Lecture Notes on Random Geometric Models — Random Graphs, Point Processes and Stochastic Geometry

Bartlomiej Blaszczyszyn

Lecture Notes on

Random Geometric Models

Random Graphs, Point Processes and Stochastic Geometry

Bartłomiej Błaszczyszyn

December 4, 2017

Preface

The goal of this sequence of lessons is to provide quick access to some popular models of random geometric structures used in many applications: from communication networks, including social, transportation, wireless networks, to geology, material sciences and astronomy. The course is composed of the following 15 lessons:

1. Bond percolation on the square lattice
2. Galton-Watson tree
3. Erdős-Rényi graph — emergence of the giant component
4. Graphs with a given node degree distribution
5. Typical nodes and random unimodular graphs
6. Erdős-Rényi graph — emergence of the full connectivity
7. Poisson point process
8. Point conditioning and Palm theory for point processes
9. Hard-core point processes
10. Stationary point processes and mass transport principle
11. Stationary Voronoi tessellation
12. Ergodicity and point-shift invariance
13. Random closed sets
14. Boolean model and coverage processes
15. Connectedness of random sets and continuum percolation

Usually, these subjects are presented in different monographs: random graphs (lessons 2–6), point processes (7-12), stochastic geometry (13-14), with percolation models presented in lesson 1 and 15 often addressed separately. Having them in one course gives us an opportunity to observe some similarities and even fundamental relations between different models. Examples of such connections are:

- Similar phase transitions regarding the emergence of big components observed in different discrete, lattice and continuous euclidean models (lessons 1–4, 15).
- Single isolated nodes being the last obstacle in the emergence of the full connectivity in some discrete and euclidean graphs exhibiting enough independence (lessons 6, 15).
- A mass transport principle as a fundamental property for unimodular random graphs and Palm theory for stationary point processes; with both theories seeking to define the typical node/point of a homogeneous structure (lessons 5, 10–12).

- Poisson-Galton-Watson tree and Poisson process playing a similar role in the theory of random graphs and point processes, respectively: for both models independence and Poisson distribution are the key assumptions, both appear as natural limits, and both rooted/conditioned to a typical node/point preserve the distribution of the remaining part of the structure (lessons 2,5, 7–8).
- Size biased sampling appearing in several, apparently different, conditioning scenarios, as unimodular trees (lesson 5), Palm distributions for point process (lesson 8), zero cell of the stationary tessellations (lessons 11).

The goal of this series of lectures is to present some spectrum of models and ideas. When doing this, we sometimes skip more technical proof details, sending the reader for them to more specialised monographs. Some theoretical and computer exercises are provided after each lesson to let the reader practice his/her skills. Regarding the prerequisites, the reader will benefit from having had some prior exposure to probability and measure theory, but this is not absolutely necessary.

The content of the course has been evolving while the author teaches it within the master programme *Probabilité et modelés aléatoires* at the University Pierre and Marie Curie in Paris. The author is grateful to Mohamed Karray for many stimulating discussions at various stages of the development of this course.

The present notes were thoroughly revised when the author was presenting them as a specially appointed professor at the School of Computing, Tokyo Institute of Technology, in the autumn term 2017. The author would like to express his gratitude to Naoto Miyoshi for the kind invitation to Tokyo Tech.

Tokyo, December 4, 2017.

Contents

Lesson 1

Bond percolation on the square lattice

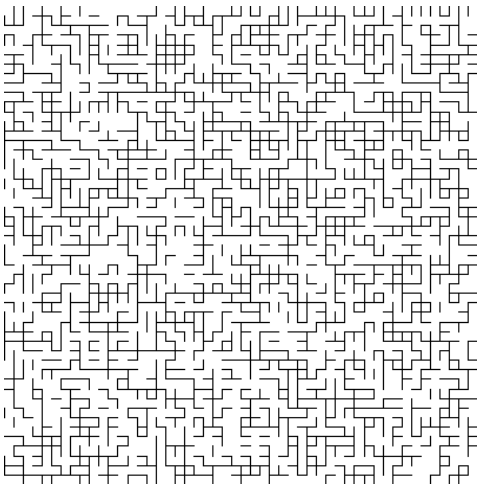


Figure 1.1: Detail of a realization of a bond percolation on the square lattice in two dimensions with percolation probability p = 0.51; by Erzbischof - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=655409

In mathematics (especially in percolation theory) by *percolation* one understands the *existence of an infinite connected component* of a (necessarily) infinite, random structure such as a graph, a tree, a random set, etc... The term comes from the study of porous media, through which a liquid can percolate (i.e. penetrate, trickle, filter) if the structure of pores (voids) is rich enough and well interconnected. Besides material sciences and geology the theory of percolation has applications to the modeling and analysis of *communication and social networks* Franceschetti and Meester ((2008)) (and also some other domains of science, e.g. astronomy) where percolation indicates some good *large scale, macroscopic connectivity*

of the model (which is not necessarily fully connected).

In what follows we shall introduce perhaps the simplest mathematical model for which we shall be able to prove the existence of the *phase transition* related to the percolation. This is the transition between two regimes: *sub-critical* one, in which there is no infinite connected component of the model and *super-critical* one, where such component exists almost surely (and moreover is unique). This is the square lattice bond percolation model introduced by Simon Broadbent and John Hammersley in 1957 Broadbent and Hammersley ((1957)).

Throughout these notes we shall make references to Grimmett ((2013)) and the reader is invited to consult this textbook for more mathematical results regarding the considered model.

1.1 Independent bond percolation on the square lattice

1.1.1 The square lattice

Denote by $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ integers numbers. The d-dimensional ($d \geq 1$) *square lattice* is a graph with vertexes (or nodes) $\mathbb{Z}^d$ and edges $\mathbb{E}^d := \{e = \{z_1, z_2\} : |z_1 - z_2| = 1, z_1, z_2 \in \mathbb{Z}^d\}$. Note that $(\mathbb{Z}^d, \mathbb{E}^d)$ is an infinite deterministic graph. In what follows we shall consider a random subset of this graph.

1.1.2 Independent bond percolation

Consider a family of independent Bernoulli random variables $\{Z(e) : e \in \mathbb{E}^d\}$ indexed by the edges of the square lattice, such that $\mathbf{P}\{\, X(e) = 1 \,\} = 1 - \mathbf{P}\{\, X(e) = 0 \,\} = p$ for some $p \in [0, 1]$, which the parameter of our model (besides the dimension d).

We say the edge $e \in \mathbb{E}^d$ is *open* if $X(e) = 1$. Otherwise it is *closed*. Denote by

$$\Xi := \Big(\mathbb{Z}^d, \{e \in \mathbb{E}^d : X(e) = 1\}\Big)$$

the random sub-graph of the square lattice $(\mathbb{Z}^d, \mathbb{E}^d)$ consisting of all its vertexes $\mathbb{Z}^d$ and only *open edges*.

Denote by C the maximal (in terms of the inclusion) connected subset of Ξ containing the origin $\mathbf{0} \in \mathbb{Z}^d$. We call C the *connected component of the origin*. Denote by

$$\theta(p) = \mathbf{P}\{\, |C| = \infty \,\}$$

the probability that the connected component of the origin is infinite (consists of an infinite number of vertexes or, equivalently, infinite number of edges). We call $\theta(p)$ the *percolation function*. If $|C| = \infty$ we say that *the origin percolates*.

Obviously $0 \leq \theta(p) \leq 1$, with $\theta(0) = 1$ and $\theta(1) = 1$. Also, $\theta(p)$ is a non-decreasing function, cf Exercise 1. Let us define the *critical value* of the parameter p

$$p_c := \sup\{p \in [0, 1] : \theta(p) = 0\}$$

as the supermom of the values of p for which the origin does not percolates.[1]

By the monotonicity of $\theta(p)$ we know that for $p \in [0, p_c)$ the connected component of the origin is almost surely finite; $\theta(p) = 0$. We call this phase (regime) *sub-critical*.

We also know that for $p \in (p_c, 1]$ the connected component of the origin is infinite with some positive probability; $\theta(p) > 0$. We call this phase (regime) *super-critical*.

We do not know however yet whether $0 < p_c < 1$, that is, whether the two regimes really exist. (Note that if $p_c = 0$ then $[0, p_c)$ is empty and similarly for $p_c = 1$.) This is the question about the existence of the non-trivial phase transition for our independent bond percolation model. We shall answer it in the next section.

1.2 Non-trivial phase transition

Theorem 1.2.1. *If dimension $d \geq 2$ then $0 < p_c < 1$.*

We shall prove this result in the next two sections. Before, let us make a few remarks.

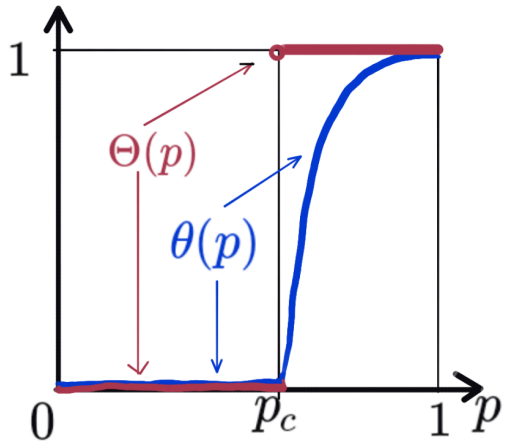
- Firstly, the assumption $d \geq 2$ is necessary. For $d = 1$ one has $p_c = 1$; cf Exercise 2.
- The exact value of p_c is not known, except for the dimension $d = 2$ for which $p_c = 1/2$. This is far from trivial to show. The known argument, cf ((Grimmett, 2013, Section 11.3)) involves some dual lattice that will be introduced in in Section 1.2.1 and the fact that it has the same structure as the original lattice. The values of p_c for higher dimensions can be estimated by simulation.[2]
- One can also ask what is the probability that there is *some* infinite connected component of Ξ, not necessarily containing the origin. Denote this probability by $\Theta(p)$. Obviously for $\Theta(p) \geq \theta(p)$ and thus $\Theta(p) > 0$ for $p > p_c$. One can show more. In fact we have a 0-1 law for $\Theta(p)$:

$$\Theta(p) = \begin{cases} 0 & \text{for } p < p_c \\ 1 & \text{for } p > p_c. \end{cases} \tag{1.2.1}$$

[1] An interesting, equivalent, definition of p_c has been recently proposed in Duminil-Copin and Tassion ((2015)).

[2] Good estimation of these values is itself a non-trivial task as one cannot simulate an infinite model. The simplest approach consists in simulating a large finite window and considering the existence of paths traversing the window in different directions as an indicator of the percolation. Even better results are obtained if the window is "wrapped around" to form a torus and one looks for the paths connecting the same pints on the opposite sides (facets) of the window. The following table shows the critical probability p_c for the bond percolation in different dimensions d. The values for $d \geq 3$, estimated by simulation, are believed to be exact to the last but one decimal; from ((Stauffer and Aharony, 2003, Tabele 1, page 17)), that presents a more physical approach to the theory of percolation. Note the values of p_c decrease with the dimension, which can be proved by a natural embedding of $\mathbb{Z}^d$ into $\mathbb{Z}^{d+1}$.

dimension d	1	2	3	4	5	6	7
critical probability p_c	1	1/2	0.2488	0.1601	0.1182	0.0942	0.0787



The intuition behind the fact that $\Theta(p) = 0$ for $p < p_c$ is as follows: the origin can be considered as the typical node [3] of the graph, properties of this node are representative for all nodes. In particular the value of $\theta(p)$ does not depend on our choice of the origin. Since this typical node does not have any chance to be in an infinite component, the same remains true for any other node. Hence there cannot be any infinite component. For a formal proof see Exercise 3.

The fact that $\Theta(p) = 1$ for $p > p_c$ is the consequence of Kolmogorov's zero-one law; see Exercise 4.

- When $p > p_c$ almost surely there is *exactly one* infinite component of Ξ. It is often called *the giant component* (often interpreted in the context of networks as the connected "core" of the network). The prove of this fact is more involved. The reader can consult ((Grimmett, 2013, Section 8.1)); cf also Exercise 5.

