left\{\rho^{-1/2}\left(\beta_{1,M}^{(r,s)}(\mathcal{P}_\rho) - \mathbb{E}\left[\beta_{1,M}^{(r,s)}(\mathcal{P}_\rho)\right]\right)\right\}_{r\leq s\leq r_f}$$

converge weakly in Skorokhod topology to centered Gaussian processes as ρ goes to infinity.

The expressions for the covariance structures of the limiting Gaussian processes can be made explicit. The proof of Theorem 2.2.1 is divided into several steps. As a first difficulty, we prove that the processes are tight. The tightness allows us to reduce our problem to finite dimensional distributions (see Lemma 3 in [53]) and then, combined with the Cramér-Wold theorem, to univariate central limit theorems (one for $\beta_{0,M}^{(\cdot)}(\mathcal{P}_\rho)$ and the other one for $\beta_{1,M}^{(\cdot,\cdot)}(\mathcal{P}_\rho)$). The key ingredient to prove the latter is a central limit theorem due to Błaszczyszyn *et al.* (Theorem 1.14 in [11]). One of the main difficulties consist in checking the assumptions of this theorem since it requires delicate estimates of the variances. Our theorem can be applied to various stationary point processes such as Log-Gaussian Cox processes and Matérn cluster processes.

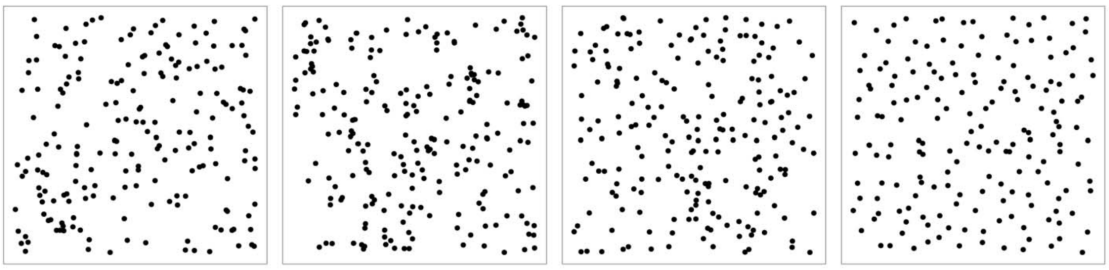


Figure 2.3: Samples from the Poisson null model, the Matérn cluster process, the Strauss process and the Baddeley-Silverman process (from left to right).

2.2.2 Simulation study

In [9], we also construct tests which are based on our functional central limit theorem for discriminating point processes. More precisely, given a stationary point process $\mathcal{P}$, we test if the latter is Poisson by considering lifetimes of connected components and loops. Under the null hypothesis, Theorem 2.2.1 ensures that the statistic

$$T_C = \int_0^{r_C} \mathrm{PD}_0(\mathcal{P}_\rho)([0,d])\mathrm{d}d$$

is asymptotically Gaussian, as ρ goes to infinity. In the above expression, we take $r_C \leq r_f$ and $\mathrm{PD}_0(\mathcal{P}_\rho)$ denotes the persistence diagram, i.e. $\mathrm{PD}_0(\mathcal{P}_\rho)([0,d]) = \beta_0^{(0)}(\mathcal{P}_\rho) - \beta_0^{(d)}(\mathcal{P}_\rho)$. Although the proof of Theorem 2.2.1 relies on the M-boundedness, we ignore this constraint in our simulations. In the same spirit, under the null hypothesis, the statistic

$$T_L = \int_0^{r_L} (d-b)\mathrm{PD}_1(\mathcal{P}_\rho)(\mathrm{d}b, \mathrm{d}d)$$

is also asymptotically Gaussian.

In our simulation study, the null model is a stationary Poisson point process of intensity 2 and the window is $W_{100} = [-5,5]^2$. As alternatives of the Poisson point process, we consider the Matérn cluster, the Strauss and the Baddeley-Silverman processes, each time with intensity 2. Figure 2.3 depicts realizations of these point processes. As observed in this figure, the Matérn cluster process is attractive, whereas the Strauss and the Baddeley-Silverman processes are more repulsive. Discriminating point process via TDA is natural since the study of lifetimes of connected components or loops allows us to detect if the point process has clusters or not.

For the alternatives introduced above, Figure 2.4 illustrates the persistence diagram. From the cluster-based diagrams, it becomes apparent that in comparison to the null model, in the Matérn cluster process, there is a pronounced peak of deaths at early times, whereas this happens very rarely in the Strauss process. When analyzing loops, we see that loops with long life times appear earlier in the null model than in the Matén cluster process. Conversely, while some loops with substantial life time emerge at later times in the null model, there are very few such cases in the Strauss model. Due to the complex higher order interaction of the Baddeley-Silverman process, its behavior is difficult to predict in advance. However, the samples in Figure 2.4 show that its topological characteristics are closer to those of a repulsive than a attractive point pattern.

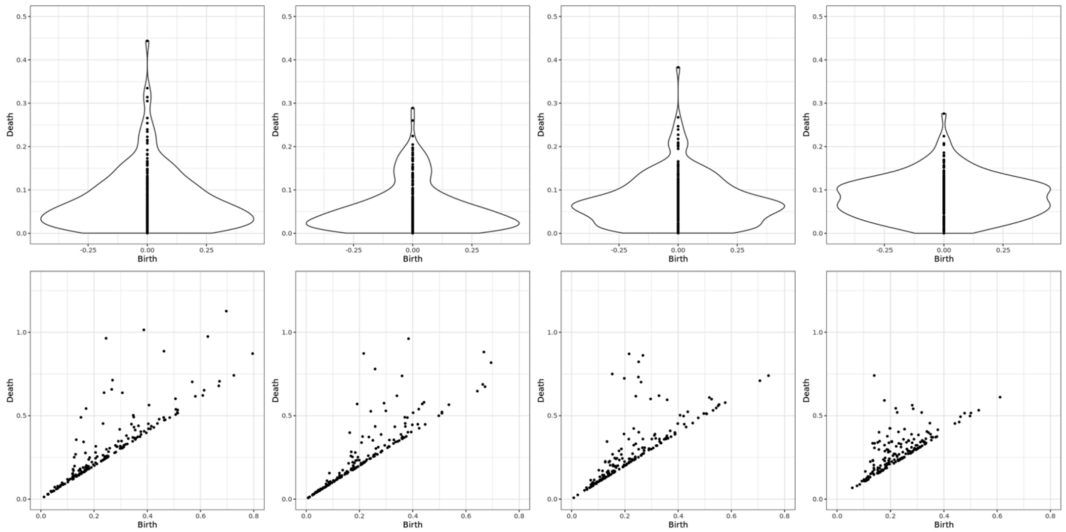


Figure 2.4: Persistence diagrams for cluster-based features with density plots (top) and loop-based features (bottom) for the Poisson null model, the Matérn cluster process, the Strauss process and the Baddeley-Silverman process (from left to right).

	Poisson	Matérn cluster	Strauss	Baddeley-Silverman
T_C	4.8%	55.7%	52.0%	65.6%
T_L	4.5%	63.0%	54.5%	84.7%

Table 2.1: Rejection rates for the test statistics T_C and T_L under the null model and the alternatives.

We estimate the means and the variances of T_C and T_L under the null hypothesis by computing the number of cluster deaths and accumulated loop life times for 10000 independent draws of the null model, for some suitable values of $r_C, r_L \leq r_f$. Then, to analyze the power of the test, we draw 1000 realizations from the null model and from the alternatives, respectively. Table 2.1 shows the rejection rates of this test setup. Under the null model, the rejection rates are close to the nominal 5%-level, thereby illustrating that already for moderately large point patterns the approximation by the Gaussian limit is accurate. Using the mean and variance from the null model, we compute the test powers for the alternatives. The statistic T_C leads to a test power of approximately 60% for both alternatives. When considering T_L, we obtain a type I error rate of 4.5%, so that the confidence level is kept. Moreover, the power analysis reveals that in the present simulation set-up, T_L is better in detecting deviations from the null hypothesis than T_C. As a concrete application, we use our tests for a point pattern in the context of neuroscience.

2.3 Extremal lifetimes of persistent cycles

In [30], we investigate extremes in TDA. We first introduce the problem in the context of Boolean models since it does not require notions of TDA. Then we extend our problem to persistent cycles in the context of Čech and Vietoris-Rips filtrations.

2.3.1 The Boolean model case

Before presenting our main results, we first give some notation. Let η be a stationary Poisson point process of intensity 1 in $\mathbf{R}^d$. For any $r \geq 0$, let $O_r(\eta) = \bigcup_{x\in\eta} B(x,r)$ be the so-called *Boolean model* associated with η and r. By a *cavity* living at time r, we mean a bounded connected component of the vacant space $V_r(\eta) = \mathbf{R}^d \setminus O_r(\eta)$. As the radius grows, cavities appear and disappear. When the growing radii create a new cavity at a radius r, we say that r is the birth time of the new cavity. Moreover, the death time of a cavity is the smallest radius $r > 0$ when it is covered completely by the Boolean model $O_r(\eta)$. We enumerate the cavities as $(J_i^*)_{i\geq 1}$ and associate with each such cavity its lifetime $L_i^* > 0$ and the point $Z_i^* \in \mathbf{R}^d$, referred to as the *nucleus*, as the last point that is covered at the death time (notice that two cavities living at different times are considered to be the same if their nuclei are equal).

Given a window $W_\rho = \rho^{1/d}[-\frac{1}{2},\frac{1}{2}]^d$, we would like to investigate the maximal lifetimes of cavities with nucleus in W_ρ, i.e. dying in W_ρ, as ρ goes to infinity. Such a problem is highly difficult since long-range interactions coming from percolation effects can occur. Indeed, if two cavities born at times which are close to the critical radius, it is possible that these cavities are very big and that their lifetimes depend on each others, even in the case where their nuclei are far. To avoid such a problem, we deal with a simpler problem by considering not all the cavities dying in W_ρ but only the one which die in W_ρ and with birth times outside a critical interval. More precisely, let r_c^O (resp. r_c^V) be the critical radius for occupied percolation (resp. for vacant percolation), i.e.

$$r_c^O = \inf\{r \geq 0 :\ \mathbb{P}\left(O_r(\eta) \text{ percolates }\right) > 0\}$$

and

$$r_c^V = \sup\{r \geq 0 :\ \mathbb{P}\left(V_r(\eta) \text{ percolates }\right) > 0\}.$$

Applying a result of percolation theory due to Duminil-Copin *et al.* [46], we can prove that, for any $\varepsilon > 0$, the events

$$E_\rho^V = \left\{ \sup_{r\notin[r_C^O-\varepsilon, r_c^V+\varepsilon]} R_\rho^V(r) \leq (\log\rho)^2 \right\}$$

and

$$E_\rho^O = \left\{ \sup_{r\notin[r_C^O-\varepsilon, r_c^V+\varepsilon]} R_\rho^O(r) \leq (\log\rho)^2 \right\}$$

occur with high probability, where $R_\rho^V(r)$ (resp. $R_\rho^O(r)$) denotes the maximal diameter of all bounded connected components in $V_r(\eta)$ (resp. $O_r(\eta)$) centered in W_ρ. In other words, with high probability, all the cavities which born outside the critical interval $[r_C^O - \varepsilon, r_c^V + \varepsilon]$ have, during all their lives, diameters lower than $(\log\rho)^2$. Such a property is important since it ensures that two distant cavities (in the sense that their nuclei are far) have lifetimes which tend to be independent when we restrict our attention only on cavities which born outside the critical interval.