- The study of the percolation problem at the *critical* value $p = p_c$ is much more difficult. It has been proved that there is *no infinite connected component* at $p = p_c$ for the dimensions $d = 2$ and $d \geq 19$, and is generally believed (especially by the physicists) that this remains true for all $d \geq 2$. Moreover, following the predictions and discoveries of more applied scientists one conjectures that $\theta(p)$ decays to 0 on the right-hand-side of the critical point in the manner of a power of $p - p_c$

$$\theta(p) \approx (p - p_c)^{\beta} \quad \text{as } p \searrow p_c, \quad \text{for some } \beta > 0,$$

with the value of β, called the *critical exponent*, depending on the dimension d. Theoretical physics predict the value $\beta = 5/36$ for $d = 2$ and are able to estimate it for

[3] The notion of the *typical element* of a random structure (typical vertex of a random graph, typical point of a point process) is one of the most important notions of this course, it will be made formal in subsequent lectures. Here we mean that the distribution of Ξ is invariant with respect to any translation in $\mathbb{Z}^d$, in other words Ξ "observed" from any node $z \in \mathbb{Z}^d$ "looks" (in distribution) the same as Ξ observed from the origin.

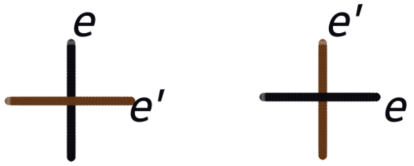
higher dimensions; cf ((Stauffer and Aharony, 2003, Tabele 2, page 52)). For more mathematical explanations consult ((Grimmett, 2013, Chapters 9,10)).

1.2.1 Proof of $p_c < 1$ using Peierls argument

By a natural embedding of $\mathbb{Z}^d$ into $\mathbb{Z}^{d+1}$ we observe that the value of $\theta(p)$ is a non-decreasing function of the dimension d. Thus, it is enough to prove that $p_c < 1$ for $d = 2$. In this section we shall prove it using the famous duality argument proposed by Rudolf Peierls.

Assume $d = 2$. We have to prove that $\theta(p) > 0$ or, equivalently, $\mathbf{P}\{\,|C| < \infty\,\} < 1$ for sufficiently large $p < 1$.

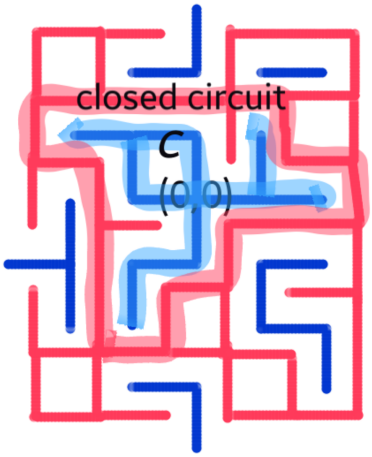
Let us consider a *dual* lattice $(\tilde{\mathbb{Z}}^2, \tilde{\mathbb{E}}^2) = (\mathbb{Z}^2 + (1/2, 1/2), \mathbb{E}^2 + (1/2, 1/2))$, which is our original square lattice on the plane translated by the vector $(1/2, 1/2)$. We consider a 1-1 mapping of the edges of $\mathbb{E}^2$ to $\tilde{\mathbb{E}}^2$: $e \in \mathbb{E}^2$ and $\tilde{e} \in \tilde{\mathbb{E}}^2$ are *dual to each other* if $\tilde{e}$ is the bisection of e.



Using the original random variables $\{X(e) : e \in \mathbb{E}^2\}$ we define random variables for the dual edges $\{\tilde{X}(\tilde{e}) : \tilde{e} \in \tilde{\mathbb{E}}^2\}$ such that $\tilde{X}(\tilde{e}) = X(e)$. Similarly to the original lattice, we say that a dual edge $\tilde{e}$ is open when $\tilde{X}(\tilde{e}) = 1$ and otherwise closed.

By a *circuit around the origin* we understand a finite, closed path (loop) on the dual lattice $(\tilde{\mathbb{Z}}^2, \tilde{\mathbb{E}}^2)$. A circuit around the origin is called *closed* if it is composed of the closed edges (edges $\tilde{e}$ such that $\tilde{X}(\tilde{e}) = 0$).

The following observation is crucial: *The connected component of the origin (open on the original lattice) is finite ($|C| = \infty$) if and only if (iff) there exists a closed circuit around the origin in the dual graph.* (In fact the implication $\Rightarrow$ is enough for our purpose.)



The (finite!) closed circuit on the dual lattice is the one that "surrounds" the open component of the origin on the original lattice, preventing it from "going" to infinity in any direction. It is somewhat tedious to prove this with complete rigour, and we shall not do so here. Note that we can also require the closed circuit to be self-avoiding, as from any circuit we can also choose a sub-circuit that is self-avoiding.

Following the above observation we have

$$\begin{aligned} 1-\theta(p) &= \mathbf{P}\{\,|C|<\infty\,\} \\ &= \mathbf{P}\{\,\text{there exists a closed self-avoiding circuit around the origin}\,\} \\ &= \mathbf{E}\,[\text{number of closed self-avoiding circuits around the origin}] \\ &= \sum_{n=4}^{\infty}\rho(n)(1-p)^n, \end{aligned}$$

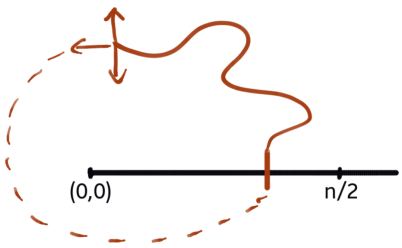
Markov inequality

where $\rho(n)$ denotes the number of self-avoiding circuits around the origin composed of exactly n edges and $(1-p)^n$ is the probability that any given such circuit is closed. Note that the above sum starts at $n=4$ as it is the length of the shortest circuit around the origin.

The following relatively loose upper bound for $\rho(n)$ can be given:

$$\rho(n) \le n3^{n-1}.$$

It can be justified in the following way: the circuit around the origin has to cross the half-line $[0,\infty)$ least once and this has to take place at some point $(k+1/2,0)\in\mathbb{Z}^2$ with $k=0,1,\ldots$, $k\le n/2$ (otherwise the circuit of length n cannot surround the origin). There are hence at most $n/2\le n$ such possible crossing points. Having placed this closest to the origin edge crossing $[0,\infty)$ we extend the path (to a circuit around the origin) adding edges one by one, each time making one of the three possible choices: go ahead, turn right or turn left (turning back is forbidden as the circuit has to be self-avoiding). This gives us 3^{n-1} possible choices. (The bound is very rough as many of such choices do not lead to the construction of a circuit).



Consequently

$$\mathbf{P}\{\,|C|<\infty\,\}\le\sum_{n=4}^{\infty}n3^{n-1}(1-p)^n\to 0\quad\text{when } 2/3\le p\to 1$$

Hence, for p large enough in the interval $(2/3,1)$

$$\mathbf{P}\{\,|C|<\infty\,\}\le\sum_{n=4}^{\infty}n3^{n-1}(1-p)^n<1$$

making $\theta(p)=1-\mathbf{P}\{\,|C|<\infty\,\}>0$. This completes the proof that $p_c<1$.

1.2.2 Proof of $p_c>0$ using open path counts

Our proof that $\theta(p)=0$ for $p>0$ small enough involves counting open paths from the origin on the original d-dimensional lattice $(\mathbb{Z}^d,\mathbb{E}^d)$. For any given $n\ge 1$ (large n will be interesting for us)

$$\begin{aligned}\theta(p)&=\mathbf{P}\{\,|C|=\infty\,\}\\&\le\mathbf{P}\{\,\text{there exists a self-avoiding open path of length } n\text{, from the origin}\,\}\\&=\mathbf{E}\,[\text{number of self-avoiding open paths of length } n\text{, from the origin}]\\&=\sigma(n)p^n,\end{aligned}$$

Markov inequality

where $\sigma(n)$ denotes the number of self-avoiding paths from the origin composed of exactly n edges and p^n is the probability that any given such path is open.

Again, the following relatively loose upper bound for $\sigma(n)$ can be given, which involves dimension d:

$$\sigma(n)\le 2d(2d-1)^{n-1}.$$

In $2d$ we recognize the number of possibilities to start the path at the origin and $(2d-1)^{n-1}$ is a (loose) bound for the number of possibilities to complete it to a self-avoiding path of length n.

This gives us the family of uppers bounds valid for all $n \geq 1$

$$\theta(p) = \mathbf{P}\{\, |C| = \infty \,\} \leq 2d(2d-1)^{n-1}p^{n-1} \to 0 \quad \text{as } n \to \infty; \text{ provided } p < 1/(2d-1).$$

Consequently $\theta(p) = 0$ for $p < 1/(2d-1)$ making $p_c \geq 1/(2d-1)$. This completes the proof of Theorem 1.2.1.

1.3 Exercises

1. Show that $\theta(p)$ is a non-decreasing function on $[0,1]$. Hint: consider independent random variables $\{U(e) : e \in \mathbb{E}^d\}$ uniform on $[0,1]$ and define random variables $\{X(e) : e \in \mathbb{E}^d\}$ such that $X(e) = \mathbf{1}U(e) \leq p$.

2. Show that for $d = 1$ $p_c = 1$; i.e., the one-dimensional lattice model does not percolate except trivially for $p = 1$.

3. Show that if $\theta(p) = 0$ then $\Theta(p) = 0$. Hint: $\Theta(p) \leq \sum_{z\in\mathbb{Z}^d} \mathbf{P}\{\text{z belongs to an infinite connected component of } \Xi\} \leq \sum_{z\in\mathbb{Z}^d} \theta(p)$.

4. Show that if $\Theta(p) > 0$ then $\Theta(p) = 1$. Hint: one can use the classical Kolmogorov's zero-one law for the *tail events* generated by independent, identically distributed (iid) random variables. Indeed, note that the event {there exits an infinite connected component of Ξ} is measurable with respect to $\{X(e) : e \in \mathbb{E}^d$ outside $[-n,n]^d\}$ for any n (it does not depend on any modification of the status of any finite number of edges).

5. Show that the number of infinite components of Ξ is almost surely constant; Hint: for any given k, the event {number of the infinite components of $\Xi = k$} is invariant with respect to the translation of Ξ by any vector $z \in \mathbb{Z}^d$. Indeed, each infinite component, if it exists, will be only translated; it cannot disappear nor be created. All such invariant events in our model generated by iid random variables $\{X(e) : e \in \mathbb{E}^d\}$ have probability 0 or 1. We shall explain it in more details in the lesson devoted to the *ergodicity* in the geometric context.

6. *1-dependent bond percolation.* Consider a modification of our bond percolation model in which the random variables $\{X(e) : e \in \mathbb{E}^d\}$ are 1-dependent; i.e., for any $k \geq 2$ and any edges $e_1, \ldots, e_k \in \mathbb{E}^d$ such that no two of them share any common vertex the random variables $X(e_1), \ldots, X(e_k)$ are independent. (Edges which share some vertex might not be independent.) Show that for such 1-dependent bond-percolation model $0 < p_c < 1$. Generalize the result to m-dependent model with $m \geq 1$. Hint: Find an upper bound for the probability that a path of length n is opened (closed). Use the same arguments as for the independent percolation model.

7. *Computer exercise.* Estimate the percolation function $\theta(p)$ and the critical probability p_c for the independent bond percolation on the planar square lattice ($d = 2$); cf remarks in Footnote 2.

Lesson 2

Galton-Watson tree

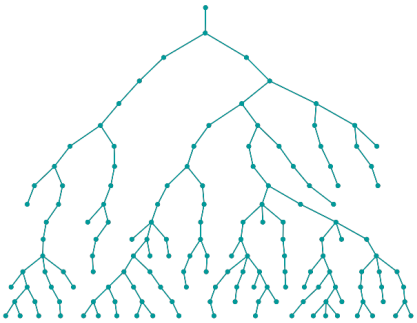


Figure 2.1: A simulation of a few generations of a genealogy tree, in which each individual independently gives birth to $0, 1, 2, 3$ children with probability, respectively, $1/8, 3/8, 3/8, 1/8$. Borrowed from Berglund ((2013)).

The Galton-Watson process takes its name from two British scientists of the 19th century: Francis Galton and Henry William Watson, who in their joint paper published in 1874 Watson and Galton ((1875)) have addressed some concerns amongst the Victorians that aristocratic family names were becoming extinct. Today, Galton-Watson process is a fundamental model of interest in the theory of *branching processes* Athreya and Ney ((2004)). Its applications go well beyond genealogical and genetic problems. It is a fundamental building block in the study of random graphs, particularly related to several models aiming to describe the structure of social networks. We shall introduce some of these models in subsequent lessons.

In this lesson we present the Galton-Watson process as a *random tree*. Its fundamental assumption is that each individual gives births *independently* to some number of children (called also offsprings) and the distribution of this offspring number is common for all individuals.

We shall study first the *extinction probability* of the Galton-Watson tree; i.e., the probability that the tree is finite. This question is of the same nature as the question about the percolation, which we have already studied on the square lattice. We shall observe similar

phase transition related to the extinction probability of the Galton-Watson tree: when the mean offspring number is smaller than 1 (*sub-critical* regime) then the tree is almost surely finite. When it is larger than 1 (*super-critical* regime) the tree has a positive probability to grow infinitely. The structure of the tree makes the study of the extinction probability more explicit than the study of the percolation function on the lattice.

In the second part of this lesson we shall study the *conditional distribution of the super-critical Galton-Watson tree, given its extinction*. In other words, how does a Galton-Watson tree look like, when it has a positive probability to grow infinitely, but this event does not happen and the tree remains finite. We shall observe that it coincides with the distribution of some *dual* Galton-Watson tree, which is sub-critical, it does not have chance to grow infinitely. This somewhat surprising observation (conditioning on the extinction could have changed the probabilistic structure of the tree, which is not the case) is an important investment in the analysis of some other random graph models that we shall see in subsequent lessons. In particular, when proving this result, we shall introduce an *exploration* process of the GW tree, that is a useful technique that will be later used to study connected components of more complicated random graphs.

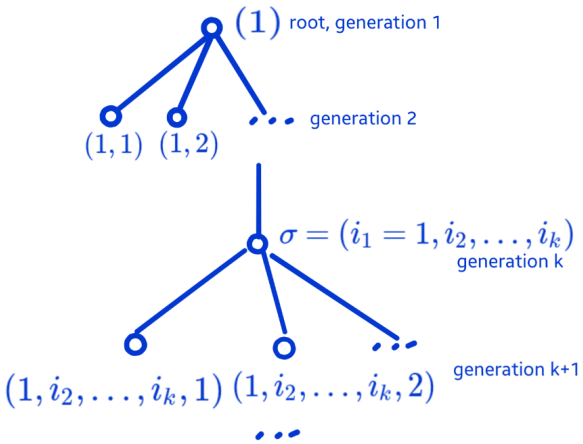
Finally we shall given some bound on the total population of the sub-critical Galton-Watson tree.

Our exposition on Galton-Watson process corresponds roughly to the results presented in ((Draief and Massoulié, 2010, Chapter 1)); see also ((Van Der Hofstad, 2017, Chapter 3)).

2.1 Extinction or percolation

2.1.1 Tree

Our tree will grow form its root downwards. We shall introduce some labeling of the nodes of the tree and distinguish its generations.



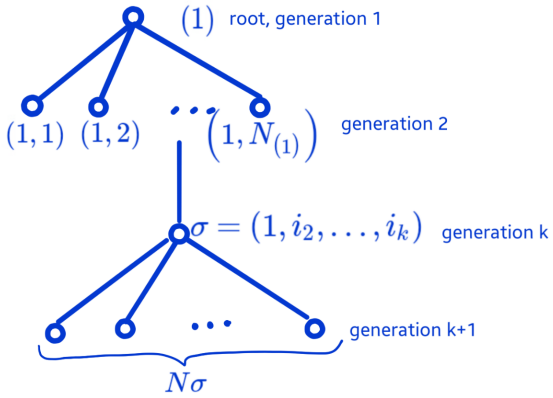
Denote by $\mathbb{S}$ the set of all possible tree nodes

$$\mathbb{S} := \Big\{(i_1 = 1, i_2, \dots, i_k) : k \geq 1, i_j \geq 1\Big\}.$$

$\sigma = (1)$ is the *tree root*.
$\sigma = (1, i_2, \dots, i_k)$ is a node of *generation* $k = |\sigma|$, where $|\sigma|$ is the length of σ.
All nodes $\{(1, i_2, \dots, i_k, i_{k+1} : i_{k+1} \geq 1\}$ are *children* of σ. They belong to the generation $k+1$.

2.1.2 The Galton-Watson tree

The Galton-Watson (GW) tree is a random tree [1] in which each node gives birth to an independent, identically distributed number of children.



Formally: let $\{N_\sigma : s \in \mathbb{S}\}$ i.i.d. r.v. taking values in $\{0, 1, \dots\}$ with $N\sigma \stackrel{\mathcal{L}}{=} N$ a generic r.v. whose law is called the *offspring distribution*. The Galton-Watson tree is a subset of $\mathbb{S}$

$$\mathrm{GW} := \left\{\sigma \in \mathbb{S} : i_1 = 1, i_2 \leq N_{(1)}, i_3 \leq N_{(1,i_2)}, \dots, i_k \leq N_{(1,i_1,\dots,i_{k-1})}\right\}. \tag{2.1.1}$$

Offspring distribution Let

$$p_k := \mathbf{P}\{\, N = k \,\},$$

$$m := \mathbf{E}\,[N] = \sum_{k=0}^{\infty} k p_k \qquad \text{mean offspring, expected number of children},$$

$$\phi(s) := \mathbf{E}\left[s^N\right] = \sum_{k=0}^{\infty} p_k s^k, \quad s \in [0,1] \qquad \text{probability generating function (pgf)}.$$

Generation recursion Let

$$X_n := \#\{\sigma \in \mathrm{GW} : |\sigma| = n\} \quad \text{number of individuals in the } n\text{ th generation of GW}.$$

[1] called also the Galton-Watson *branching process*

$$
\begin{aligned}
X_1 &= 1\,,\\
X_2 &= N_{(1)}\,,\\
&\vdots\\
X_n &= \sum_{\substack{\sigma\in \mathrm{GW}\\ |\sigma|=n-1}} N_\sigma \qquad n\geq 2.
\end{aligned}
$$

Stochastic recursion for $n \geq 2$

$$
X_n \stackrel{\mathcal{L}}{=} \sum_{i=1}^{X_{n-1}} N_i \qquad N_i \stackrel{\mathcal{L}}{=} N, \text{i.i.d., independent of } X_{n-1}\,. \tag{2.1.2}
$$

Denote

$$
S := \sum_{n=1}^{\infty} X_n \qquad \text{total number of individuals (size) of the GW tree}\,.
$$

$S < \infty$ if and only if (iff) $X_n = 0$ for sufficiently large n; we say then there is *extinction* of the process (tree). If $S = \infty$ then GW is an infinite tree, we observe a *non-extinction*; one can one can say the GW tree *percolates*.

2.1.3 Extinction probability

Denote

$$
p_{\mathrm{ext}} := \mathbf{P}\{\, S < \infty \,\} \qquad \text{extinction probability}\,.
$$

Theorem 2.1.1. *p_{ext} is the smallest solution in s of the equation*

$$
s = \phi(s) \quad \textit{for } s \in [0,1]\,. \tag{2.1.3}
$$

Proof. Denote $\phi_n(s) := \mathbf{E}\left[s^{X_n}\right]$. By (2.1.2) we have for $n \geq 2$

$$
\begin{aligned}
\phi_n(s) &= \sum_{k=0}^{\infty} \mathbf{P}\{\, X_{n-1} = k \,\}\, \mathbf{E}\left[s^{\sum_{i=1}^{k} N_i} \,|\, X_{n-1} = k\right]\\
\boxed{\text{independence}} \quad &= \sum_{k=0}^{\infty} \mathbf{P}\{\, X_{n-1} = k \,\}\, (\mathbf{E}\left[s^{N}\right])^k\\
&= \mathbf{E}\left[(\phi(s))^{X_{n-1}}\right]\\
&= \phi_{n-1}(\phi(s))\\
\boxed{\text{by the induction, observing } \phi_2(s) = \phi(s)} \quad &= \underbrace{\phi(\phi(\ldots(\phi(s))\ldots))}_{n-1 \text{ times}}
\end{aligned}
$$

$$= \phi^{n-1}(s)\,.$$

Moreover

$$\begin{aligned} p_{\text{ext}} &= \mathbf{P}\left(\bigcup_{n\geq 1}\{\,X_n = 0\,\}\right) \\ \boxed{\text{increasing events}} &= \lim_{n\to\infty} \mathbf{P}\{\,X_n = 0\,\} \\ &= \lim_{n\to\infty} \phi_{n-1}(0) \\ &= \lim_{n\to\infty} \phi_n(0)\,. \end{aligned}$$

Consequently

$$\begin{aligned} p_{\text{ext}} &= \lim_{n\to\infty} \phi_n(0) \\ \boxed{\text{continuity of } \phi(\cdot)} &= \phi(\lim_{n\to\infty} \phi_{n-1}(0)) \\ &= \phi(p_{\text{ext}})\,. \end{aligned}$$

In order to see that p_{ext} is *the smallest* solution of (2.1.3) in $[0,1]$ assume that for some η, with $0 \leq \eta \leq 1$, $\eta = \phi(\eta)$. We have

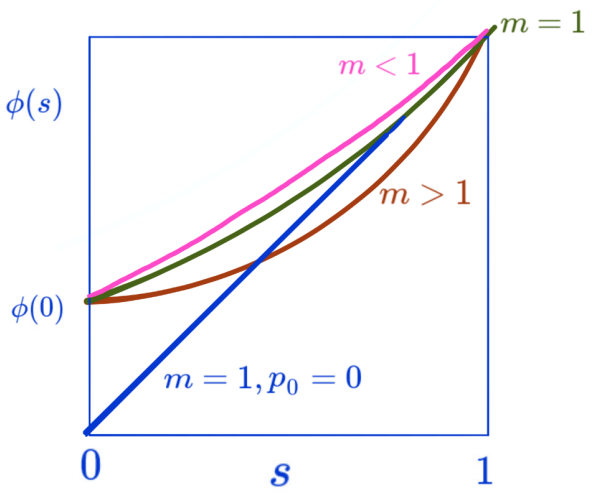
$$\begin{aligned} 0 &\leq \eta \\ \boxed{\phi(\cdot) \text{ is increasing}} \qquad \phi(0) &\leq \phi(\eta) = \eta \\ \phi^n(0) &\leq \eta \\ p_{\text{ext}} = \lim_{n\to\infty} \phi^n(0) &\leq \eta\,. \end{aligned}$$

□

Recall $m = \mathbf{E}[N]$ expected offspring (number of children of a given node).

Corollary 2.1.2.
If $m < 1$ then $p_{\text{ext}} = 1$ (sub-critical case).
If $m > 1$ then $p_{\text{ext}} < 1$ (super-critical case).
If $m = 1$ (critical case) then
if $p_0 = 0$ (hence $p_1 = 1$) then $p_{\text{ext}} = 0$
if $p_0 > 0$ then $p_{\text{ext}} = 1$.

Proof. Recall $\phi(0) = p_0$, $\phi(1) = 1$ and $\phi(\cdot)$ is convex.
Note also that $m = \mathbf{E}[N] = \frac{d}{ds}\phi(s)|_{s=1} = \phi'(1)$.


$m = 1$
$m < 1$
$\phi(s)$
$m > 1$
$\phi(0)$
$m = 1, p_0 = 0$
0
s
1

□

2.2 Conditional distribution under extinction

Consider a super-critical GW tree with positive extinction probability. We shall see that the conditional distribution of GW given extinction coincides with the distribution of some sub-critical GW tree, called the *dual* one.

2.2.1 Duality of the Galton-Watson trees

Let $\{p_k\}$ be an offspring distribution generating a super-critical GW tree of non-null extinction probability. Hence, throughout this section $p_0 > 0$ and $m = \sum_{k=0}^{\infty} k p_k > 1$.

Lemma 2.2.1. *There exists a unique $s_0 \in (0,1)$ such that*

$$\sum_{k=0}^{\infty} k p_k s_0^k = \phi(s_0)\,. \tag{2.2.1}$$

Proof. Consider

$$f(x) = \sum_{k=0}^{\infty} k p_k x^k - \phi(x) = -p_0 + \sum_{k=1}^{\infty} (k-1) p_k x^k\,.$$

Note $f(0) = -p_0 < 0$ and $f(1) = m - 1 > 0$. Moreover $f(\cdot)$ is increasing and continuous. □

Define a family of "tilted" distributions $\{p_k(\lambda)\}$, parametrized by $\lambda > 0$

$$p_k(\lambda) := p_k \frac{(s_0\lambda)^k}{\phi(s_0\lambda)}\,. \tag{2.2.2}$$

Note that $\sum_{k=0}^{\infty} p_k(\lambda) = 1$, hence $\{p_k(\lambda)\}$ are indeed (offspring) distributions.