Let $(L_i)_{i\geq 1}$ (resp. $(Z_i)_{i\geq 1}$) be the lifetimes (resp. nuclei) of all cavities with birth time outside $[r_C^O-\varepsilon, r_c^V+\varepsilon]$, where $\varepsilon > 0$ is fixed. We introduce concepts of EVT in the same spirit as Chapter 1. First, given $\tau > 0$, we consider a threshold $v_\rho = v_\rho(\tau)$ in such a way that the mean number of exceedances is τ, i.e.

$$\tau = \mathbb{E}\left[\#\{i \geq 1 :\ Z_i \in W_\rho \text{ and } L_i > v_\rho\}\right].$$

Then we consider the normalized point process of exceedances, i.e.

$$\Phi_{W_\rho} = \rho^{-1/d} \{Z_i : \ L_i > v_\rho\}_{Z_i \in W_\rho} .$$

The first main result in [30] is stated below.

Theorem 2.3.1. *For any $\tau > 0$,*

(i) $v_\rho \underset{\rho\to\infty}{\sim} (\kappa_d^{-1} \log \rho)^{1/d}$,

(ii) $\Phi_{W_\rho} \underset{\rho\to\infty}{\xrightarrow{\mathcal{D}}} \Phi$,

where κ_d is the volume of the unit ball and where Φ is a stationary Poisson point process of intensity τ in $[-\frac{1}{2}, \frac{1}{2}]^d$.

As a consequence of the above theorem we obtain $\kappa_d^{1/d}(\log \rho)^{-1/d} \max_{Z_i \in W_\rho} L_1 \underset{\rho\to\infty}{\xrightarrow{\mathbb{P}}} 1$. Such a convergence is not classical in EVT since, in general, we give a more precise result on the maximum, e.g. the limit distribution (see Chapter 1). However, we did not make explicit this distribution since it requires a more precise estimate for the scaling. To get *(i)*, the main ideas are the following. For the upper bound, we prove that v_ρ cannot be too big if not a too big portion of the window has no point of the Poisson point process. For the lower bound, we construct a template occuring with non-negligeable probability for which there exists at least one cavity with a large lifetime. To get *(ii)*, we adapt several arguments of Section 1.2 by showing that, with high probability, it is not possible to have a pair of exceedances in the same neighborhood.

2.3.2 Extremes for Vietoris-Rips and Čech complexes

As mentioned in Section 2.1.1, the nerve theorem claims that the Boolean model and the Čech complex are equal in terms of topological invariants. In particular, the study of maximal lifetimes for cavities in the Boolean model can be seen as a study on extremes of lifetimes for $(d-1)$-features in the context of a Čech complex. As an extension of Section 2.3.1, the paper [30] also deals with the minimal lifetimes for k-features, $1 \leq k \leq d-1$, centered in the window W_ρ, in the Čech and in the Vietoris-Rips filtrations, as ρ goes to infinity. As previously, the point set on which the simplicial complexes are based is a stationary Poisson point process in $\mathbf{R}^d$. Here again, to avoid percolation effects, we only consider features outside a critical interval. Similarly to Theorem 2.3.1, we obtain:

(i) estimates of the threshold chosen in such a way that the mean number of exceedances equals some $\tau > 0$;

(ii) Poisson approximation for the point process of exceedances.

The case $k = 0$ is not discussed since a 0-feature can be seen as a connected component. In particular, the longest lifetime corresponds to the longest edge in the minimum spanning tree whose asymptotic was established in Penrose [82], and the smallest lifetime corresponds to the minimum of interpoint distances. An interesting observation is that the orders which we obtain for the minimal lifetimes in the Čech and in the Vietoris-Rips filtrations are radically different. Indeed, for Čech, the order is ρ^{-2} whereas, for Vietoris-Rips, it is ρ^{-1}. This can be explained by the fact that cycles with 0 lifetime in the Vietoris-Rips filtration (and thus which are not counted) can have a very small positive lifetime in the Čech-filtration.

In a previous work, Bobrowski *et al.* have also investigated extremes of lifetimes but in the context of multiplicative persistence. More precisely, [12] defines the lifetime of a k-feature as the ratio between the death time and the birth time, whereas we work with the difference. The advantage of their paper is that, as opposed to ours, they do not have to consider critical intervals to avoid percolation effects. However, our results are more precise in the sense that we establish Poisson approximation with the order of the scaling whereas they provide estimates of the expectations up to multiplicative constants.

Chapter 3

The Internal Diffusion Limited Aggregation forest

Sommaire

3.1 Introduction

3.1.1 IDLA model and known results

The Internal Diffusion Limited Aggregation (IDLA) is a random growth model first introduced for chemical applications in 1986 by Meakin and Deutch and then, in a mathematical framework, by Diaconis and Fulton. In this model, the aggregate is recursively defined by adding to the aggregate the first site out of the current aggregate visited by a random walk starting from some source point. The standard IDLA model is constructed in $\mathbf{Z}^d$ as follows. We start with $A_0 = \emptyset$. At step N, a simple symmetric random walk starts from the origin 0 until it exits the current aggregate A_{N-1}, say at some vertex z, which is added to A_{N-1} to get $A_N = A_{N-1} \cup \{z\}$. In this manuscript, the word *particle* is used to refer to the random walk which is stopped when it exits the current aggregate A_{N-1}, and settled on the new vertex z.

A first shape theorem was established by Lawler *et al.* in [69] for the standard IDLA model. It asserts that the aggregate A_N (when it is suitably normalized) converges a.s. to an Euclidean ball as N goes to infinity, with fluctuations (w.r.t. the limit shape) which are at most linear.

Theorem 3.1.1. *(Lawler, Bramson, Griffeath) Let $\varepsilon > 0$ be fixed. Then, a.s., for $N = \lfloor \kappa_d n^d \rfloor$ and for n large enough,*

$$\mathbf{B}(0, n(1-\varepsilon)) \subset A_N \subset \mathbf{B}(0, n(1+\varepsilon)).$$

In the above theorem, $\mathbf{B}(0,r) = \{x \in \mathbf{Z}^d : \ |x| < r\}$ denotes the d-dimensional "lattice ball" of radius r and κ_d is the volume of the unit ball in $\mathbf{R}^d$. Roughly, the spherical shape of the aggregate can be explained by the facts that, asymptotically and under suitable scaling, a simple random walk looks like a Brownian motion and that a Brownian motion is isotropic.

Since then, many papers by Lawler (see e.g. [68]), Asselah and Gaudillière (see e.g. [3]), and Jerison, Levine and Sheffield (see e.g. [61]) have improved the bounds for fluctuations which are known to be logarithmic in $2D$ and sublogarithmic in higher dimensions. Recently, many variants of this problem have been considered. In particular, IDLA on discrete groups with polynomial or exponential growth have been studied in [10], with multiple sources in [72], on supercritical percolation clusters in [93], on cylinder graphs in [73], constructed with drifted random walks in [74] or with uniform starting points in [7].

One of the important ingredients of the IDLA model is the so-called *Abelian property*. The latter states that the distribution of any aggregate based on an IDLA protocol does not depend on the order in which the particles are sent. As a consequence, one can realize the cluster by sending many exploration waves. To illustrate this notion, let us consider N random walks starting from the origin and a real number $R > 0$. As a first wave, we send the particles associated with the random walks with the following constraint: if a particle reaches the (exterior) boundary of $\mathbf{B}(0,R)$, i.e.

$$\partial\mathbf{B}(0,R) = \{x \notin \mathbf{B}(0,R) : \ \exists y \in \mathbf{B}(0,R), |y-x| = 1\},$$

before settling, then we stop it on $\partial\mathbf{B}(0,R)$. The settled particles make up a cluster $A_R(N)$ (which is included in $\mathbf{B}(0,R)$). Let ζ_R be the positions of the stopped (but not settled) particles. As a second wave, we send particles with respect to the configuration ζ_R and with initial aggregate $A_R(N)$. Then the Abelian property implies that the random aggregate which is obtained is equal in distribution to $A(N)$. Such an observation is strongly used to derive shape theorems.

A random infinite tree $\mathcal{T}_\infty$ can be associated with the sequence of IDLA models (A_N) defined above in a very natural way. To our knowledge, this object has not been introduced in the literature. The tree $\mathcal{T}_1$ only consists of the root 0. By induction, $\mathcal{T}_N$ is obtained by adding to $\mathcal{T}_{N-1}$ the new vertex z such that $A_N = A_{N-1} \cup \{z\}$ and the edge used by the N-th particle to reach z from A_{N-1} (see Figure 3.1). Hence, we can define a.s. a random graph

$$\mathcal{T}_\infty = \bigcup_{N \geq 1} \uparrow \mathcal{T}_N,$$

which actually is a random tree (since each vertex of $\mathbf{Z}^2$ may only be added once) rooted at the origin. The lower bound for the shape theorem specifies that its edge set spans the whole set $\mathbf{Z}^2$. A natural question concerning this tree is the existence of (many) infinite branches with asymptotic directions. However, such a question is highly difficult since any branch of the IDLA tree $\mathcal{T}_\infty$ is not produced by a single particle but by many particles, each of them adding exactly one edge depending on the shape of the current aggregate. Another difficulty is the radial character of $\mathcal{T}_\infty$ (its branches are directed to the origin).

3.1.2 Main problem

In [25], we define three IDLA models in $\mathbf{Z}^2$ which are based on particles sent from each site of an infinite vertical axis. We investigate various properties of these aggregates, such as stationarity, mixing, stabilization and we establish shape theorems. The protocol defining one of our aggregates allows us to define a new (directed) random forest which is invariant w.r.t. vertical translations. The main aim of [25] is to prove the existence of such a forest.

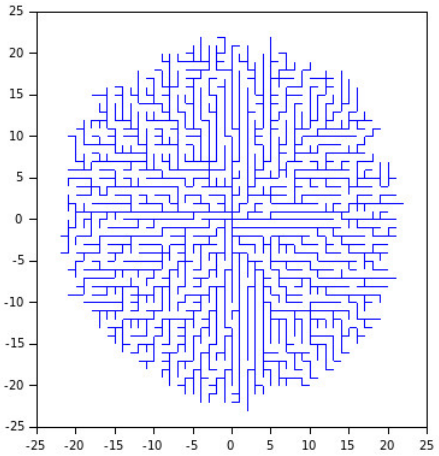


Figure 3.1: A realization of $\mathcal{T}_{1500}$.

3.2 Aggregates with an infinite number of sources

3.2.1 Construction

We define three aggregates in $\mathbf{Z}^2$ based on particles which are sent from each site of the vertical axis $\{0\} \times \mathbf{Z}$.