Observe the following properties: $\{p_k(1)\}$ generates a critical GW tree.
$\{p_k(\lambda)\}$ with any $\lambda > 1$ generates a super-critical GW tree,
$\{p_k(\lambda)\}$ with any $\lambda < 1$ generates a sub-critical GW tree.
Indeed, observe that the function $f(x)$ defined in the proof of Lemma 2.2.1 is *strictly increasing* with $f(1) = 0$. Obviously, $\{p_k(1/s_0)\} = \{p_k\}$ is the original distribution.

Denote by $p_{\text{ext}}(\lambda)$ the extinction probability of the GW tree with offspring distribution $\{p_k(\lambda)\}$.

Lemma 2.2.2. *For any $\lambda > 1$ there exists a unique $\mu > 0$, $\mu \neq \lambda$ such that*

$$\frac{\phi(s_0\lambda)}{\lambda} = \frac{\phi(s_0\mu)}{\mu}\,. \tag{2.2.3}$$

Moreover, $\mu \in (0,1)$ and

$$\mu = \lambda p_{\text{ext}}(\lambda)\,. \tag{2.2.4}$$

The GW trees with the offspring distributions $\{p_k(\lambda)\}$ and $\{p_k(\mu)\}$, where λ and μ are related by (2.2.3) and (2.2.4), are called *dual* to each other.

Remark 2.2.3. In particular, the original super-critical distribution $\{p_k\}$ with $p_0 > 0$ has as its dual sub-critical distribution $\{\tilde{p}_k\}$ with

$$\tilde{p}_k = p_{\text{ext}}^{k-1} p_k \qquad k = 0, 1, \dots . \tag{2.2.5}$$

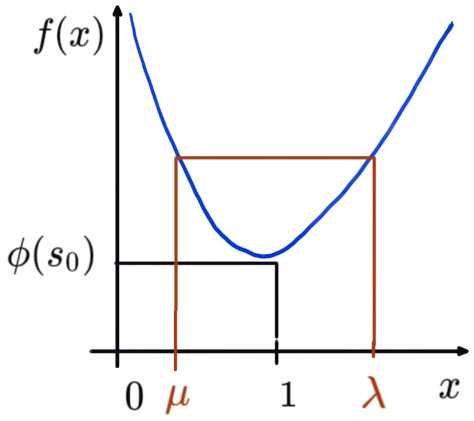
Indeed, $\{\tilde{p}_k\}$ corresponds to $\{p_k(\lambda)\}$ with $\lambda = 1/s_0$. By (2.2.3) and (2.2.4) the dual parameter μ satisfies then $s_0 = \phi(s_0\mu)/\mu$ and $s_0\mu = p_{\text{ext}}$ giving $\mu = \phi(p_{\text{ext}})/s_0 = p_{\text{ext}}/s_0$, where the last equality follows from the caricaturisation of p_{ext} given in Theorem 2.1.1. Consequently, by (2.2.2), $p_k(\mu) = p_k p_{\text{ext}}^k/\phi(p_{\text{ext}}) = p_k p_{\text{ext}}^{k-1}$.

Proof of Lemma 2.2.2. Denote

$$f(x) = \frac{\phi(s_0 x)}{x} = \frac{p_0}{x} + \sum_{k=1}^{\infty} p_k s_0^k x^{k-1} .$$

Note that $\lim_{x\downarrow 0} f(x) = \lim_{x\to\infty} f(x) = \infty$. Also, note $f(x)$ is strictly convex for $x > 0$ (as a sum of convex functions with p_0/x being strictly convex). Moreover

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x)|_{x=1} = \sum_{k=0}^{\infty} (k-1) p_k s_0^k = 0 \quad \text{by (2.2.1)}.$$



This proves the first statement. For the second statement, note $\mu/\lambda < 1$ and by Theorem 2.1.1, it is enough to prove that μ/λ solves the equation $s = \phi_\lambda(s)$, where $\phi_\lambda(\cdot)$ is the pgf of the distribution $\{p_k(\lambda)\}$. Indeed,

$$\phi_\lambda(\mu/\lambda) = \sum_{k=0}^{\infty} p_k \frac{(s_0\lambda)^k}{\phi(s_0\lambda)} \left(\frac{\mu}{\lambda}\right)^k$$

$$= \sum_{k=0}^{\infty} p_k \frac{(s_0\mu)^k}{\phi(s_0\lambda)}$$

$$= \frac{\phi(s_0\mu)}{\phi(s_0\lambda)}$$

$$\boxed{\text{by (2.2.3)}} = \frac{\mu}{\lambda}\,.$$

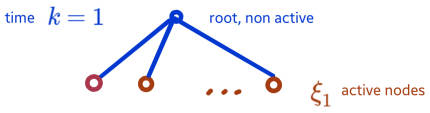
□

2.2.2 Exploration of the Galton-Watson tree

Starting from the root, in successive time instances $k = 0, 1, \ldots$ we shall be "discovering" children of different nodes of the GW tree. This *exploration* of the GW tree is a useful technique that will be later used to study connected components of more complicated random graphs.

- Time $k = 0$: There is $A_0 = 1$ *active* node, namely the root.

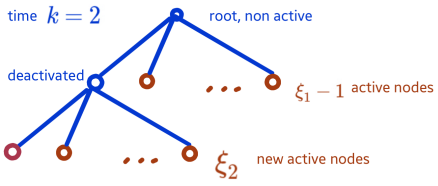
- Time $k = 1$: We pick the root (the only active node), we *deactivate it*, and we *discover all* $\xi_1 := N_{(1)}$ *children of this node, making them active.*



After the step at time 1, we have

$$A_1 = A_0 + \xi_1 - 1 = \xi_1 \qquad \text{active nodes}\,.$$

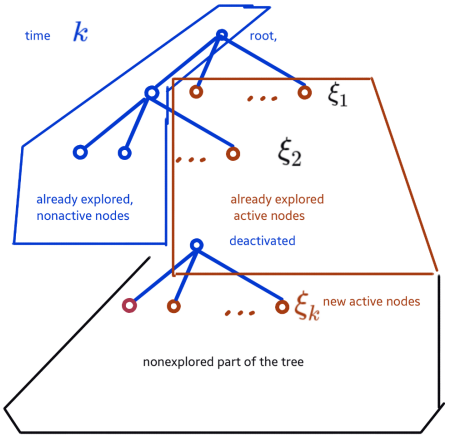
- Possible situation after step $k = 2$



$$A_2 = A_0 + \xi_1 + \xi_2 - 2 \qquad \text{active nodes}\,.$$

- In general, at time k we pick some active node σ [2], we deactivate σ and discover and activate all children of this node. We denote by $\xi_k := N_\sigma$ the number of newly activated nodes.

[2] The way of choosing an active node is irrelevant. We can fix some particular policy related to tree node labeling $\sigma \in \mathbb{S}$ (cf. Section 2.1.1), For example smaller $|\sigma|$ first and in lexicographic order among σ of the same length.



This makes the number of active nodes

$$A_k = A_{k-1} + \xi_k - 1 = \sum_{i=1}^{k} \xi_i - (k-1)\,. \tag{2.2.6}$$

Note ξ_i in (2.2.6) are i.i.d. and have the offspring distribution of the GW tree.

We continue the above exploration as long as there are active nodes. This can be a finite or infinite procedure, depending whether the GW is finite or infinite. In this regard, denote

$$T := \min\{k \geq 1 : A_k = 0\}\,, \tag{2.2.7}$$

with the convention $\min \emptyset = \infty$. Note

$$T < \infty \quad \text{iff extinction holds for GW}\,.$$

Given $T < \infty$, $A_T = 0$ and by (2.2.6)

$$T = \sum_{k=1}^{T} \xi_k + 1\,. \tag{2.2.8}$$

The above equation remains valid also for $T = \infty$. Recall, ξ_k are i.i.d., have the offspring distribution. Obviously T is *not* independent of $\{\xi_k\}$.

Remark 2.2.4. Note $T = S$, the total size of the GW tree. This can be deduced from the fact that at each step of the exploration procedure one deactivates one node, and at time $T < \infty$ all nodes are explored and deactivated. The equation remains trivially true also in case $T = S = \infty$.

2.2.3 History of the Galton-Watson tree

Denote

$$\mathcal{H} := (\xi_1, \ldots, \xi_T) \qquad \textit{history of the GW tree}$$

with $\{\xi_k\}$ and T as in (2.2.6), (3.1.6) and (2.2.8). Note that, given the exploration policy (cf. Footnote 4), there is one-to-one mapping between $\mathcal{H}$ and the GW tree (2.1.1).
A vector $(x_1, \dots, x_k)$, with $x_i \geq 0$, $1 \leq k \leq \infty$ is a *(valid) realization of the history* $\mathcal{H}$ iff

$$\sum_{i=1}^{l} x_i + 1 > l \quad \forall l = 1, \dots, k-1 \qquad \text{and} \qquad \sum_{i=1}^{k} x_i + 1 = k\,. \tag{2.2.9}$$

Also, by the independence of $\{\xi_k\}$, for any valid realization $(x_1, \dots, x_k)$ of the history $\mathcal{H}$

$$\mathbf{P}\{\,\mathcal{H} = (x_1, \dots, x_k)\,\} = \prod_{i=1}^{k} p_{x_i}\,. \tag{2.2.10}$$

Theorem 2.2.5. *Let $\mathcal{H}_\lambda$ by the history of the GW tree with offspring distribution $\{p_k(\lambda)\}$ given by (2.2.2) with $\lambda > 1$, with $\sum_{k=1}^{\infty} k p_k(\lambda) > 1$ and $p_0(\lambda) > 0$. Denote by T_λ the end-time of its exploration process. Then the conditional law of $\mathcal{H}_\lambda$ given extinction*

$$\mathcal{L}(\mathcal{H}_\lambda | T_\lambda < \infty) = \mathcal{L}(\mathcal{H}_\mu)$$

where $\mathcal{H}_\mu$ is the history of a GW tree with the dual offspring distribution $\{p_k(\mu)\}$, where λ and μ are related by (2.2.3) and (2.2.4).

Proof. Let $(x_1, \dots, x_k)$ with $1 \leq k < \infty$ be a valid, finite realization of the history of the extincted GW tree; cf (2.2.9). Then

$$\begin{aligned}
\mathbf{P}\{\,\mathcal{H}_\lambda = (x_1, \dots, x_k)\,|\,\text{extinction}\,\} &= \frac{1}{p_{\text{ext}}(\lambda)} \prod_{i=1}^{k} p_{x_i}(\lambda) \\
&= \frac{1}{p_{\text{ext}}(\lambda)} \prod_{i=1}^{k} p_{x_i} \frac{(s_0\lambda)^{x_i}}{\phi(s_0\lambda)} \\
&= \frac{1}{p_{\text{ext}}(\lambda)} \frac{(s_0\lambda)^{\sum_{i=1}^{k} x_i}}{(\phi(s_0\lambda))^k} \prod_{i=1}^{k} p_{x_i} \\
\text{by (2.2.9)} \quad &= \frac{1}{p_{\text{ext}}(\lambda)} \frac{(s_0\lambda)^{k-1}}{(\phi(s_0\lambda))^k} \prod_{i=1}^{k} p_{x_i} \\
&= \frac{1}{\lambda p_{\text{ext}}(\lambda)} \frac{(s_0\lambda)^{k}}{(\phi(s_0\lambda))^k} \prod_{i=1}^{k} p_{x_i} \\
\text{by (2.2.3)} \quad &= \frac{1}{\lambda p_{\text{ext}}(\lambda)} \frac{(s_0\mu)^{k}}{(\phi(s_0\mu))^k} \prod_{i=1}^{k} p_{x_i} \\
\text{by (2.2.4)} \quad &= \frac{(s_0\mu)^{k-1}}{(\phi(s_0\mu))^k} \prod_{i=1}^{k} p_{x_i}
\end{aligned}$$

$$\boxed{\cdots} = \prod_{i=1}^{k} p_{x_i}(\mu)$$
$$= \mathbf{P}\{\, \mathcal{H}_\mu = (x_1, \ldots, x_k) \,\} \ .$$

□

Remark 2.2.6. Following Remark 2.2.3, the conditional distribution of a super-critical GW tree with the offspring distribution $\{p_k\}$ satisfying $p_0 > 0$ (hence $p_{\text{ext}} > 0$), given its extinction, is equal to the distribution of the sub-critical GW tree having the offspring distribution $\{\tilde{p}_k\} = \{p_{\text{ext}}^{k-1} p_k\}$.

2.3 Total population bound

We shall see that the total number of individuals in the sub-critical GW tree has an exponential tail, provided the offspring distribution has finite exponential moments. By Theorem 2.2.5, the result remains true for the super-critical GW tree conditioned on extinction.