The first one is a natural extension of the standard IDLA. To construct it, let n be fixed. We first introduce a family of finite random aggregates $A_n[M]$, $M \geq 0$, with sources in the interval $\{0\} \times [\![-M, M]\!]$. When $M = 0$, the random set $A_n[0]$ is the standard IDLA cluster with volume n. Given a realization of $A_n[M-1]$, we send n particles from the site $(0, M)$ then n particles from the site $(0, -M)$. The set $A_n[M]$ denotes the aggregate which is produced by these $2n$ particles and by the aggregate $A_n[M-1]$ w.r.t. the IDLA protocol. As an illustration, Figure 3.2 gives a realization of $A_{90}[200]$ when it is observed in the strip $\mathbf{Z}_{20} = \{0\} \times [\![-20, 20]\!]$. By construction, the sequence of random aggregates $(A_n[M])_{M\geq 0}$ is increasing. We then define a first infinite aggregate as $A_n[\infty] = \bigcup_{M\geq 0} A_n[M]$. The aggregate is based on a specific order: the particles are first sent from the origin, then from levels ± 1, then from levels ± 2, and so on.

A second aggregate, say $A_n^*[\infty] = \bigcup_{M\geq 0} A_n^*[M]$, is constructed in the same spirit as above but this time the number of particles which are sent from each site of $\{0\} \times \mathbf{Z}$ is no longer equal to n but is a Poisson random variable with parameter n (the Poisson random variables are assumed to be independent). Here again, the protocol is based on the same order, namely 0, ± 1, ± 2, ...

A third aggregate is defined as follows. The number of particles from each site is still Poisson but this time the order for which they are sent is modified: the particles are not sent w.r.t. the specific order 0, ± 1, ± 2 but w.r.t. a family of random clocks. To do it, let $(\mathbf{N}_i)_{i\in\mathbf{Z}}$ be a family of independent and identically distributed Poisson point processes (PPP's) in $[0, n]$, with intensity 1. Each PPP $\mathbf{N}_i$ provides an increasing sequence $(\tau_{i,j})_{j\geq 1}$ of random clocks. Then, we attach to the collection $\{\tau_{i,j} : i \in \mathbf{Z}, j \geq 1\}$ a family of independent and identically distributed symmetric random walks $\{S_{i,j} : i \in \mathbf{Z}, j \geq 1\}$ which are also independent of the PPP's. In other words, at time $\tau_{i,j}$, the j-th particle from level i starts and its trajectory, associated with $S_{i,j}$, is instantaneously realized and adds a new site to the current aggregate. Hence, for any M, we can define an aggregate $A_n^\dagger[M]$ by sending particles from the source set $\{0\} \times [\![-M, M]\!]$ according to the clocks given by the corresponding PPP's up to time n. This construction ensures that, at

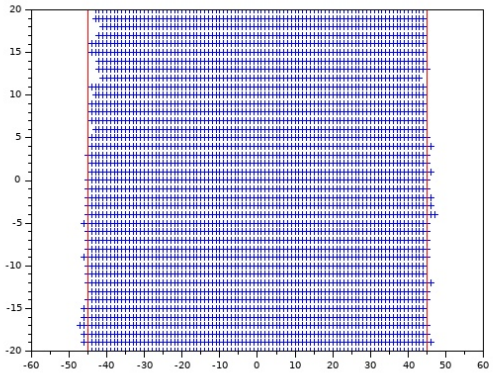


Figure 3.2: A realization of the aggregate $A_{90}[200] \cap \mathbf{Z}_{20}$ based on 90 particles per site $(0, i)$, with $|i| \leq 200$, and intersected by the strip $\mathbf{Z}_{20}$.

each time, the next particle (if it exists) is sent from a source chosen uniformly on $\{0\} \times [\![-M, M]\!]$. Similarly to the first two infinite aggregates, we define a third infinite aggregate $A_n^\dagger[\infty]$ as the increasing union of the $A_n^\dagger[M]$'s.

Motivation The reason for which we consider the three above aggregates is discussed below. Our main motivation is to construct a random forest in $\mathbf{Z}^2$ which is invariant w.r.t. vertical translations. The protocols defining the aggregates $A_n[\infty]$ and $A_n^*[\infty]$ can be used to define random forests. But these forests are not invariant w.r.t. vertical translations since they are based on the specific order 0, ± 1, ± 2.

The most natural approach is to use the protocol defining $A_n^\dagger[\infty]$ since, roughly, it consists in sending a first particle which is chosen uniformly at random, then a second one (also chosen uniformly at random on the same axis), and so on. More precisely, for each integer M, we can define a random forest $\mathcal{F}_n^\dagger[M]$ based on particles that are sent from each site of $\{0\} \times [\![-M, M]\!]$ w.r.t. the Poisson clocks, in the same spirit as we did for building the IDLA random tree (see Section 3.1.1). In particular, the set of vertices of $\mathcal{F}_n^\dagger[M]$ is $A_n^\dagger[M]$. Then, to define a random forest with an infinite number of sources, it is natural to take the limit of $\mathcal{F}_n^\dagger[M]$ over M. However, the existence of such an infinite random forest is not at all trivial since the sequence $(\mathcal{F}_n^\dagger[M])_{M \geq 0}$ *is not consistent*. Indeed, it is possible that a vertex in $\mathcal{F}_n^\dagger[M]$ is reached by a particle P_1 starting from $\{0\} \times [\![-M, M]\!]$ when we only send particles from this interval, but that the same vertex is reached by another particle, say P_2, when we send particles from $\{0\} \times [\![-M-1, M+1]\!]$; the last edge visited by P_2 being different from the one visited by P_1. Such a configuration can occur if a particle starting from $(0, M+1)$ (which works for $\mathcal{F}_n^\dagger[M+1]$ but not for $\mathcal{F}_n^\dagger[M]$) is sent before particles starting from the interval $\{0\} \times [\![-M, M]\!]$. In other words, if it is true that the sequence of aggregates $(A_n^\dagger[M])_{M \geq 0}$ (and thus the sets of vertices of the finite forests) is increasing, it is not true that $\mathcal{F}_n^\dagger[M]$ is included in $\mathcal{F}_n^\dagger[M+1]$. As an illustration, Figure 3.3 depicts a configuration in which two random forests have common vertices with different edges.

Working with the aggregate $A_n^\dagger[\infty]$ directly is delicate since there is no specific order. The main idea is first to work with $A_n[\infty]$ and $A_n^*[\infty]$ since they are based on simpler protocols. Then, as a key ingredient, we use the Abelian property to observe that the aggregates $A_n^*[\infty]$

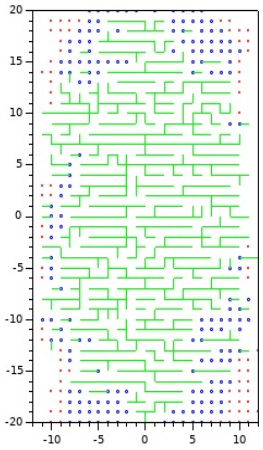


Figure 3.3: Realizations of the forests $\mathcal{F}^{\dagger}_{20}[20]$ and $\mathcal{F}^{\dagger}_{20}[50]$, defined on the same time interval $[0, 20]$, with different sets of sources, and restricted to the strip $\mathbf{Z}_{20}$, are depicted. The associated aggregates are coupled in the sense that they are based on the same clocks and random walks with level $|i| \leq 20$. In particular, $A^{\dagger}_{20}[20]$ is included in $A^{\dagger}_{20}[50]$. The edges created in both forests by the same particles are depicted in green. The red points are vertices of $A^{\dagger}_{20}[50] \backslash A^{\dagger}_{20}[20]$. The blue circles represent vertices in $A^{\dagger}_{20}[20]$ (and then also in $A^{\dagger}_{20}[50]$) which are reached by different particles in both aggregates and whose corresponding edges may differ in both forests $\mathcal{F}^{\dagger}_{20}[20]$ and $\mathcal{F}^{\dagger}_{20}[50]$. These blue vertices are possible discrepancies between forests $\mathcal{F}^{\dagger}_{20}[20]$ and $\mathcal{F}^{\dagger}_{20}[50]$.

and $A^{\dagger}_{n}[\infty]$ have the same distribution. In particular, all the results which hold for $A^{*}_{n}[\infty]$ also hold for $A^{\dagger}_{n}[\infty]$ and can be used to prove the existence of our random forest.

3.2.2 Main results

As a first result, we prove in [25] that the random aggregates $A_n[\infty]$ and $A^{*}_{n}[\infty]$ (and thus $A^{\dagger}_{n}[\infty]$) are *invariant w.r.t. vertical translations* in distribution. Such a result uses the concept of Choquet capacity (see e.g. p. 21 in [90]) and the Abelian property. A consequence of this fact is a *mass transport principle*. The latter states that, for any level $i \in \mathbf{Z}$, the expected number of sites in $A_n[\infty] \cap (\mathbf{Z} \times \{i\})$ equals the (expected) number of particles emitted from level i, i.e.

$$\mathbb{E}\left[\#A_n[\infty] \cap (\mathbf{Z} \times \{i\})\right] = n.$$

A similar result holds for the aggregate $A^{*}_{n}[\infty]$.

As a second important result, we prove that (for fixed n) the sequences $(A_n[M])_{M \geq 0}$ (resp. $(A^{*}_{n}[M])_{M \geq 0}$) satisfies a *strong stabilization* result. More precisely, given $\alpha > 1$, we show that a.s. there exists a random integer $M_0 = M_0(n) \geq 1$ such that, for any $M \geq M_0$, the trajectory of any particle contributing to $A_n[\infty]$ (resp. $A^{*}_{n}[\infty]$) and starting from $(0, i)$, with $|i| > M^{\alpha}$, does not visit the horizontal strip $\mathbf{Z}_M$. In other words, far particles do not touch central strips. The main idea to prove such a result is to proceed as follows. First, given a realization of $A_n[M^{\alpha}]$, where we assume that M^{α} is an integer for the sake of simplicity, we discretize the aggregate $A_n[M^{\alpha}]$ into annuli with suitable size (say, for instance, that the first one is at the top; the second one is below the first one, and so on). Then we send a particle far from the origin, say at level $M^{\alpha} + 1$. If the particle hits the strip $\mathbf{Z}_M$, it necessarily has to intersect all the annuli from the first one to one of those which intersect the strip. But, because we can bound the size

of $A_n[M^\alpha]$ (resp. $A_n^*[M^\alpha]$), we know that a non-negligeable proportion of such annuli are thin. Moreover, using a crossing's lemma due to Duminil-Copin *et al.* [45], we can prove that, with a non-negligeable probability, it is hard for the particle to cross a thin annulus. We deduce that it is unlikely that the particle hits a strip $\mathbf{Z}_M$ which is close to the origin since, if not, it has to cross many thin annuli, each time with a small probability.

On the reciprocal, we also prove that central particles do not touch far levels. The underlying stabilization results allow us to prove that the aggregate $A_n[\infty]$ has a *mixing property* w.r.t. vertical translations. In other words, for any events $\mathcal{A}$, $\mathcal{B}$, we have

$$\lim_{|k|\to\infty} \mathbb{P}\left(A_n[\infty] \in \mathcal{A}, A_n[\infty] \in \tau_k\mathcal{B}\right) = \mathbb{P}\left(A_n[\infty] \in \mathcal{A}\right)\mathbb{P}\left(A_n[\infty] \in \mathcal{B}\right),$$

where τ_k denotes the translation w.r.t. the vector $(0,k)$. The same property holds for the aggregates $A_n^*[\infty]$ and $A_n^\dagger[\infty]$. Furthermore, we can show that, with positive probability, the horizontal line $\mathbf{Z}\times\{0\}$ does not intersect the aggregate $A_n^*[\infty]$. Combined with the mixing property and thus the ergodicity, we deduce that the event $\{A_n^*[\infty]\cap(\mathbf{Z}\times\{i\})=\emptyset\}$ occurs infinitely often. In other words, we get the following result.