For a given random variable X (its distribution) consider the function [3]

$$h_X(x) := \sup_{\theta>0}\{\theta x - \log \mathbf{E}\left[e^{\theta X}\right]\} \quad x \in \mathbb{R}\,. \tag{2.3.1}$$

We shall need the following Lemma.

Lemma 2.3.1 (Chernoff bound)**.** *For X_i, $i = 1, \dots, n$ i.i.d. random variables, $X_i \stackrel{\mathcal{L}}{=} X$, and $a \in \mathbb{R}$*

$$\mathbf{P}\left\{\sum_{i=1}^{n} X_i \geq na\right\} \leq e^{-nh_X(a)}\,, \tag{2.3.2}$$

where $h_X(\cdot)$ is the rate function (2.3.1) of X. Moreover, if $\mathbf{E}\left[e^{\theta X}\right] < \infty$ for some $\theta > 0$ then $h_X(a) > 0$ for $a > \mathbf{E}[X]$.

Proof. For any $\theta > 0$

$$\begin{aligned}
\mathbf{P}\left\{\sum_{i=1}^{n} X_i \geq na\right\} &= \mathbf{P}\left\{\theta\sum_{i=1}^{n} X_i \geq \theta na\right\} \\
&= \mathbf{P}\left\{e^{\theta\sum_{i=1}^{n} X_i} \geq e^{\theta na}\right\} \\
&= \mathbf{P}\left\{e^{\theta\sum_{i=1}^{n} X_i - \theta na} \geq 1\right\} \\
\text{Markov inequality} \quad &\leq e^{-\theta na}\mathbf{E}\left[e^{\theta\sum_{i=1}^{n} X_i}\right] \\
\text{independence} \quad &= e^{-\theta na}\left(\mathbf{E}\left[e^{\theta X}\right]\right)^n \\
&= e^{-n(\theta a - \log \mathbf{E}\left[e^{\theta X}\right])}\,.
\end{aligned}$$

Since the chain of inequalities is valid for all $\theta > 0$ we can take $\inf_{\theta>0}$ on the right hand side

$$\begin{aligned}
\mathbf{P}\left\{\sum_{i=1}^{n} X_i \geq na\right\} &\leq \inf_{\theta>0} e^{-n(\theta a - \log \mathbf{E}\left[e^{\theta X}\right])} \\
&= e^{-n\sup_{\theta>0}(\theta a - \log \mathbf{E}\left[e^{\theta X}\right])} \\
&= e^{-nh_X(a)}\,.
\end{aligned}$$

[3] Called *rate* or *Cramer* function.

For the second statement, observe that if $\mathbf{E}\left[e^{\theta X}\right] < \infty$ for some $\theta > 0$ then $\log \mathbf{E}\left[e^{\theta X}\right] = \theta \mathbf{E}[X] + o(\theta)$, when $\theta \to 0$. Consequently, for $x > \mathbf{E}[X]$ and sufficiently small $\theta > 0$, $\theta x - \log \mathbf{E}\left[e^{\theta X}\right] > 0$. □

Theorem 2.3.2. *Consider a sub-critical GW tree with the offspring variable N, satisfying $\phi(s) < \infty$ for some $s > 1$. We have the following bound on the tail of the distribution of the total size S of the GW tree.*

$$\mathbf{P}\{\, S > k \,\} \le e^{-k h_N(1)} \,. \tag{2.3.3}$$

where $h_N(\cdot)$ is given by (2.3.1), and $h_N(1) > 0$.

Proof. Consider the numbers of active nodes A_k (2.2.6) in the exploration of GW, the completion time T (3.1.6) and the fact that $S = T$ (cf. Remark 2.2.4)

$$\begin{aligned}
\mathbf{P}\{\, S > k \,\} &= \mathbf{P}\{\, T > k \,\} \\
&= \mathbf{P}\{\, A_1 > 0, \ldots, A_k > 0 \,\} \\
&\le \mathbf{P}\{\, A_k > 0 \,\} \\
\boxed{(2.2.6)} \quad &= \mathbf{P}\left\{ \sum_{i=1}^{k} \xi_i - (k-1) > 0 \right\} \\
&= \mathbf{P}\left\{ \sum_{i=1}^{k} \xi_i \ge k \right\} \\
\boxed{\text{Lemma 2.3.1}} \quad &\le e^{-k h(1)} \,.
\end{aligned}$$

□

2.4 Exercises

1. Let M be total number of generations in the GW tree. Show that $\mathbf{P}\{\, M \le m \,\} = \phi^m(0)$ for $m = 1, 2, \ldots$.

2. Denote by $\phi_S(\cdot)$ the pgf of the total number of nodes in the GW tree; $\phi_S(t) := \mathbf{E}\left[t^S\right]$. Show that $\phi_S(\cdot)$ satisfies the equation

$$\phi_S(t) = t\phi(\phi_S(t)) \qquad s \in [0, 1].$$

3. Prove that when $\{p_k\}$ is Poisson, with parameter (mean) m i.e., $p_k = e^{-m}m^k/k!$, $k = 1, 2, \ldots$, then $\{p_k(\lambda)\}$ is Poisson distribution with parameter λ. Moreover, Poisson distributions $\{p_k(\lambda)\}$ and $\{p_k(\lambda)\}$ are dual to each other in the sense of Lemma 2.2.2 iff

$$\lambda e^{-\lambda} = \mu e^{-\mu}\,.$$

4. Prove that for X_i, $i = 1, \ldots, n$ independent (not necessarily identically distributed) random $\{0, 1\}$-random variables, with $X := \sum_{i=1}^n X_i$, $\bar{X} := \mathbf{E}[X]$, we have for any $\epsilon > 0$

$$\mathbf{P}\{\, X - \bar{X} \ge \epsilon\bar{X} \,\} \le e^{-\bar{X}h(\epsilon)}, \qquad \mathbf{P}\{\, X - \bar{X} \le -\epsilon\bar{X} \,\} \le e^{-\bar{X}h(-\epsilon)}\,, \tag{2.4.1}$$

where $h(x) := (1 + x)\log(1 + x) - x$.

5. *Computer exercise.* Consider the Galton-Watson tree presented on Figure 2.1, with the offspring distribution $p_k = 1/8, 3/8, 3/8, 1/8$ for $k = 0, 1, 2, 3$.

   (a) Calculate its mean offspring and conclude it is super-critical.

   (b) Calculate numerically the probability of extinction p_{ext}. Observe how $\phi^n(0)$ converges to p_{ext} when $n \to \infty$.

   (c) Calculate numerically the probabilities that the tree will have no more than $1, 2, 3, \ldots, 10$ generations. Hint: use Exercise 1.

   (d) Calculate numerically the conditional probabilities that the tree will have no more than $1, 2, 3, \ldots, 10$ generations given its extinction. (Hint: calculate the distribution of the dual sub-critical tree.) Compare to the unconditional probabilities. Do experiments with the distribution of p_k tilted by different values of λ as in (2.2.2), increasing or decreasing the expected offspring.

Lesson 3

Erdős-Rényi graph — emergence of the giant component

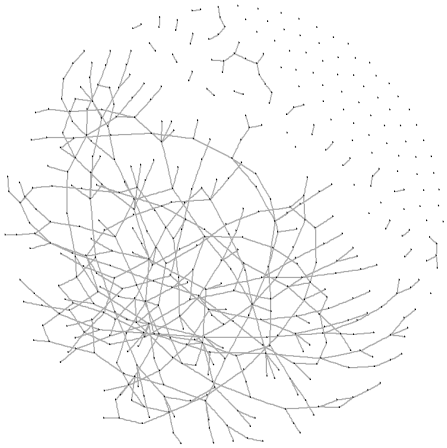


Figure 3.1: A realization of the Erdős-Rényi graph on $n = 500$ vertexes with edge probability $p = 0.0035$ (mean vertex degree $np \approx 1.751 > 1$). For transparency, vertexes which are connected by an edge are placed on this figure close to each other, however that any two vertexes have the same probability of being connected by an edge. Figure borrowed from Laszlo Gadar.

The Erdős-Rényi graph is named after two Hungarian mathematicians who set the foundations of modern random graph theory: Paul Erdős, one of the most prolific mathematicians of the 20th century[1], and Alfréd Rényi, who introduced a version of this models in their joint

[1] Erdős had over 500 collaborators around the world. The "collaborative distance" between him and another person, as measured by authorship of mathematical papers, is known as the Erdős number. The American Mathematical Society provides a free online tool to determine this number for every mathematical author listed in the Mathematical Reviews catalogue; see http://www.ams.org/mathscinet/collaborationDistance.html.

paper Erds and Rényi ((1959)) in 1959.

As is customary, in what follows we shall consider however a slightly different variant of this model, more simple to study, introduced independently in 1959 in Gilbert ((1959)) by Edgar Nelson Gilbert, an American mathematicians with wide interest in communication theory.[2] In the context of our first lesson, on the bond percolation, this model can be seen as an *independent bond percolation on the complete graph* with finite total number of vertexes. When this number goes to infinity and the probability of the edge to be open goes to zero in some appropriate way, one observes various phase transitions. In the present lesson we consider the phase transitions related to the emergence, of the *giant component*. In a separate lesson we shall study the emergence of the full connectivity.

Our exposition in this lesson corresponds to the results presented in ((Draief and Massoulié, 2010, Chapter 2)); see also ((Van Der Hofstad, 2017, Chapter 4)).

An inhomogeneous generalization of the classical model (presented here), which allows one for different edge probabilities between different groups of nodes (cf ((Van Der Hofstad, 2017, Chapter 6))), has become recently popular under the name of *stochastic block model* and used to study the problem of community detection in social networks, cf Holland et al. ((1983)).

3.1 Erdős-Rényi graphs

3.1.1 Gilbert's variant of the Erdős-Rényi graph

The Gilbert's variant of the Erdős-Rényi graph is a graph with n vertexes and edges creted independently, with probability p_n, between any pair of vertexes. Depending on the choice of p_n different macroscopic phenomena appear when $n \to \infty$.

Formally, let $\{\delta_{\{u,v\}} : u, v \in \mathbb{N}, u \neq v\}$ be i.i.d. Bernoulli r.v. with $\mathbf{P}\{\, \delta_{\{u,v\}} = 1 \,\} = 1 - \mathbf{P}\{\, \delta_{\{u,v\}} = 0 \,\} = p_n$.

ER graph is defined as $\mathrm{ER} := \mathrm{ER}_n = (V_n, E_n)$, where $V_n := \{1, \dots, n\}$ (graph vertexes) and

$$E_n := \big\{\{u, v\} \subset V_n : \delta_{\{u,v\}} = 1, u \neq v\big\} \qquad \text{(random edges)}\,.$$

Notation: for $u, v \in V_n$, we write $u \sim v$ if $\delta_{\{u,v\}} = 1$; i.e., there is an edges between u and v. In this case we say u, v are *neighbours*.

3.1.2 Sparse Erdős-Rényi graph

We consider here the case when

$$p_n := \frac{\lambda}{n} \qquad \text{for some } \lambda > 0; \tag{3.1.1}$$

λ is the (unique) parameter of the family $\mathrm{ER}_n(\lambda)$.

[2] He is probably best known for his contributions to coding theory. In this series of lessons we shall hear about him in the context of the Voronoi tessellation and continuum percolation, where his model, called today Gilbert graph, is a fundamental model of interest.

The choice (3.1.1) means that ER is (asymptotically) *sparse graph*; i.e., such that the number of its edges is small with respect to the number $n(n-1)/2$ of edges in the complete graph. Indeed,

$$\mathbf{E}\left[|E_n|\right] = \mathbf{E}\left[\sum_{\substack{u,v=1\\u<v}}^{n} \delta_{\{u,v\}}\right] = \frac{n(n-1)}{2} p_n = \frac{(n-1)}{2}\lambda\,.$$

3.1.3 Degree distribution of the typical vertex

Take some $v \in V_n$ (the choice should not depend on the realization of the edges). Denote by $D_v = |\{u \in V_n : v \sim m\}|$ the *number of neighbours* or *degree* of v in $\mathrm{ER}_n(\lambda)$. Note

$$\mathbf{E}\left[D_v\right] = (n-1)p_n = \lambda\frac{n-1}{n} \to \lambda \quad \text{as } n \to \infty\,;$$

i.e., the mean degree of a vertex of $\mathrm{ER}_n(\lambda)$ is asymptotically equal to λ. Moreover, its distribution

$$D_v \stackrel{\mathcal{L}}{=} \mathrm{Binomial}(n-1, p_n) \Longrightarrow \mathrm{Poisson}(\lambda) \quad \text{as } n \to \infty. \tag{3.1.2}$$

[3] Hence, the direct neighbouhood of any vertex of $\mathrm{ER}_n(\lambda)$ looks asymptotically (as $n \to \infty$) as this of the first generation of the GW tree with the offspring distribution $\mathrm{Poisson}(\lambda)$.