Theorem 3.2.1. *Let $n\geq 1$. With probability 1, for any integer M there are (infinitely many) levels $i\geq M$ and $j\leq -M$ such that the aggregate $A_n^*[\infty]$ does not intersect the axes $\mathbf{Z}\times\{i\}$ and $\mathbf{Z}\times\{j\}$.*

In particular, a.s. $A_n^*[\infty]$ (and thus $A_n^\dagger[\infty]$) has only finite connected components included in disjoint strips. The above theorem is one of the key ingredients to define our random forest.

Inspired by the method of Asselah and Gaudillière [3], we also establish shape theorems. More precisely, we prove the following result: there exists $A>0$ such that, for any $\alpha>0$, a.s. there exists $N\geq 1$ such that for any $n\geq N$,

$$R_{n/2-A\log(n)}\cap\mathbf{Z}_{n^\alpha}\subset A_n[\infty]\cap\mathbf{Z}_{n^\alpha}\subset R_{n/2+A\log(n)}\cap\mathbf{Z}_{n^\alpha},$$

where $R_r=[\![-r,r]\!]\times\mathbf{Z}$. Roughly, the above assertion claims that the aggregate $A_n[\infty]$, when it is restricted to a strip $\mathbf{Z}_{n^\alpha}$, is close to a rectangle with extremities $-\frac{n}{2}$ and $\frac{n}{2}$, up to logarithmic fluctuations (see Figure 3.2 for an illustration). Similarly, we obtain shape theorems for $A_n^*[\infty]$ and $A_n^\dagger[\infty]$.

3.3 Construction of the IDLA forest

Let $n,M\geq 0$. The protocol defining the aggregate $A_n^\dagger[M]$ allows us to define a random forest, say $\mathcal{F}_n^\dagger[M]$, in finite volume. The latter is based on particles which are sent from the interval $\{0\}\times[\![-M,M]\!]$ w.r.t. Poisson clocks, with in average n particles from each site of the interval. To construct a random forest spanning all $\mathbf{Z}^2$, which is invariant w.r.t. vertical translations, a natural approach is to proceed into two steps: first, we take the limit over M (vertical limit) and then we take the limit over n (horizontal limit). But, as mentioned in Section 3.2, taking the limit over M is an obstacle since the sequence $(A_n^\dagger[M])_{M\geq 1}$ is not consistent. Actually, a bad configuration can occur through a mechanism that we call a *chain of changes* which we describe below.

Assume that a particle, referred to as particle 1, starts at time $t_1\in(0,n)$ (from a level $M<|i_1|\leq M'$) and adds a site z_1 to $A_{t_1-}^\dagger[M']$. The aggregate at time t_1 becomes

$$A_{t_1}^\dagger[M'] = A_{t_1-}^\dagger[M']\cup\{z_1\}$$

while $A^\dagger_{t_1}[M]$ remains unchanged. In the above equation, the set $A^\dagger_{t_1-}[M']$ denotes the (current) aggregate produced just before sending particle 1. The site z_1 is a *discrepancy* at time t_1 between aggregates $A^\dagger_{t_1}[M]$ and $A^\dagger_{t_1}[M']$. If there is no other particles starting from a level $|i| \leq M$, at time $t \in (t_1, n)$ and going through z_1 then, at the final time n, the site z_1 constitutes a discrepancy (created by particle 1) between the aggregates $A^\dagger_n[M]$ and $A^\dagger_n[M']$. It also defines a discrepancy between the forests $\mathcal{F}^\dagger_n[M]$ and $\mathcal{F}^\dagger_n[M']$. Otherwise, we set

$$t_2 = \min\left\{t \in (t_1, n) : \text{a particle, starting from a level } |i| \leq M \text{ at time } t, \text{ goes through } z_1\right\}.$$

The particle starting from time t_2 is referred to as particle 2. This particle works for both aggregates. By definition, it adds the site z_1 to $A^\dagger_{t_2-}[M]$, so that the aggregate at time t_2 becomes $A^\dagger_{t_2}[M] = A^\dagger_{t_2-}[M] \cup \{z_1\}$. Thus, it continues its trajectory until adding a site z_2 (but only) to $A^\dagger_{t_2-}[M']$ which then becomes $A^\dagger_{t_2}[M'] = A^\dagger_{t_2-}[M'] \cup \{z_2\}$. At this time:

- the site z_1 now belongs to both aggregates but it could be reached via two different edges respectively in $A^\dagger_{t_2}[M]$ and $A^\dagger_{t_2}[M']$ so that the forests $\mathcal{F}^\dagger_n[M]$ and $\mathcal{F}^\dagger_n[M']$ may differ at the edge leading to z_1;
- the site z_2 is become a discrepancy between both aggregates at time t_2. This discrepancy is generated via a relay between particle 1 and particle 2.

Thus, we iterate this step while the current discrepancy is visited by a new particle starting from a level $|i| \leq M$. After a random number ℓ of steps (a.s. finite), we finally get the set of possible discrepancies between the forests $\mathcal{F}^\dagger_n[M]$ and $\mathcal{F}^\dagger_n[M']$, generated by particle 1. This set consists of edges leading to $z_1, \ldots, z_\ell$ and the final vertex z_ℓ itself. The mechanism producing this set of discrepancies is called a chain of changes, initiated by particle 1, between the forests $\mathcal{F}^\dagger_n[M]$ and $\mathcal{F}^\dagger_n[M']$. Notice that the aggregates $A^\dagger_n[M]$ and $A^\dagger_n[M']$ may have other chains of changes initiated by other particles starting from levels $M < |i| \leq M'$.

Roughly speaking, the existence of an infinite chain of changes involving an infinite number of relaying particles and initiated by a "Big Bang particle", *i.e.* a particle coming from a level arbitrarily far from the origin and born arbitrarily early, could modify infinitely often (in M) the forests $\mathcal{F}^\dagger_n[M]$, for $M \geq 0$, in the neighborhood of the origin. Proving that such infinite chain of changes does not exist with probability 1 leads to the next stabilization result and to the existence of the random forest $\mathcal{F}_n$.

Proposition 3.3.1. *Let $K \geq 1$. Then, a.s. there exists some (random) integer $M_0(K)$ such that, for any $M' > M \geq M_0(K)$, we have*

$$\mathcal{F}^\dagger_n[M] \cap \mathbf{Z}_K = \mathcal{F}^\dagger_n[M'] \cap \mathbf{Z}_K.$$

The above proposition is a consequence of Theorem 3.2.1. It allows us to define a.s. a random forest $\mathcal{F}^\dagger_n$, with set of sources $\{0\} \times \mathbf{Z}$, as the increasing union

$$\mathcal{F}^\dagger_n = \bigcup_{K \geq 0} \uparrow \mathcal{F}^\dagger_n[M_0(K)] \cap \mathbf{Z}_K.$$

The set of vertices of $\mathcal{F}^\dagger_n$ is $A^\dagger_n[\infty]$. It can be easily proved that the sequence of forests $(\mathcal{F}^\dagger_n)_{n \geq 0}$ is increasing. Such an observation allows us to define a new random forest $\mathcal{F}^\dagger$, referred to as the *directed IDLA forest*, as the increasing union of the $\mathcal{F}^\dagger_n$'s. Thanks to our shape theorems, the random forest $\mathcal{F}^\dagger$ spans $\mathbf{Z}^2$. Moreover, $\mathcal{F}^\dagger$ is invariant w.r.t. vertical translations and has a mixing property. The existence of such a model is the main topic of [25].

Chapter 4

Various works, works in progress and perspectives

Sommaire

4.1 Extremes of transient random walks in random sceneries

In [26], we extend a result due to Franke and Saigo [50] on extremes of random walks in random sceneries. We briefly recall the framework of [50]. Two quantities are considered:

- *A random walk* $(S_n)_{n\geq 0}$ in $\mathbf{Z}$, with $S_n = X_1 + \cdots + X_n$. The sequence X_k's are centered, integer-valued i.i.d. random variables and are in the domain of attraction of a stable law, i.e. for each $x \in \mathbf{R}$,

$$\mathbb{P}\left(n^{-\frac{1}{\alpha}} S_n \leq x\right) \underset{n\to\infty}{\longrightarrow} F_\alpha(x),$$

where F_α is the distribution function of a stable law with characteristic function given by

$$\varphi(\theta) = \exp(-|\theta|^\alpha (C_1 + iC_2 \mathrm{sgn}\theta)),\ \alpha \in (0,2].$$

When $\alpha < 1$ (resp. $\alpha > 1$), the random walk (S_n) is transient (resp. recurrent); see [70].

- *A random scenery* $(\xi(s))_{s\in\mathbf{Z}}$, where the $\xi(s)$'s are $\mathbf{R}$-valued i.i.d. random variables.

The sequences $(S_n)_{n\geq 0}$ and $(\xi(s))_{s\in\mathbf{Z}}$ are assumed to be independent and $(\xi(S_n))_{n\geq 0}$ is called a *random walk in random scenery*. Franke and Saigo derive limit theorems for the maximum of the first n terms of $(\xi(S_n))_{n\geq 0}$ as n goes to infinity. The results they obtain concern the recurrent and the transient cases. The first case is more delicate than the second one since long range dependence problems occur. An adaptation of Theorem 1 in [50] shows that, in the transient case, if u_n is a threshold chosen in such a way $n\mathbb{P}\left(\xi(0) > u_n\right) \underset{n\to\infty}{\longrightarrow} \tau$ for some $\tau > 0$, then

$$\mathbb{P}\left(\max_{k\leq n} \xi(S_k) \leq u_n\right) \underset{n\to\infty}{\longrightarrow} e^{-\tau q}, \tag{4.1.1}$$

where

$$q = \mathbb{P}\left(\forall k \in \mathbf{N}_+, S_k \neq 0\right). \tag{4.1.2}$$

The term q is (strictly) positive because the random walk $(S_n)_{n\geq 0}$ is transient. According to [70], the number q can be expressed as

$$q = \lim_{n\to\infty} \frac{R_n}{n} \quad \text{a.s.},$$

where $R_n = \#\{S_1, \ldots, S_n\}$ is the *range* of the random walk. In the sense of Equation (1.1.4), the quantity q is the extremal index and can be interpreted as the reciprocal of the mean size of a cluster of exceedances.

In [26], we consider the same problem as Franke and Saigo but this time we do not assume that the $\xi(s)$'s are i.i.d.. More precisely, we assume that the $\xi(s)$'s only satisfy a slight modification of the $D(u_n)$ and $D'(u_n)$ conditions (see Definitions 1.1.1 and 1.1.2). The result which is obtained only deals with the transient case and is similar to (4.1.1). Our proof is mainly based on an adaptation of [71]. We think that our method combined with Kallenberg's theorem ensures that the point process of exceedances converges to a Poisson point process, in the same spirit as Theorem 3 in [50]. More precisely, if the threshold is of the form $u_n = u_n(x) = a_n x + b_n$, for some $x \in \mathbf{R}$, and if we let $\tau_k = \inf\{m \in \mathbf{N}_+, \#\{S_1, \ldots, S_m\} \geq k\}$, then the point process

$$\Phi_n = \left\{ \left(\frac{\tau_k}{n}, \frac{\xi(S_{\tau_k}) - b_{\lfloor qn \rfloor}}{a_{\lfloor qn \rfloor}} \right), k \geq 1 \right\}$$

should converge to a Poisson point process with explicit intensity measure.

4.2 First digit phenomenon

4.2.1 Introduction

A sequence of positive numbers (x_n) is said to satisfy the *first digit phenomenon* in base $b \geq 2$ if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{F(x_n)=k} = \log_b\left(1 + \frac{1}{k}\right), \quad 1 \leq k < b,$$

where $F(x_n)$ is the first digit of x_n and where $\log_b$ denotes the logarithm in base b. Such a phenomenon was observed by Benford and Newcomb on real life numbers, e.g. electricity bills,

street addresses, stock prices and lengths of rivers. It is extensively used in various domains, such as fraud detection, computer design and image processing. As an extension of the first digit phenomenon, the notion of Benford sequence is introduced as follows. Let μ_b be the measure on the interval $[1, b)$ defined by

$$\mu_b([1, a)) = \log_b a$$

for any $a \in [1, b)$. Let $\mathcal{M}_b(x)$ be the *mantissa* in base b of a positive number x, i.e. $\mathcal{M}_b(x)$ is the unique number in $[1, b)$ such that there exists an integer k satisfying $x = \mathcal{M}_b(x)b^k$. A set of numbers (x_n) is referred to as a *Benford sequence* if for any $1 \leq a < b$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}_b(x_n)\in[1,a)} = \mu_b([1, a)).$$

In particular, each Benford sequence satisfies the first digit phenomenon since $F(x) = k$ if and only if $\mathcal{M}_b(x) \in [k, k + 1)$, with $x > 0$, $k \in [1, b)$. For instance, the sequences $(n!)$, (n^n) and (c^n) (with $\log_b c$ irrational) are Benford but the sequences (n) and $(\log n)$ are not. For various examples of sequences of positive numbers whose mantissae are (or approach to be) distributed with respect to μ_b, see e.g. [39].

It is straightforward that a sequence (x_n) of positive numbers is Benford in base b if and only if the sequence of fractional parts $(\{\log_b x_n\})$ is uniformly distributed in $[0, 1)$, i.e.

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[0,c)}(\{\log_b x_n\}) = c,$$

with $c \in [0, 1)$. Combining this with the Weyl's criterion (see p.7 in [66]), a sequence (x_n) is Benford if and only if, for any $h \in \mathbf{Z}^*$,

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2i\pi h \log_b x_n} = 0.$$

4.2.2 Products of random variables and the first digit phenomenon

In [31], we consider the following problem. Let (X_n) be a sequence of positive numbers and let $Y_n = \prod_{k=1}^{n} X_k$, $n \geq 1$. We discuss two concepts on the sequence of mantissae $(\mathcal{M}_b(Y_n))$:

(i) *the sequence* $(\mathcal{M}_b(Y_n))$ *is a.s. a Benford sequence*, i.e. for almost all ω and for all $1 \leq a < b$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}_b(Y_n(\omega))\in[1,a)} = \mu_b([1, a));$$

(ii) *the sequence* $(\mathcal{M}_b(Y_n))$ *converges to the Benford's law*, i.e. for all $1 \leq a < b$,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{M}_b(Y_n) \leq a\right) = \mu_b([1, a)).$$

Although the above concepts are connected, we provide counter-examples which show that the latter are different if we have no assumption on the sequence (X_n) (see Section 2.1 in [31]). However, when the X_n's are i.i.d., the property *(i)* (resp. *(ii)*) holds if and only if $\mathbb{E}\left[e^{2i\pi h \log_b X_1}\right] \neq 1$

(resp. $\left|\mathbb{E}\left[e^{2i\pi h \log_b X_1}\right]\right| \neq 1$) for every $h \in \mathbf{Z}^*$. In particular, in the i.i.d. case, if the sequence $(\mathcal{M}_b(Y_n))$ converges to the Benford's law then it is a.s. Benford.

Under the assumption that the sequence (X_n) is stationary, we prove that $(\mathcal{M}_b(Y_n))$ is a.s. a Benford sequence if and only if

$$\forall\, h \in \mathbf{Z}^*, \quad \lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}\left[e^{2i\pi h \log_b Y_n}\right] = 0. \tag{4.2.1}$$

The direct part of this result is a consequence of the Weyl's criterion whereas the reciprocal relies on applications of Van der Corput inequality (see e.g. p25 in [66]), Riesz's summation methods and Birkhoff's theorem. Since $(\mathcal{M}_b(Y_n))$ converges to the Benford's law if and only if $(\{\log_b Y_n\})$ converges to the uniform distribution, i.e. if and only if $\mathbb{E}\left[e^{2i\pi h \log_b Y_n}\right] \underset{n\to\infty}{\longrightarrow} 0$ for any $h \in \mathbf{Z}^*$, a consequence of (4.2.1) gives the following result:

Proposition 4.2.1. *Let (X_n) be a stationary sequence of positive random variables and let $Y_n = \prod_{k=1}^{n} X_k$. If $(\mathcal{M}_b(Y_n))$ converges to the Benford's law then $(\mathcal{M}_b(Y_n))$ is a.s. a Benford sequence.*

The above result extends our observation on the i.i.d. case. Other criterions ensuring that $(\mathcal{M}_b(Y_n))$ is a.s. a Benford sequence, or converges to the Benford's law, are also provided in [31]. These criterions, as well as (4.2.1), are illustrated through various examples including log-normal, exchangeable and 1-dependent random variables.

4.2.3 Discrepancy of powers of random variables

The main topic of [34] is to provide general conditions over a sequence of positive and independent random variables (X_n) to ensure that $(X_n^{d_n})$ is a.s. Benford in base 10 (in the sense of Section 4.2.2) for any (deterministic) sequence (d_n) converging to infinity at a rate at most polynomial. Such a question extends a work of Eliahou *et al.* [48] in which it is proved that several deterministic sequences at a power d tend to be Benford when d is large enough.

To define a deviation between the sequence $(X_n^{d_n})$ and the Benford's law, we introduce the following (random) quantity:

$$D_N((X_n^{d_n})) = \sup_{1\leq s<t<10} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[s,t)}(\mathcal{M}_{10}(X_n^{d_n})) - \mu_{10}([s,t)) \right|.$$

The above term is called the *discrepancy*. Our main result (Theorem 1 in [34]) is an upper bound for $D_N((X_n^{d_n}))$, depending on N and on some integrable random variable, which holds for any sequence (d_n) under suitable assumptions on the tails and on the characteristic functions of the X_n's. Such a result comes from the so-called Erdös-Turán inequality (see e.g. [85]). As a consequence, we prove that $D_N((X_n^{d_n}))$ converges a.s. to 0 as N goes to infinity, and therefore that $(X_n^{d_n})$ is a.s. Benford, when $d_n \underset{n\to\infty}{\longrightarrow} \infty$ and $d_n = O(n^\alpha)$ for some $\alpha > 0$. A second consequence is that $D_N((X_n^d))$ converges to 0 a.s. as $N, d \to \infty$. Roughly, it means that the sequence (X_n^d) is likely Benford as d goes to infinity.

Our main result is theoretically illustrated through various examples including geometric, (discrete and continuous) uniform, exponential and Fréchet distributions. Table 4.1 gives the frequencies of the first significant digits of $X_1^d, \ldots, X_N^d$ when X_n follows the continuous uniform distribution on $[1, n]$ for all $n \geq 1$, with $N = 1000$ and $d = 2$. As observed in this table, the latter are close to the frequencies of the first significant digit of the Benford's law.

First digit	(X_n^d)	Benford's law
1	0.293	0.306
2	0.183	0.184
3	0.130	0.116
4	0.099	0.106
5	0.081	0.082
6	0.065	0.055
7	0.058	0.050
8	0.047	0.053
9	0.043	0.048

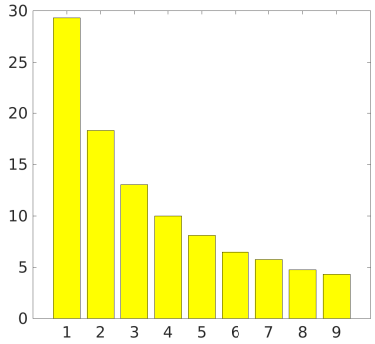


Table 4.1: A simulation of the frequencies of the first significant digits of $X_1^d, \dots, X_N^d$, where X_n has a uniform distribution on $[1, n]$ for each $n \geq 1$, with $N = 1000$ and $d = 2$.

4.3 Recent works and works in progress

In this section, we present three works. The first one has been recently submitted (September 2022) and is the most significant. The two others are works in progress which should be submitted in 2022 or in 2023.

4.3.1 Composite likelihood estimators for Brown-Resnick random fields in a fixed domain

In a common work with C. Y. Robert, we estimate parameters of a widely used class of stationary max-stable random fields: the Brown-Resnick random field. The latter can be defined w.r.t. its spectral representation as follows (see Section 1.1.2). Let (U_i) be the decreasing enumeration of the points of a Poisson point process on $(0, \infty)$ with intensity $u^{-2}\mathrm{d}u$. Let (W_i) be a family of i.i.d. fractional Brownian fields in $\mathbf{R}^2$ with scale parameter $\sigma > 0$ and range parameter $\alpha \in (0, 2)$ (the term $H = \alpha/2$ is also known as the *Hurst parameter*), i.e. W_i is a centered Gaussian field with $W(0) = 0$ and with variance

$$\mathbb{V}\left[W_i(x)\right] = \sigma^2 |x|^\alpha =: 2\gamma(x), \tag{4.3.1}$$

for any $x \in \mathbf{R}^2$. The (isotropic) *Brown-Resnick random field* is defined as

$$\eta(x) = \bigvee_{i \geq 1} U_i \exp\left(W_i(x) - \gamma(x)\right),$$

where $\bigvee$ denotes the pointwise maximum. While the W_i's have stationary increments (i.e. the distribution of $(W_i(x + x_0) - W_i(x_0))_{x \in \mathbf{R}^2}$ does not depend on $x_0 \in \mathbf{R}^2$), it can be shown that the Brown-Resnick process is stationary (i.e. the distribution of $(\eta(x + x_0))_{x \in \mathbf{R}^2}$ is the same as η). The random field η is simple in the sense that it has standard unit Fréchet marginals, i.e. $\mathbb{P}\left(\eta(x) \leq z\right) = \exp(-z^{-1})$ for any $z > 0$.

The goal of our work is to infer the parameters σ and α in *infill* asymptotic. More precisely, we consider more and more datas which are observed in some *fixed bounded sampling domain*. Making inference in infill is much more difficult than if we do it in an increasing window (tending to infinity) since we do not have mixing properties. From a theoretical point of view, the maximum likelihood method is the best approach. Nevertheless, the evaluation of the likelihood

function is computationally impractical for large datasets. Indeed, although there exists a theoretical formula for the joint distributions (see Equation (1.1.7)), the latter is not useable from a practical point of view. To overcome this difficulty, we apply *composite likelihood* methods. These methods use objective functions which are based on the likelihood of lower dimensional marginals or conditional events and, in general, provide a good balance between computational complexity and statistical efficiency, see e.g. [97].