Questions:
Can we go further (deeper in generations) with the GW approximation of ER? (Yes)
Can we derive, or at least guess, some macroscopic properties of ER using this approximation? (Yes, but ...)

3.1.4 Exploration of node's neighbourhood

In what follows we shall be using basically the exploration technique already applied to Galton-Watson tree. Let $v \in V_n$ be any vertex (chosen whiteout any dependence on the realization of the edges).

Starting from v, in successive time instances $k = 0, 1, \ldots$ we shall be discovering vertexes and (some) edges in the neighbourhood of v.

- Time $k = 0$: $\mathcal{A}_0 := \{v\}$, the *active* node. $\mathcal{B}_0 := \emptyset$, the set of *inactive nodes* (already discovered but not active any more).
- Time $k = 1$: We pick $v_1 := v$ (the only active node), we *deactivate it*, and we *discover all neighbours* $\mathcal{D}_1 := \mathcal{D}(v_1)$ of v_1. At the end of the step 2, we have

$$\mathcal{A}_1 := \mathcal{A}_0 \cup \mathcal{D}_1 \setminus \{v_1\} = \mathcal{D}(v_1) \qquad \text{active nodes}\,,$$

$$\mathcal{B}_1 := \{v_1\} \qquad \text{(discovered and) inactive nodes.}$$

[3] recall, $\mathrm{Binomial}(n,p)$ is a distribution on $k = 0, 1, \ldots, n$ with probabilities $\binom{n}{k}p^k(1-p)^{n-k}$, while $\mathrm{Poisson}(\lambda)$ is a distribution on $k = 0, 1, \ldots,$ with probabilities $e^{-\lambda}\lambda^k/k!$ and (3.1.2) results from Poisson limit theorem.

- In general, at time k we pick some active node $v_k \in \mathcal{A}_{k-1}$ [4], we deactivate v_k and discover and activate all neighbours $\mathcal{D}_k := \mathcal{D}(v_k)$ of this v_k, *which are not in* $\mathcal{A}_{k-1} \cup \mathcal{B}_{k-1}$. This makes

$$\mathcal{A}_k := \mathcal{A}_{k-1} \cup \mathcal{D}_k \setminus \{v_k\} \qquad \text{active nodes}\,,$$

$$\mathcal{B}_k := \mathcal{B}_{k-1} \cup \{v_k\} \qquad \text{(discovered and) inactive nodes.}$$

Denote the cardinalities of the above sets by $A_k := |\mathcal{A}_k|$, $B_k := |\mathcal{B}_k|$ and $\xi_k := |\mathcal{D}_k|$. This makes

$$A_0 = 1,$$

$$B_0 = 0,$$

$$A_k = \sum_{i=1}^{k} \xi_k - k + 1, \tag{3.1.3}$$

$$B_k = k. \tag{3.1.4}$$

Note that, unlike in the exploration of the GW tree, variables ξ_k, $k = 1, \ldots$ are *not* independent *nor* identically distributed. They form a Markov chain with

$$\begin{aligned}\mathcal{L}(\xi_k | \xi_1, \ldots, \xi_{k-1}) &= \text{Bernoulli}(n - A_{k-1} - B_{k-1}, p_n) \\ &= \text{Bernoulli}(n - \sum_{i=1}^{k-1} \xi_i - 1, p_n)\,. \end{aligned} \tag{3.1.5}$$

We continue the above exploration as long as there are active nodes. Denote

$$T := \min\{k \geq 1 : A_k = 0\}\,, \tag{3.1.6}$$

with the convention $\min \emptyset = \infty$. Note $T \leq n$ since the graph has n nodes.

Remark 3.1.1. *Observe that at time T we have discovered all nodes in the connected component of v in* ER_n but not all edges. *In fact, we have discovered a* spanning tree *of this component in which only edges between v_k and its neighbours not in $\mathcal{A}_{k-1} \cup \mathcal{B}_{k-1}$ are revealed.*

3.1.5 Local Poisson tree-like structure

Denote

$$\mathcal{H} = \mathcal{H}_n = \mathcal{H}_n(v) := (\xi_1, \ldots, \xi_T) \qquad \textit{history of the exploration of the neighbourhood of } v\,.$$

[4] The way of choosing an active node is irrelevant. We can fix some particular policy related to tree node labeling $\sigma \in \mathbb{S}$ (cf. the lesson on Galton-Watson tree).

A vector $(x_1, \dots, x_k)$ is a possible realization of the *(partial) history* of the exploration up to k th generation of neighbours iff

$$\begin{cases} \sum_{i=1}^{j} x_i - j + 1 > 0 & \text{for all } 1 < j < k, \\ \sum_{i=1}^{k} x_i - k + 1 \geq 0 & \text{with the equality iff } k = T. \\ \sum_{i=1}^{k} x_i + 1 \leq n. \end{cases} \tag{3.1.7}$$

Fact 3.1.2. *Let $\mathcal{H}_n := (\xi_1, \dots, \xi_T)$ be the history of the exploration of the neighbourhood in* ER_n *with edge probability $p_n = \lambda/n$ and let $(x_1, \dots, x_k)$ be a possible partial realization of this history up to k th generation as in* (3.1.7)*. Then (for fixed k)*

$$\lim_{n\to\infty} \mathbf{P}\{\, \xi_1 = x_1, \dots, \xi_k = x_k \,\} = \prod_{i=1}^{k} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \,. \tag{3.1.8}$$

Proof. Let $(x_1, \dots, x_k)$ with $1 \leq k < \infty$ be a valid, partial realization of the history of the exploration (3.1.7). Denote $a_i := \sum_{j=1}^{i} x_j - i + 1$, $b_i = i$, $i = 1, \dots, k$. Note a_i, b_i are realizations of A_j and B_j in (3.1.3), (3.1.4), respectively, when $\xi_1 = x_1, \dots, \xi_k = x_k$. Then by (3.1.5)

$$\mathbf{P}\{\, \xi_1 = x_1, \dots, \xi_k = x_k \,\} = \prod_{i=1}^{k} \binom{n - a_{i-1} - b_{i-1}}{x_i} p_n^{x_i} (1 - p_n)^{n - a_{i-1} - b_{i-1} - x_i} \,.$$

Recall, $p_n = \lambda/n$, a_i and b_i do not vary with n and thus for all $i \leq k$

$$\lim_{n\to\infty} (1 - p_n)^{n - a_{i-1} - b_{i-1} - x_i} = \lim_{n\to\infty} (1 - \lambda/n)^n = e^{-\lambda} \,,$$

$$n^{x_i} p_n^{x_i} = \lambda^{x_i} \,,$$

$$\lim_{n\to\infty} n^{-x_i} \binom{n - a_{i-1} - b_{i-1}}{x_i} = \frac{1}{x_i!} \lim_{n\to\infty} \frac{\prod_{j=1}^{x_i} (n - a_{i-1} - b_{i-1} - j + 1)}{n^{x_i}} = \frac{1}{x_i!}$$

□

Remark 3.1.3. *Fact 3.1.2 says that the exploration process of* ER_n *has asymptotically, for large n, the same finite dimensional distributions as the exploration process of the GW tree with Poisson(λ) offspring distribution. Proving further that the number of undiscovered edges in the exploration of* ER_n *(cf Remark 3.1.1) tends to 0 when $n \to \infty$ (cf Exercise 1) one can conclude that* (sparse) ER graph is (asymptotically) locally tree-like, *with Poisson(λ) offspring distribution. The above statement can be formalized on the ground of the* local weak convergence *of random graphs, which will be introduced in the lesson on the Unimodular random graphs.*

3.2 Emergence of the giant component

3.2.1 Phase transition

Denote by C_1 and C_2 a largest and a second largest connected component of ER_n. Possible lack of uniqueness is not a problem as we are going to consider only the statements regarding the size (number of vertexes) $|C_1|$ and $|C_2|$ of these components.

Theorem 3.2.1. *Consider the ER graph* $\mathrm{ER}_n(\lambda)$ *with edge probability* $p_n = \lambda/n$.

1. (Sub-critical regime) *If* $\lambda < 1$ *then there exists a constant* $a = a(\lambda)$ *such that*

$$\mathbf{P}\{\, |C_1| \le a \log(n) \,\} \to 1 \quad \text{as } n \to \infty.$$

2. (Super-critical regime) *If* $\lambda > 1$ *then there exists a constant* $a' = a'(\lambda) > 0$ *such that for all* $\delta > 0$

$$\mathbf{P}\left\{ \left| \frac{|C_1|}{n} - (1 - p_{\mathrm{ext}}(\lambda)) \right| \le \delta,\ |C_2| \le a' \log(n) \right\} \to 1 \quad \text{as } n \to \infty,$$

where $p_{\mathrm{ext}}(\lambda)$ *is the extinction probability in GW tree with Poisson*(λ) *offspring distribution.*

3. (Critical regime) *If* $\lambda = 1$ *then there exists a constant* $\kappa > 0$ *such that for all* $\zeta > 0$

$$\mathbf{P}\Big\{ |C_1| \ge \zeta n^{2/3} \Big\} \le \frac{\kappa}{\zeta^2}\,.$$

We refer to ((Draief and Massoulié, 2010, Chapter 2)) for the detailed proof of Theorem 3.2.1.

In what follows we will show how the *statements regarding sub- and super-critical regimes can be concjectured from the local Poisson tree-like structure of the graph* observed in Remark 3.1.3. These observations inspire the proof idea.

3.2.2 Tree-approximation heuristic

Sub-critical regime

Let $\lambda < 1$. For a given $v \in V_n$ denote by $C(v)$ the connected component of v in ER_n. By the GW approximation of $C(v)$ (neglecting the fact that the approximation is valid only up to a finite generation of neighbours of v), for fixed $a > 0$

$$\mathbf{P}\{\, |C(v)| > a \log n \,\} \overset{\mathrm{GW}}{\approx} \mathbf{P}\{\, S > a \log n \,\}\,,$$

where S is the total size of the GW tree with Poisson(λ) offspring distribution. By the total population bound in the sub-critical GW tree (cf. Theorem 3.2 in the *Lesson on the GW tree*) we have

$$\mathbf{P}\{\, S > a \log n \,\} \le e^{-ah(1)\log n} = n^{-ah(1)}\,,$$

where

$$h(x) = h_N(x) = \sup_{\theta>0}\{\theta x - \log \mathbf{E}\left[e^{\theta N}\right]\} \tag{3.2.1}$$

is the Cramer (rate) function of associated with N having Poisson(λ) distribution. It is easy to see that in this Poisson case

$$h(x) = x\log(x/\lambda) - x + \lambda \tag{3.2.2}$$

(cf Exercise 2) and thus

$$\mathbf{P}\{\,|C(v)| > a\log n\,\} \overset{\text{GW}}{\lesssim} n^{-a(\lambda-1-\log\lambda)}\,,$$

with $\lambda - 1 - \log\lambda > 0$. Multiplying by n the right-hand size of the above expression we obtain bound on the probability that *there exists* a vertex v for which $|C(v)| > a\log n$

$$\mathbf{P}\{\,|C(v)| > a\log n \quad \text{for some } v \in V_n\,\} \overset{\text{GW}}{\lesssim} n^{1-a(\lambda-1-\log\lambda)}\,.$$

For $a > (\lambda - 1 - \log\lambda)^{-1}$ the right-hand side converges to 0.

Super-critical regime

Let $\lambda > 1$. Note that $|C_1|/n$ is the fraction of vertexes in the largest component and consequently can be interpreted as the probability that an arbitrarily selected node v (before generating the edges of the graph) belongs to C_1 (after the edges have been generated)

$$\frac{|C_1|}{n} = \mathbf{P}\{\,v \in C_1\,\}\,.$$

The probability that v is in a largest component C_1 is interpreted in GW (heuristic) approximation as the probability that the GW approximation of the connected component $C(v)$ of v percolates (does not extinct)

$$\mathbf{P}\{\,v \in C_1\,\} \overset{\text{GW}}{\simeq} \mathbf{P}\{\,S = \infty\,\} = 1 - p_{\text{ext}}(\lambda)\,.$$

This suggest that

$$\frac{|C_1|}{n} \overset{\text{GW}}{\simeq} 1 - p_{\text{ext}}(\lambda)\,.$$

Regarding all other components of ER_n (including the second-largest), $\text{ER}_n \setminus C_1$ can be approximated by an ER graph on $n' := n - |C_1|$ nodes and the same $p_n = \lambda/n$ edge probability. By the above approximation of $|C_1|/n$ we have n' is of order $n - n(1 - p_{\text{ext}}(\lambda)) = np_{\text{ext}}(\lambda)$ and consequently $n'p_n \approx \mu := \lambda p_{\text{ext}}(\lambda)$. Hence $\text{ER}_{n'}(\lambda)$ is approximately for large n equal in distribution to $\text{ER}_n(\mu)$ where $\mu < 1$ is the parameter of the dual (sub-critical) GW tree, with Poisson(μ) offspring distribution, cf Section 2.1 in the *Lesson on GW trees* and Exercise 2 there. The statement regarding C_2 follows now by the GW approximation for the sub-critical $\text{ER}_n(\mu)$.