We consider a Poisson stochastic spatial sampling scheme and use the Poisson-Delaunay graph to select the pairs and triples of sites with their associated marginal distributions that will be integrated into the composite likelihood (CL) objective functions, i.e. we exclude pairs that are not edges of the Delaunay graph or triples that are not vertices of triangles in the graph. More precisely, we consider a stationary Poisson point process $\mathcal{P}_N$ with intensity N in $\mathbf{R}^2$ and a fixed window, say $[0,1]^2$. The pairwise (log) CL function is defined as

$$\ell_{2,N}(\sigma,\alpha) = \sum_{(x_1,x_2)\in E_N} \log f_{x_1,x_2}(\eta(x_1),\eta(x_2))$$

while the triplewise (log) CL function is defined as

$$\ell_{3,N}(\sigma,\alpha) = \sum_{(x_1,x_2,x_3)\in DT_N} \log f_{x_1,x_2,x_3}(\eta(x_1),\eta(x_2),\eta(x_3)).$$

In the above expressions, f_{x_1,x_2} (resp. f_{x_1,x_2,x_3}) denotes the density of $(\eta(x_1),\eta(x_2))$ (resp. $(\eta(x_1),\eta(x_2),\eta(x_3))$) and E_N (resp. DT_N) denotes the set of couples of points (x_1,x_2) (resp. of triples of points (x_1,x_2,x_3)) such that

- x_1 and x_2 are Delaunay neighbors (resp. x_1,x_2,x_3 are the vertices of a Delaunay triangle) in the Delaunay graph associated with $\mathcal{P}_N$;
- x_1 is the leftmost point, with $x_1 \in [0,1]^2$.

To the best of our knowledge, this is the first time that a Delaunay graph is used to select the pairs and triples. Using this graph is natural since we only use the distributions of pairs and triples. Moreover, the Delaunay graph appears to be the most regular graph in the sense that it is the one which maximises the minimum of the angles of the triangles. When α is assumed to be known, the pairwise and triplewise maximum CL estimators of σ are respectively defined as

$$\hat{\sigma}_{j,N} = \mathrm{argmax}_{\sigma>0}\, \ell_{j,N}(\sigma,\alpha), \quad j=2,3$$

whereas, when σ is assumed to be known, the pairwise and triplewise maximum CL estimators of α are respectively defined as

$$\hat{\alpha}_{j,N} = \mathrm{argmax}_{\alpha\in(0,2)}\, \ell_{j,N}(\sigma,\alpha), \quad j=2,3.$$

To avoid heavy notation, we only present our results on the pairwise estimation of σ. Using a known expression of f_{x_1,x_2} (see e.g. [59]), we prove that

$$\lim_{d\to 0} \frac{\partial}{\partial\sigma} \log f_{x_1,x_2}(z_1,z_2) = \frac{1}{\sigma}(u^2-1),$$

where u is fixed, z_1, z_2 are such that $u = d^{-\alpha/2}\sigma^{-1}\log(z_2/z_1)$ and $d = |x_2-x_1| > 0$. Since the typical distance between two Delaunay neighbors tends to 0 as N goes to infinity, the above expression roughly ensures that the estimator $\hat{\sigma}_{2,N}$ satisfies the following property:

$$\frac{1}{\hat{\sigma}_{2,N}} \sum_{(x_1,x_2)\in E_N} \left(\frac{\sigma^2}{\hat{\sigma}_{2,N}} U^2_{x_1,x_2} - 1 \right) \simeq 0,$$

where

$$U_{x_1,x_2} = |x_2 - x_1|^{-\alpha/2}\sigma^{-1}\log\left(\eta(x_2)/\eta(x_1)\right).$$

According to [42], we can prove that U_{x_1,x_2} is asymptotically a standard Gaussian random variable as $|x_2-x_1|$ converges to 0, where x_1 is fixed. Our main theorem deals with the asymptotic distribution of squared increment sums for the max-stable Brown-Resnick random field and can be stated as follows.

Theorem 4.3.1. *Let $\alpha \in (0,1)$ and let*

$$V_{2,N} = \frac{1}{\sqrt{|E_N|}} \sum_{(x_1,x_2)\in E_N} \left(U_{x_1,x_2}^2 - 1\right).$$

Then there exists a constant c_2 and there exist a family of positive random variables $L_{Z_{k\setminus j}}(0)$, $k, j \geq 1$, such that

$$N^{-(2-\alpha)/4} V_{2,N} \xrightarrow[n\to\infty]{\mathbb{P}} c_2 \sum_{j\geq 1}\sum_{k>j} L_{Z_{k\setminus j}}(0).$$

The constant c_2 can be made explicit and is (strictly) negative. To define the random variables $L_{Z_{k\setminus j}}(0)$, we proceed as follows. First, following an idea of Dombry and Kabluchko [44], we define for each k, j a random cell

$$C_{k,j} = \left\{ x \in [0,1]^2 :\ Z_k(x) \wedge Z_j(x) > \bigvee_{i\neq j,k} Z_i(x) \right\},$$

where $Z_i(x) = \log U_i + (W_i(x) - \gamma(x))$, $x \in \mathbf{R}^2$. In other words, the random cell $C_{k,j}$ is the set of points for which the two largest trajectories in η are the k-th and the j-th. The family of cells defines a random tessellation of $[0,1]^2$, not in the sense of Stochastic Geometry (the cells are not necessarily convex and connected), but in the sense that each point belongs a.s. to a unique cell. Besides, there exists a.s. a finite number of cells which are non-empty. Given a realization of η, and thus a realization of the tessellation, the random variable $L_{Z_{k\setminus j}}(0)$ is the *local time* at level 0 of the random field $Z_{k\setminus j} = Z_k - Z_j$. This random variable is defined as the density, evaluated in 0, w.r.t. the Lebesgue measure of the random measure

$$\nu^{(k\setminus j)}(A) = \int_{C_{k,j}} \mathbb{1}_{Z_{k\setminus j}(x)\in A} \mathrm{d}x,$$

with $A \subset \mathbf{R}$, provided that $C_{k,j}$ has a positive Lebesgue measure (if not, we take $L_{Z_{k\setminus j}}(0) = 0$). In other words, $L_{Z_{k\setminus j}}(0) = \frac{\mathrm{d}\nu^{(k\setminus j)}}{\mathrm{d}\ell}(0)$. Roughly, the local time $L_{Z_{k\setminus j}}(0)$ measures the set of points x such that $Z_{k\setminus j}(x) = 0$.

Theorem 4.3.1 is not at all classical in the sense that the limit distribution is not a Gaussian random variable but a (finite) sum of local times. Local times already appeared in the context of max-stable random fields. Indeed, Robert [87] recently establishes a biased central limit theorem, whose bias depends on local times, in the context of power variations for a class of Brown-Resnick processes defined on $\mathbf{R}$. The proof of Theorem 4.3.1 is divided into three steps. First, applying a result on normal approximations due to Nourdin and Peccati (Theorem 6.3.1 in [78]), we establish a central limit theorem for the sum of squared increments when we consider only one trajectory, i.e. only one fractional Brownian random field. More precisely, we obtain the

asymptotic behaviour of the random variable defined in the same spirit as $V_{2,N}$ when we replace the term $\log(\eta(x_2)/\eta(x_1))$ by $W_1(x_2) - W_1(x_1)$ in the expression of U_{x_1,x_2}. Then, we extend our result to the case where we consider the maximum of two trajectories. Here, the random variable which appears in the limit is a local time. Finally we get Theorem 4.3.1 by considering the random tessellation described above.

As a consequence of Theorem 4.3.1, we can derive asymptotic properties of the maximum CL estimators. More precisely, we get

$$\sqrt{|E_N|}N^{-(2-\alpha)/4}(\hat{\sigma}^2_{2,N} - \sigma^2) \xrightarrow[n\to\infty]{\mathbb{P}} c_2\sigma^2 \sum_{j\geq 1}\sum_{k>j} L_{Z_{k\setminus j}}(0).$$

In a similar way, we obtain the asymptotic behaviours of $\hat{\sigma}^2_{3,N}$ (i.e. the triplewise CL estimator of σ based on Delaunay triangles) and of $\hat{\alpha}^2_{2,N}$ and $\hat{\alpha}^2_{3,N}$ (i.e. the pairwise/triplewise CL estimators of α).

4.3.2 Compound Poisson process approximation with explicit rate of convergence

In a common work with M. Otto, we establish compound Poisson approximations with examples in Stochastic Geometry.

The framework is the following. Let η be a Poisson point process in some locally compact separable metric space $(\mathbf{X}, d)$ with finite intensity measure and let $g : \mathbf{X} \times \mathbf{N_X} \to \{0, 1\}$ be a measurable function, where $\mathbf{N_X}$ denotes the space of all σ-finite counting measures on $\mathbf{X}$. Assume that g is localized in the sense that, for any $\omega \in \mathbf{N_X}$ and $x \in \mathbf{X}$, there exists a set $S(x) \subset \mathbf{X}$ such that, for any $S \subset S(x)$, we have

$$g(x, \omega) = g(x, \omega \cap S).$$

We are interested by the following discrete random measure:

$$\xi[\eta] = \sum_{x\in\eta} g(x, \eta)\delta_x.$$

The function g has to be seen as the indicator function that an extreme event occurs. For instance, if $g(x, \omega) = \mathbb{1}_{R_k(x,\omega)>v}$, where R_k denotes the distance between x and its k-th nearest neighbor in ω (provided that such a quantity exists and is unique) and where v is a suitable threshold, then $\xi[\eta]$ counts the number of points with distances to the k-th nearest neighbors larger than v.

Our main result is an upper bound for the Kantorovich distance between $\xi[\eta]$ and a compound Poisson point process in $\mathbf{X}$ (see e.g. [5] for a definition of the Kantorovich distance). The proof is mainly based on a paper of Barbour and Månsson [5] and strongly uses the Chen-Stein method. As an application, we obtain compound Poisson approximations for the maximum (resp. minimum) spheres in the k-nearest neighbor graph. Our work complements [30] in which only a Poisson approximation is established. We try to apply our result to more examples and to adapt it in the context of binomial point process.