3.2.3 Martingale bound for the critical case

Recall random variables ξ_i in the history of the exploration of the neighbourhood $\mathcal{H}$ have conditional distribution

$$\mathcal{L}(\xi_k|\xi_1,\ldots,\xi_{k-1}) = \text{Bernoulli}(n - \sum_{i=1}^{k-1}\xi_i - 1, p_n)\,.$$

They can be stochastically (and a.s. on some probability space [5]) upper-bounded by i.i.d. random variables ξ'_k with

$$\mathcal{L}(\xi'_k) = \text{Binomial}(n, \lambda/n)$$

We have

$$A'_k := \sum_{i=1}^{k}\xi' - k + 1$$

and in case $\lambda = 1$ the sequence A'_k is a *martingale*. Indeed

$$\mathbf{E}\left[A'_k|A'_1\ldots,A'_{k-1}\right] = A'_{k-1}\mathbf{E}\left[\xi' - 1\right] = A'_{k-1}\,.$$

Studying this martingale allows one to obtain bound on the size of any component in the critical regime.

[5] ξ is *stochastically smaller* than ξ' if $\mathbf{P}\{\,\xi > t\,\} \le \mathbf{P}\{\,\xi' > t\,\}$ for all t. By Strassen's theorem one can construct then a probability space on which $\xi \le \xi'$ almost surely. This is called a *coupling* of random variables.

3.3 Exercises

1. Prove that the expected number of undiscovered edges in the exploration of the neighbourhood of a node in ER_n graph up to k th generation (k fixed) converges to 0 when $n \to \infty$. Using Fact 3.1.2 conclude that the graph $\mathrm{ER}_n(\lambda)$ asymptotically, locally looks like the Galton-Watson tree with Poisson offspring distribution of parameter λ. The above statement can be formalized on the ground of the *local weak convergence* of random graphs, which will be introduced in the lesson on the Unimodular random graphs.

2. Show that the Cramer (rate) function (3.2.1) associated to Poisson(λ) distribution of N is equal to (3.2.2).

3. *Computer exercise.* Using *igraph*, `http://igraph.org/redirect.html`, which is a library collection for creating and manipulating graphs and analyzing networks. It is written in C and also exists as Python and R packages, with an interface for Mathematica. For example, the plot on Figure 3.1 was produced using the following *igraph* commands

```
g <- erdos.renyi.game(500, 0.0035, type = "gnp")
plot(g, vertex.label= NA, edge.arrow.size=0.02,vertex.size = 0.5)
```

   (a) Generate and plot several instances of the sub- and super-critical ER graphs on $n = 500$ nodes.

   (b) Generate and plot an instance of the ER graph on $n = 500$ nodes with $p_n = 0.0035$.

   (c) For the observed realization of the graph calculate the empirical mean node degree.

   (d) Plot the histogram of the degree distribution. Compare to Binomial(n, p_n) and Poisson(np_n) distribution.

   (e) Calculate the fraction of nodes in the biggest component and compare to the survival (non-extinction) probability of the Galton-Watson tree with Poisson(np_n) offspring distribution.

   (f) For every node of the graph calculate the depth of the maximal tree-like neighborhood. Plot the histogram. Hint: Use commands `ego` to obtain node neighborhood, `mst` to calculate its minimum spanning tree and `difference` to obtain the tree excess.

Lesson 4

Graphs with a given node degree distribution

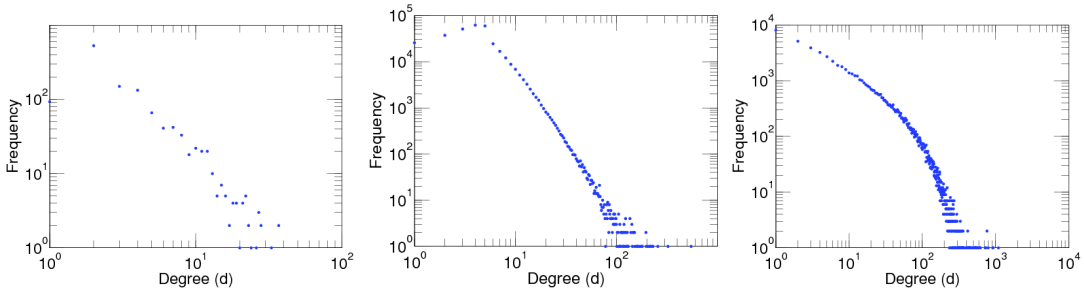


Figure 4.1: Examples of node degree distributions observed in three different real networks; from left to right: air traffic control, co-purchase of different items on Amazon, Facebook friendships [2]. A liner plot on the log-log scale indicates a power law, that is a probability function $p_k \sim x^{-\beta}$ for some $\beta > 1$ and large k.

The node degree in the Erdős-Rényi graph is asymptotically Poisson. This seriously limits the possibility to fit this model to really observed networks, where this distribution is not Poisson-like. In particular, in many examples of large networks this distribution resembles a power law; see Figure 4.1. These networks cannot be reasonable modeled by the Erdős-Rényi graph with the fitted mean node degree.

In this lesson we present a random graph model, whose node degree distribution can be chosen arbitrarily. This is an extension of the Erdős-Rényi graph in the sense that is also a sparse graph, with *local tree asymptotic* behaviour. However, the Galton-Watson tree, which describes the asymptotic structure of the neighbourhood of the typical (randomly selected)

[2] For the explanations and for more examples see *KONECT* (Koblenz Network Collection) database at http://konect.uni-koblenz.de/plots/degree_distribution.

node, is *inhomogeneous*: its root has a different offspring distribution than all other nodes. [3] The Configuration model exhibits also a similar phase transition, as the Erdős-Rényi graph, related to the emergence (or not) of the giant component, which can be conjectured from its local Galton-Watson tree asymptotic properties. In Section 4.2.2 we also present a new approach to this percolation problem via the law of large numbers for the exploration process.

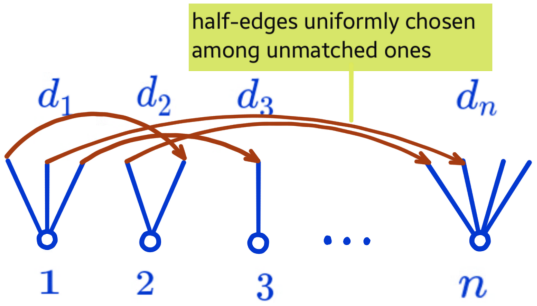
The Configuration model was introduced in 1980 by Béla Bollobás, yet another famous Hungarian mathematician, in his paper Bollobás ((1980)) to study *uniform random graphs with a given degree sequence*. We formulate and prove this result as well. The emergence of the giant component of this model was studied by Molloy and Reed in Molloy and Reed ((1995, 1998)). For further reading on the Configuration model we refer to ((Van Der Hofstad, 2017, Chapter 7)). and ((Van Der Hofstad, 2014, Chapter 4)).

4.1 Configuration model

4.1.1 Model construction

Consider a given sequence (in fact an array) $\mathbf{d} := (d_i^{(n)} : i = 1, \ldots, n)_{n=1}^{\infty}$, where $d_i^{(n)} \in \{0, 1, \ldots\}$. We call $\mathbf{d}$ *degree sequence*.

A *Configuration model with degree sequence* $\mathbf{d}$ (CM($\mathbf{d}$)) is a sequence of graphs $\mathrm{CM}_n = \mathrm{CM}_n(d_i^{(n)} : i = 1, \ldots, n)$, $n \geq 1$ with vertexes $V_n = \{1, \ldots, n\}$ and edges created between vertexes is such a way that $d_i^{(n)}$ is the degree of the vertex $i \in V_n$. The edges are created in a process of a *uniform matching of half-edges*: until there is no more unmatched half-edges, take one (in an arbitrary way) and match it to another one which is uniformly chosen among the remaining unmatched half-edges.



As a result of the uniform matching of half-edges one obtains a graph CM_n, that is possibly a *multi-graph*; i.e., there might exist

- multiple edges (more than one edge between two different nodes), 

[3] This phenomenon will be further considered in a more general context in the next lesson on the *unimodularity*.

- self-loops (edges coming back to the same node). 

Under appropriate conditions on $\mathbf{d}$, for large n, multiple edges and self-loops are not observed locally (adjacent to a given node). There are also ways to transform a CM into a simple graph (not having multiple edges and self-loops at all).

4.1.2 Conditions on the degree sequence

We assume the following conditions are satisfied regarding $\mathbf{d}$:

(0) *Feasibility of matching:* For each n the total number of half-edges

$$\ell_n := \sum_{i=1}^{n} d_i^{(n)} \qquad \text{is an even number,}$$

so as to be able to match all half-edges.

(i) *Consistency for large n.* The empirical distribution of the node degree converges to some distribution: for all $k = 0, 1, \ldots$

$$\lim_{n\to\infty} \frac{n_k}{n} = p_k \qquad \text{for some } p_k \geq 0,$$

where

$$n_k := \#\{i \in V_n : d_i^{(n)} = k\}$$

is the number of nodes of degree k in the graph of size n.

(ii) Further technical assumptions. The mean asymptotic degree is non-null and finite:

$$\lambda := \sum_{k=1}^{\infty} k p_k \in (0, \infty).$$

(iii)

$$\sum_{i=1}^{n} (d_i^{(n)})^2 = O(n)\,;$$

implies in particular a bound on the growth of the maximal degree; cf Exercise 1. It implies also that the average node degree converges to λ; cf Section 4.1.3.

(iv) The graph has leafs (nodes with degree 1) asymptotically with positive probability

$$p_1 > 0\,,$$

which prevents some special (strange?) phenomena from happening.

4.1.3 Degree distribution

For any n, let D_n be a random variable with distribution $\{n_k/n\}$; i.e., $\mathbf{P}\{\, D_n = k \,\} = n_k/n$, $k = 0, 1, \ldots$ Node that this is the distribution of the degree of the node uniformly selected in V_n. We call D_n (its distribution) the *degree distribution* of CM_n. The sequence of distributions of D_n, $n \geq 1$, characterizes $\mathbf{d}$ and hence $\mathrm{CM}(\mathbf{d})$.

Let D be a random variable with the distribution $\{p_k\}$; i.e., $\mathbf{P}\{\, D = k \,\} = p_k$, $k = 0, 1, \ldots$ We call D *(asymptotic) degree distribution*. The distribution of D characterizes many (but not all [4]) asymptotic properties of $\mathrm{CM}(\mathbf{d})$.

Observe:

- Condition (i) of Section 4.1.2 is equivalent to $D_n \Rightarrow D$, as $n \to \infty$.
- $\lambda = \mathbf{E}\,[D]$ and condition (ii) is equivalent to $0 < \lambda < \infty$.
- Condition (iii) implies $\mathbf{E}\,[D_n] \to \mathbf{E}\,[D] = \lambda$; cf Exercise 2.

Example 4.1.1 (CM with iid degrees)**.** Let $\{p_k\}$ be a given distribution satisfying $0 < \sum_{k=0}^{\infty} k p_k < \infty$ and $\sum_{k=0}^{\infty} k^2 p_k < \infty$. Let $\xi_1, \xi_2, \ldots$ be iid rv with distribution $\{p_k\}$. For all $n \geq 1$, define

$$d_i^{(n)} = \xi_i \qquad \text{for } i = 1, \ldots, n-1,$$

$$d_n^{(n)} = \begin{cases} \xi_n & \text{if } \sum_{i=1}^n \xi_i \text{ is even} \\ \xi_n + 1 & \text{otherwise.} \end{cases}$$

Then $\mathbf{d} = (d_i^{(n)})$ satisfies conditions (0)–(iii) of Section 4.1.2.