4.3.3 Properties of extremes for simple random walks in random sceneries

In a common work with A. Darwiche and A. Rousselle, we give a more specific treatment of extremes of random walks in random sceneries (see Section 4.1). Similarly to [26], we consider

a random walk $(S_n)_{n\geq 0}$ in $\mathbf{Z}$ and a random scenery $(\xi(s))_{s\in\mathbf{Z}}$ which consist of real random variables. This time, the sequence $(S_n)_{n\geq 0}$ is assumed to be the *simple* random walk, i.e. $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. and $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = -1) = 1 - p$. Here again we assume that $(S_n)_{n\geq 0}$ is transient, i.e. $p \neq 1/2$. As opposed to [26], we do not assume that $(\xi(s))_{s\in\mathbf{Z}}$ satisfies the $D'(u_n)$ condition (thus, clusters of exceedances for the $\xi(s)$'s can occur). However, throughout our work, we require that $(\xi(s))_{s\in\mathbf{Z}}$ satisfies a slight stronger assumption than the $D(u_n)$ condition, referred to as the $\Delta(u_n)$ condition. Roughly, the latter is a condition which ensures a mixing-type behaviour for the tails of the joint distributions of a stationary sequence of random variables (see [56] for a precise definition). Similarly to Section 4.1 (see also Chapter 1), the threshold u_n is chosen in such a way that $n\mathbb{P}(\xi(0) > u_n) \underset{n\to\infty}{\longrightarrow} \tau$, for $\tau > 0$. In addition, we assume that the following property holds:

$$\max_{A\in\tilde{\mathcal{B}}_1^n(u_n), B\in\tilde{\mathcal{B}}_1^n(u_n)} \{|\mathbb{E}[\mathbb{P}(A|\mathcal{S}_{1:n})\,\mathbb{P}(B|\mathcal{S}_{1:n})] - \mathbb{P}(A)\,\mathbb{P}(B)|\} \underset{n\to\infty}{\longrightarrow} 0, \tag{4.3.2}$$

where $\tilde{\mathcal{B}}_1^n(u_n)$ stands for the σ-algebra generated by events of the form $\{\xi(S_i) \leq u_n\}$, $1 \leq i \leq n$ and $\mathcal{S}_{1:n} = \{S_1, \ldots, S_n\}$. Our main results can be stated as follows.

(i) The sequence $(\xi(S_n))_{n\geq 0}$ satisfies the $\Delta(u_n)$ condition.

(ii) If the extremal index of $(\xi(s))_{s\in\mathbf{Z}}$ exists and equals σ, then the extremal index of $(\xi(S_n))_{n\geq 0}$ also exists and equals $\theta = \sigma q$, where q is as in (4.1.2).

(iii) Under suitable assumptions, the point process of exceedances $\Phi_n = \left\{\frac{i}{n} : \xi(S_i) > u_n, i \leq n\right\}$ converges to a compound Poisson point process.

Although the condition given in (4.3.2) is restrictive, we think that all our results remain true when we do not assume such a condition. Assertion (i) is natural and can be understood as follows: because $(\xi(s))_{s\in\mathbf{Z}}$ has a mixing property for the tails and because $(S_n)_{n\geq 0}$ has a drift, then also $(\xi(S_n))_{n\geq 0}$ has a mixing property for the tails. Assertion (ii) is, in some sense, a generalization of [26]. Indeed, in this paper we prove that the extremal index of $(\xi(S_n))_{n\geq 0}$ exists and equals $\theta = q$ provided that the sequence $(\xi(s))_{s\in\mathbf{Z}}$ satisfies the $D'(u_n)$ condition. Since this condition ensures that the extremal index of $(\xi(s))_{s\in\mathbf{Z}}$ exists and equals $\sigma = 1$ (see Section 1.1.2), we obtain as a special case of (ii) that $\theta = \sigma q$. We have in mind potential examples for which we could make explicit the cluster size distribution of the compound Poisson point process appearing in Assertion (iii). However, this requires to check condition (4.3.2), which is one of the matters that we currently investigate.

4.4 Perspectives

We give below several lines of research. The latter are classified according to the order of our chapters.

Extremes in Stochastic Geometry A first perspective is to improve our results on extremes in Stochastic Geometry. As described in Section 1.2.2, we obtained Poisson approximations for the number of exceedances in the context of STIT and Poisson line tessellations. However, our results only concern the inradius and only hold in $\mathbf{R}^2$. A natural question is to consider more general geometric characteristics and to extend the Poisson approximations in higher dimension. Another extension could concern the maximal degree for random graphs. As stated in Theorem 1.4.1, such a quantity is concentrated, with high probability, on two consecutive integers in the

context of planar Poisson-Delaunay graph. Although we prove in [14] that the maximal degree is also concentrated on a finite (and deterministic) number of consecutive integers in higher dimension, we do not show that such a number equals two. It could be interesting to do it and to consider more general geometric random graphs, e.g. Gabriel graphs. Another question concerns our work on the stretch factor in a planar Delaunay graph (Section 1.4.2). As stated in Theorem 1.4.2, as the intensity of the underlying Poisson point process goes to infinity and for fixed points $s, t \in \mathbf{R}^2$, the latter is larger than $(1+\varepsilon)|s-t|$ for some explicit value of $\varepsilon > 0$. However, the value which we obtain for ε is far from optimal. A natural question is to improve it and to extend our result in $\mathbf{R}^d$, $d \geq 3$. In the light of [33] (see also Section 1.3), it would be also very interesting to make explicit the cluster size distribution for (non-trivial) examples which are discussed in this paper, namely the maximum of circumradius for Poisson-Voronoi and Poisson-Delaunay tessellations. Another perspective is to give more examples of computations of extremes for random tessellations; for instance the minimum of angles in a planar Poisson-Delaunay tessellation and the minimum of areas in a planar Poisson-Voronoi tessellation (including the study of the shape of the cell minimizing the underlying geometric characteristic), or more generally geometric characteristics discussed in this manuscript (Chapter 1) but for random tessellations not necessarily based on a Poisson point process.

Topological Data Analysis A second perspective is to strengthen our research in TDA. With T. Owada, we envisage to work together and to deepen a result that he recently obtained with Z. Wei. The problem which is considered is the following. Given n i.i.d. points $X_1, \ldots, X_n$ in $\mathbf{R}^d$ distributed w.r.t. some isotropic density f, and given an increasing sequence (R_n) of positive numbers tending to infinity, we construct the Čech complex with set of vertices $\{X_1, \ldots, X_n\} \cap B(0, R_n)^c$, i.e. outside of an expanding ball. In [80], T. Owada and Z. Wei establish various functional strong law of large numbers for Betti numbers of the simplicial complex. Their results depend on the decay rate of the density f and on how rapidly R_n diverges. With T. Owada, we would like to deepen the latter by establishing large deviation principles. One of the main difficulties is that the point process $\{X_1, \ldots, X_n\}$ is not stationary.

IDLA forest and IDLA tree A third perspective is to deepen our work on the IDLA forest (see Section 3.3) with D. Coupier, A. Rousselle and our future PhD student (K. Penner). Many questions arise from [25]. A first one is to extend the construction of our random forest in higher dimension. Such an extension is not trivial since one of the key results which allows us to define our model in $\mathbf{R}^2$ (see Theorem 3.2.1) does not hold in $\mathbf{R}^d$, $d \geq 3$. A natural approach is to deal with the chains of discrepancies as described in Section 3.3. A second type of questions concerns properties of the random forest. For instance, is it true that a.s. all the trees are finite? In a suitable direction, namely the x-axis, can we say (in a sense which has to be specified) that the random forest coincides asymptotically with the IDLA tree? If so, can we deduce properties of the IDLA tree from the one of the IDLA forest? Quite recently, it was proved by Jerison *et al.* [62] that, under a suitable normalization, the fluctuations (taken over time and space) of the classical IDLA model scale to a variant of the Gaussian free field. Do we obtain the same type of results for the process which generates our random forest? Other questions concern properties of the IDLA tree. Are there many infinite branches? If so, can we say that they are tight and that they have asymptotic directions? These questions are delicate because the random tree has a radial property and is based on particles which depend on each others.

Inference for Brown-Resnick on a grid in a fixed window A fourth perspective is to extend the problem considered in Section 4.3.1 with C. Y. Robert. More precisely, given a Brown-Resnick random field with parameters α, σ (see Equation (4.3.1)), we observe the latter

in a fixed window, say $[0,1]^2$. Instead of fixing a parameter and estimating the other one as we do in Section 4.3.1, we want to estimate the parameters α and σ simultaneously. Taking a Delaunay graph for the sampling scheme seems very technical since it requires to take account the distances between pairs of nodes. To simplify the problem, we can use instead a regular grid. More precisely, we assume that the Brown-Resnick random field is observed on the set of points $\{(i/n, j/n) : 0 \leq i, j \leq n\}$. We use the grid for selecting pairs of points and for defining a pairwise composite likelihood estimator of (α, σ). Here, the pairs which are selected are not at distance 1 in terms of distance of graph (as we do in Section 4.3.1) but at distance 2 for identifying the parameters. Our aim is to show that our estimator is consistent and to deal with its asymptotic behaviour as n goes to infinity.

Bibliography

[1] C. W. Anderson. Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probability*, 7:99–113, 1970.

[2] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

[3] A. Asselah and A. Gaudillière. Sublogarithmic fluctuations for internal DLA. *Ann. Probab.*, 41(3A):1160–1179, 2013.

[4] F. Avram and D. Bertsimas. On central limit theorem in geometrical probability. *The Annals of Applied Probability*, 3(4):1033–1046, 1993.

[5] A. D. Barbour and M. Mansson. Compound Poisson process approximation. *Ann. Probab.*, 30(3):1492–1537, 2002.

[6] B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Process. Appl.*, 119(4):1055–1080, 2009.

[7] I. Benjamini, H. Duminil-Copin, G. Kozma, and C. Lucas. Internal diffusion-limited aggregation with uniform starting points. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(1):391–404, 2020.

[8] M. W. Bern, D. Eppstein, and F. F. Yao. The expected extremes in a Delaunay triangulation. *ICALP*, pages 674–685, 1991.

[9] C. A. N. Biscio, N. Chenavier, C. Hirsch, and A. M. Svane. Testing goodness of fit for point processes via topological data analysis. *Electron. J. Stat.*, 14(1):1024–1074, 2020.

[10] S. Blachère and S. Brofferio. Internal diffusion limited aggregation on discrete groups having exponential growth. *Probab. Theory Related Fields*, 137(3-4):323–343, 2007.

[11] B. Blaszczyszyn, D. Yogeshwaran, and J. E. Yukich. Limit theory for geometric statistics of point processes having fast decay of correlations. *Ann. Probab.*, 47(2):835–895, 2019.

[12] O. Bobrowski, M. Kahle, and P. Skraba. Maximally persistent cycles in random geometric complexes. *Ann. Appl. Probab.*, 27(4):2032–2060, 2017.

[13] G. Bonnet, P. Calka, and M. Reitzner. Cells with many facets in a Poisson hyperplane tessellation. *Advances in Mathematics*, 324:203–240, 2018.

[14] G. Bonnet and N. Chenavier. The maximal degree in a Poisson-Delaunay graph. *Bernoulli*, 26(2):948–979, 2020.

[15] P. Bose and P. Morin. Online routing in triangulations. *SIAM J. Comput.*, 33(4):937–951, 2004.

[16] N. Broutin, O. Devillers, and R. Hemsley. The maximum degree of a random Delaunay triangulation in a smooth convex. *AofA - 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, 2014.

[17] P. Calka. The distributions of the smallest disks containing the Poisson-Voronoi typical cell and the Crofton cell in the plane. *Adv. in Appl. Probab.*, 34(4):702–717, 2002.

[18] P. Calka. *Ecole d'été-Autriche, ch.7.* Lecture Notes, 2010.

[19] P. Calka, A. Chapron, and N. Enriquez. Poisson-Voronoi tessellation on a Riemannian manifold. *Int. Math. Res. Not. IMRN*, (7):5413–5459, 2021.

[20] P. Calka and N. Chenavier. Extreme values for characteristic radii of a Poisson-Voronoi tessellation. *Extremes*, 17(3):359–385, 2014.

[21] P. Calka, Y. Demichel, and N. Enriquez. Large planar Poisson-Voronoi cells containing a given convex body. *Ann. H. Lebesgue*, 4:711–757, 2021.

[22] F. Chazal and B. Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence 4*, 2021.

[23] N. Chenavier. A general study of extremes of stationary tessellations with examples. *Stochastic Process. Appl.*, 124(9):2917–2953, 2014.

[24] N. Chenavier. The extremal index for a random tessellation. In *Geometric science of information*, volume 9389 of *Lecture Notes in Comput. Sci.*, pages 171–178. Springer, Cham, 2015.

[25] N. Chenavier, D. Coupier, and A. Rousselle. The bi-dimensional directed IDLA forest. *To appear in the Annals of Applied Probability; available at https://arxiv.org/pdf/2009.12090.pdf*, 2022+.

[26] N. Chenavier and A. Darwiche. Extremes for transient random walks in random sceneries under weak independence conditions. *Statist. Probab. Lett.*, 158:108657, 6, 2020.

[27] N. Chenavier and O. Devillers. Stretch factor in a planar Poisson-Delaunay triangulation with a large intensity. *Adv. in Appl. Probab.*, 50(1):35–56, 2018.

[28] N. Chenavier and R. Hemsley. Extremes for the inradius in the Poisson line tessellation. *Adv. in Appl. Probab.*, 48(2):544–573, 2016.

[29] N. Chenavier, N. Henze, and M. Otto. Limit laws for large k th-nearest neighbor balls. *J. Appl. Probab.*, 59(3):880–894, 2022.

[30] N. Chenavier and C. Hirsch. Extremal lifetimes of persistent cycles. *Extremes*, 25(2):299–330, 2022.

[31] N. Chenavier, B. Massé, and D. Schneider. Products of random variables and the first digit phenomenon. *Stochastic Process. Appl.*, 128(5):1615–1634, 2018.

[32] N. Chenavier and W. Nagel. The largest order statistics for the inradius in an isotropic STIT tessellation. *Extremes*, 22(4):571–598, 2019.

[33] N. Chenavier and C. Y. Robert. Cluster size distributions of extreme values for the Poisson-Voronoi tessellation. *The Annals of Applied Probability*, 28(6):3291–3323, 2018.

[34] N. Chenavier and D. Schneider. On the discrepancy of powers of random variables. *Statist. Probab. Lett.*, 134:5–14, 2018.

[35] R. Cowan. Properties of ergodic random mosaic processes. *Math. Nachr*, 97:89–102, 1980.

[36] L. de Haan. A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204, 1984.

[37] L. de Haan and A. Ferreira. *Extreme value theory. An introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[38] O. Devillers, S. Pion, and M. Teillaud. Walking in a triangulation. volume 13, pages 181–199. 2002. Volume and surface triangulations.

[39] P. Diaconis. The distribution of leading digits and uniform distribution mod 1. *Ann. Probability*, 5(1):72–81, 1977.

[40] D. P. Dobkin, S. J. Friedman, and K. J. Supowit. Delaunay graphs are almost as good as complete graphs. *Discrete Comput. Geom.*, 5:399–407, 1990.

[41] C. Dombry and F. Eyi-Minko. Strong mixing properties of max-infinitely divisible random fields. *Stochastic Process. Appl.*, 122(11):3790–3811, 2012.

[42] C. Dombry and F. Eyi-Minko. Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab.*, 18:no. 7, 21, 2013.

[43] C. Dombry, F. Éyi Minko, and M. Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124, 2013.

[44] C. Dombry and Z. Kabluchko. Random tessellations associated with max-stable random fields. *Bernoulli*, 24(1):30–52, 2018.

[45] H. Duminil-Copin, C. Lucas, A. Yadin, and A. Yehudayoff. Containing internal diffusion limited aggregation. *Electron. Commun. Probab.*, 18:no. 50, 8, 2013.

[46] H. Duminil-Copin, A. Raoufi, and V. Tassion. Subcritical phase of d-dimensional Poisson-Boolean percolation and its vacant set. *Ann. H. Lebesgue*, 3:677–700, 2020.

[47] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. volume 28, pages 511–533. 2002. Discrete and computational geometry and graph drawing (Columbia, SC, 2001).

[48] S. Eliahou, B. Massé, and D. Schneider. On the mantissa distribution of powers of natural and prime numbers. *Acta Math. Hungar.*, 139(1-2):49–63, 2013.

[49] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.

[50] B. Franke and T. Saigo. The extremes of random walks in random sceneries. *Advances in Applied Probability*, 41(2):452–468, 2009.

[51] L. Györfi, N. Henze, and H. Walk. The limit distribution of the maximum probability nearest-neighbour ball. *J. Appl. Probab.*, 56(2):574–589, 2019.

[52] L. Heinrich, H. Schmidt, and V. Schmidt. Central limit theorems for Poisson hyperplane tessellations. *Ann. Appl. Probab.*, 16(2):919–950, 2006.

[53] L. Heinrich and V. Schmidt. Normal convergence of multidimensional shot noise and rates of this convergence. *Adv. in Appl. Probab.*, 17(4):709–730, 1985.

[54] N. Henze. The limit distribution for maxima of "weighted" rth-nearest-neighbour distances. *J. Appl. Probab.*, 19(2):344–354, 1982.

[55] C. Hirsch, D. Neuhäuser, and V. Schmidt. Moderate deviations for shortest-path lengths on random segment processes. *ESAIM Probab. Stat.*, 20:261–292, 2016.

[56] T. Hsing, J. Hüsler, and M. R. Leadbetter. On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields*, 78(1):97–112, 1988.

[57] D. Hug, M. Reitzner, and R. Schneider. Large Poisson-Voronoi cells and Crofton cells. *Adv. in Appl. Probab.*, 36(3):667–690, 2004.

[58] D. Hug, M. Reitzner, and R. Schneider. The limit shape of the zero cell in a stationary Poisson hyperplane tessellation. *Ann. Probab.*, 32(1B):1140–1167, 2004.

[59] R. Huser and A. C. Davison. Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518, 2013.

[60] S. R. Jammalamadaka and S. Janson. Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. *Ann. Probab.*, 14(4):1347–1358, 1986.

[61] D. Jerison, L. Levine, and S. Sheffield. Internal DLA in higher dimensions. *Electron. J. Probab.*, 18:No. 98, 14, 2013.

[62] D. Jerison, L. Levine, and S. Sheffield. Internal DLA and the Gaussian free field. *Duke Math. J.*, 163(2):267–308, 2014.

[63] Z. Kabluchko. Extremes of independent Gaussian processes. *Extremes*, 14(3):285–310, 2011.

[64] Z. Kabluchko and M. Schlather. Ergodic properties of max-infinitely divisible processes. *Stochastic Process. Appl.*, 120(3):281–295, 2010.

[65] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042–2065, 2009.

[66] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

[67] R. Lachièze-Rey and S. Vega. Boundary density and Voronoi set estimation for irregular sets. *Trans. Amer. Math. Soc.*, 369(7):4953–4976, 2017.

[68] G. F. Lawler. Subdiffusive fluctuations for internal diffusion limited aggregation. *Ann. Probab.*, 23(1):71–86, 1995.

[69] G. F. Lawler, M. Bramson, and D. Griffeath. Internal diffusion limited aggregation. *Ann. Probab.*, 20(4):2117–2140, 1992.

[70] J.-F Le Gall and J. Rosen. The range of stable random walks. *Ann Proba*, 19(2):650–705, 1991.

[71] M. R. Leadbetter. Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete*, 65(2):291–306, 1983.

[72] L. Levine and Y. Peres. Scaling limits for internal aggregation models with multiple sources. *J. Anal. Math.*, 111:151–219, 2010.

[73] L. Levine and V. Silvestri. How long does it take for internal DLA to forget its initial profile? *Probab. Theory Related Fields*, 174(3-4):1219–1271, 2019.

[74] C. Lucas. The limiting shape for drifted internal diffusion limited aggregation is a true heat ball. *Probab. Theory Related Fields*, 159(1-2):197–235, 2014.

[75] S. Martínez and W. Nagel. STIT tessellations have trivial tail σ-algebra. *Adv. in Appl. Probab.*, 46(3):643–660, 2014.

[76] S. Martínez and W. Nagel. The β-mixing rate of STIT tessellations. *Stochastics 88*, 3:396–414, 2016.

[77] W. Nagel and V. Weiss. Crack STIT tessellations: characterization of stationary random tessellations stable with respect to iteration. *Adv. in Appl. Probab.*, 37(4):859–883, 2005.

[78] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. From Stein's method to universality.

[79] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, second edition, 2000.

[80] T. Owada and Z. Wei. Functional strong law of large numbers for betti numbers in the tail. *Available at https://arxiv.org/pdf/2103.05799.pdf*, 2021.

[81] M. Penrose. *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[82] M. D. Penrose. The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.*, 7(2):340–361, 1997.

[83] R. Perfekt. Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab.*, 4(2):529–548, 1994.

[84] P. N. Rathie. On the volume distribution of the typical Poisson-Delaunay cell. *J. Appl. Probab.*, 29(3):740–744, 1992.

[85] J. Rivat and G. Tenenbaum. Constantes d'Erdös-Turán. *Ramanujan J.*, 9(1-2):111–121, 2005.

[86] C. Y. Robert. Inference for the limiting cluster size distribution of extreme values. *Ann. Statist.*, 37(1):271–310, 2009.

[87] C. Y. Robert. Power variations for a class of Brown-Resnick processes. *Extremes*, 23(2):215–244, 2020.

[88] H. Rootzén. Maxima and exceedances of stationary Markov chains. *Adv. in Appl. Probab.*, 20(2):371–390, 1988.

[89] M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.

[90] R. Schneider and W. Weil. *Stochastic and Integral Geometry.* Probability and its Applications (New York). Springer-Verlag, Berlin, 2008.

[91] M. Schulte. A central limit theorem for the Poisson-Voronoi approximation. *Adv. in Appl. Math.*, 49(3-5):285–306, 2012.

[92] M. Schulte and C. Thäle. The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Process. Appl.*, 122(12):4096–4120, 2012.

[93] E. Shellef. IDLA on the supercritical percolation cluster. *Electron. J. Probab.*, 15:no. 24, 723–740, 2010.

[94] R.L. Smith. Max-stable processes and spatial extremes. *unpublished*, 1990.

[95] R.L. Smith and I. Weissman. Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B*, 56(3):515–528, 1994.

[96] S. A. Stoev. Max-stable processes: representations, ergodic properties and statistical applications. In *Dependence in probability and statistics*, volume 200 of *Lect. Notes Stat.*, pages 21–42. Springer, Berlin, 2010.

[97] C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statist. Sinica*, 21(1):5–42, 2011.

[98] G. Xia. The stretch factor of the delaunay triangulation is less than 1.998. *SIAM J. Comput.*, 42:1620–1659, 2013.

[99] G. Xia and L. Zhang. Toward the tight bound of the stretch factor of delaunay triangulations. *Proceedings 23th Canadian Conference on Computational Geometry*, 2011.

[100] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

[101] S. A. Zuyev. Estimates for distributions of the Voronoi polygon's geometric characteristics. *Random Structures Algorithms*, 3(2):149–162, 1992.