4.1.4 Distribution of the configuration model

CM might be a multi-graph. Let us represent a given realization of CM_n by a (symmetric) matrix $(x_{ij} : i, j \in V_n)$, where $x_{ij} = x_{ji}$ denotes the number of edges between i and j. In particular, x_{ii} denotes the number of self-loops at i. Observe that

$$d_i^{(n)} = x_{ii} + \sum_{j=1}^{n} x_{ij}\,. \tag{4.1.1}$$

Note that (x_{ij}) corresponds to a simple graph (no multiple edges and self-loops) iff $x_{i,j} \in \{0, 1\}$ and $x_{ii} = 0$ for all $i, j = 1, \ldots, n$.

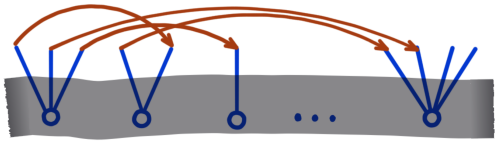
Proposition 4.1.2. *Let G be a multi-graph with n vertexes corresponding to $(x_{ij} : i, j \in V_n)$ satisfying (4.1.1). Then*

$$\mathbf{P}\Big\{\, \mathrm{CM}_n(d_i^{(n)}) = G \,\Big\} = \frac{\prod_{i=1}^n d_i^{(n)}!}{(\ell_n - 1)!! \prod_{i=1}^n 2^{x_{ii}} \prod_{1 \leq i \leq j \leq n} x_{ij}!}\,, \tag{4.1.2}$$

[4] For example $\mathbf{P}\{\, \mathrm{CM}_n \text{ is a fully connected graph} \,\} \to 1$ when $n \to \infty$ when $d_i^{(n)} \geq 3$ for all $i = 1, \ldots, n$ and n large enough, but not necessarily when $p_0 = p_1 = p_2 = 0$; cf. ((Van Der Hofstad, 2014, Theorem 10.14)).

where $\ell_n = \sum_{i=1}^n d_i^{(n)}$ *and* $a!! = a(a-2)\dots 3\cdot 1$ *for odd* a.

Proof. Forget vertexes and consider ℓ_n half-edges. Call any matching of them in pairs (giving $\ell_n/2$ pairs) a configuration.



There are $(\ell_n - 1)!!$ possible configurations. Each of these configurations is equally-likely as a result of the uniform matching of half-edges. Restore vertexes and observe

$$\mathbf{P}\Big\{ \mathrm{CM}_n(d_i^{(n)}) = G \Big\} = \frac{1}{(\ell_n - 1)!!} N(G)\,,$$

where $N(G)$ is the number of configurations which give the muliti-graph G. Obviously, several different configurations may give the same graph G. In fact

$$G(N) = \frac{\prod_{i=1}^n d_i^{(n)}!}{\prod_{i=1}^n 2^{x_{ii}} \prod_{1\le i\le j\le n} x_{ij}!}\,.$$

The factor $\prod_{i=1}^n d_i^{(n)}!$ in the numerator steams form the fact that any permutation of the half-edges at any of the vertexes leads to the same graph G. This number needs to be however adjusted when the graph G is not simple. The factor $\prod_{1\le i\le j\le n} x_{ij}!$ in the denominator compensates that fact that we should not count twice the permutations of x_{ij} of half-edges that contribute to the multiple edges between i and j when. And the factor $\prod_{i=1}^n 2^{x_{ii}}$ adjusts the calculation of configurations with self-loops. □

Corollary 4.1.3. *For a simple graph* G *we have* $x_{ii} = 0$ *for all* $i \in V_n$ *and* $x_{ij} \in \{0,1\}$, $i,j \in V_n$, $i \ne j$. *Hence*

$$\mathbf{P}\Big\{ \mathrm{CM}_n(d_i^{(n)}) = G \Big\} = \frac{\prod_{i=1}^n d_i^{(n)}!}{(\ell_n - 1)!!}\,. \tag{4.1.3}$$

This expression dosed not depend on G *(corresponding to* $(x_{ij} : i,j)$*) but only on* $d_i^{(n)}, i = 1,\dots,n$. *Hence the* conditional distribution of CM_n given it is simple is a uniform distribution *on all simple graphs with the given degree sequence.*

Remark 4.1.4. Corollary 4.1.3 says that the one can simulate a graph uniformly distributed on all simple graphs with a given degree sequence via the following *rejection sampling*: repeat the construction of the configuration model until you observe (say for the fist time) a simple graph; this is a realization sampled from the conditional distribution of CM_n given it is simple, hence from the desired uniform distribution.

Regarding the asymptotic feasibility of the above rejection sampling one has to know whether $\mathbf{P}\Big\{ \mathrm{CM}_n(d_i^{(n)}) \text{ is simple} \Big\}$ is asymptotically positive. We cite the following result to this regard.

With the multi-graph representation $(x_{ij}, i, j \in V_n)$ denote by $S_n = \sum_{i=1}^n x_{ii}$ the number of self-loops and $M_n = \sum_{i<j=1}^n (x_{ij} - 1)^+$ the number of multiple edges in CM_n. Denote

$$\nu := \frac{\mathbf{E}\,[D(D-1)]}{\mathbf{E}\,[D]} = \frac{\sum_{k=2}^\infty k(k-1)p_k}{\sum_{k=1}^\infty k p_k}\,. \tag{4.1.4}$$

By condition (ii) the denominator in (4.1.4) is finite and $\nu > 0$ provided $p_0 + p_1 < 1$.

Proposition 4.1.5. *Assume $p_0 + p_1 < 1$ and $\nu < \infty$. Then*

$$(S_n, M_n) \Rightarrow (S, M), \qquad n \to \infty,$$

where S and M are independent Poisson variables with parameters, $\nu/2$ and $\nu^2/4$, respectively. Consequently

$$\mathbf{P}\Big\{\, \mathrm{CM}_n(d_i^{(n)}) \text{ is simple} \,\Big\} \to e^{-\nu/2-\nu^2/4} > 0 \qquad n \to \infty\,.$$

Obviously if $p_0 + p_1 = 1$ then $S_n \to 0$, $M_n \to 0$ in probability.

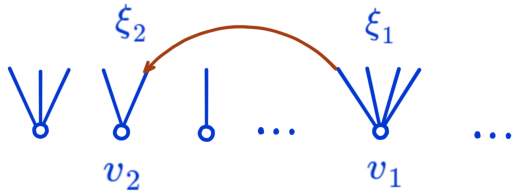
Cf ((Van Der Hofstad, 2017, Proposition 7.9)).

Remark 4.1.6 (Repeated CM)**.** A *repeated CM* is a sequence of conditional realizations of CM_n, $n \geq 1$ each given being a simple graph. By Proposition 4.1.5 it is asymptotically well defined model (one can study asymptotic properties of this sequence of conditional random graphs). Naturally, all properties observed asymptotically with probability 1 for CM hold asymptotically with probability 1 also for the repeated CM.

4.1.5 Tree-like local structure of the configuration model

Let $v_1 = v_1^{(n)}$ be a vertex uniformly selected from V_n. Let $\xi_1 := d_{v_1}^{(n)}$ be its degree. Then ξ_1 has the distribution of $D_n \Rightarrow D$, $n \to \infty$, cf. Section 4.1.3.

If $\xi_1 > 0$, take any of the half-edges of v_1 and denote by ξ_2 the degree of the node to which v_1 is connected by this half-edge.



When $\xi_1 > 0$, the conditional distribution of ξ_2 given ξ_1 can be written as follows: for $k > 0$

$$\mathbf{P}\{\, \xi_2 = k \,|\, \xi_1 \,\} = \begin{cases} \frac{kn_k}{\ell_n - 1} & \text{if } k \neq \xi_1, \\ \frac{k(n_k - 1) + (k-1)}{\ell_n - 1} & \text{if } k = \xi_1 \end{cases}$$

and $\mathbf{P}\{\, \xi_2 = 0 \,|\, \xi_1 \,\} = 0$. Note, given $\xi_1 > 0$,

$$\lim_{n\to\infty} \mathbf{P}\{\, \xi_2 = k \,|\, \xi_1 \,\} = \lim_{n\to\infty} \frac{kn_k}{n} \frac{n}{\ell_n} = \frac{kp_k}{\lambda}$$

for $k \geq 0$. Call by $\tilde{D}$ a random variable with distribution $\{\tilde{p}_k := kp_k/\lambda\}$

$$\mathbf{P}\Big\{\,\tilde{D} = k\,\Big\} = \tilde{p}_k = \frac{kp_k}{\lambda} = \frac{k\mathbf{P}\{\,D = k\,\}}{\mathbf{E}\,[D]}. \tag{4.1.5}$$

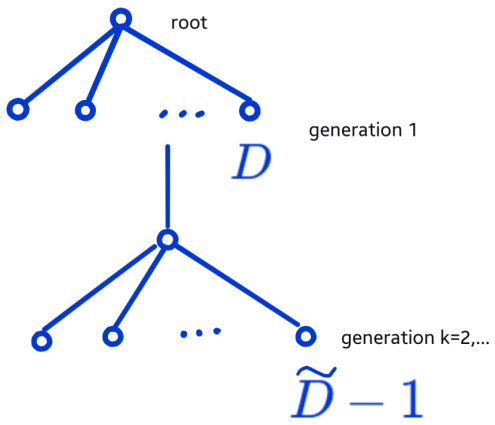
This is so-called *size-biased* modification of the distribution D. Observe that $\mathbb{E}\,[\tilde{D}] = \nu + 1$, where ν is given by (4.1.4).

The arguments presented above say that the first variable of the history of the exploration of the neighbourhood of the randomly, uniformly selected node in CM_n has the typical degree distribution D. However, the second variable of this history has the *size-biased distribution* $\tilde{D}$. This observation can be extended to $\xi_3, \xi_4, \ldots, \xi_k$ for finite k (cf Fact 1.2 for ER graph). Proving further that the number of undiscovered edges in the exploration of CM_n tends to 0 when $n \to \infty$ (cf Remark 1.1 and Exercise 1 for ER graph) one can conclude with the following observation.

Remark 4.1.7.

- *CM is (asymptotically for $n \to \infty$) locally* $\mathrm{GW}(D, \tilde{D})$ *tree-like, with the offspring distribution D of the root and the offspring distribution $\tilde{D} - 1$ of all other nodes. The above statement can be formalized on the ground of local weak convergence of random graphs; cf. Section 2 for of the lesson on ER graphs.*

- *A CM with Poisson degree distribution D converge locally weakly to the Poisson GW tree. In this case (and only in this case) D and $\tilde{D} - 1$ have the same distribution; cf Exercise 4. Hence CM with Poisson degree distribution is locally, asymptotically similar to the ER graph having the same mean degree.*



4.2 Emergence of the giant component

4.2.1 Phase transition via a local tree approximation

The above tree approximation, which describes the *local structure* of a connected component of CM, can be used to study (or at least conjecture) the conditions of the emergence of the

giant component in CM and calculate its relative size, similarly as in ER case.

Denote by $p_{\text{ext}}(D, \tilde{D})$ the extinction probability of $\text{GW}(D, \tilde{D})$. Observe that $\text{GW}(D, \tilde{D})$ is finite iff all sub-trees rooted at the nodes of the first generation are finite. By the independence

$$\begin{aligned} p_{\text{ext}}(D, \tilde{D}) &= \mathbf{E}\left[p_{\text{ext}}(\tilde{D}-1)^D\right] \\ &= \phi_D(p_{\text{ext}}(\tilde{D}-1))\,, \end{aligned}$$

where $p_{\text{ext}}(\tilde{D}-1)$ is the extinction probability of the "homogeneous" $\text{GW}(\tilde{D}-1)$ and $\phi_D(s)$ is the p.g.f. of D. Recall that $p_{\text{ext}}(\tilde{D}-1)$ is the smallest solution of the equation $s = \phi_{\tilde{D}-1}(s)$ in $s \in [0,1]$.

Observe also that $p_{\text{ext}}(D, \tilde{D}) < 1$ iff $p_{\text{ext}}(\tilde{D}-1) < 1$ (since $p_1 > 0$) and hence iff

$$\nu := \mathbf{E}\left[\tilde{D}-1\right] = \frac{\mathbf{E}\left[D(D-1)\right]}{\mathbf{E}\left[D\right]} > 1 \tag{4.2.1}$$

(since $\mathbf{P}\Big\{\,\tilde{D}-1=0\,\Big\} = \tilde{p}_1 = p_1/\lambda > 0$). Note that condition (4.2.1) can be equivalently rewritten as $\mathbf{E}\left[D(D-2)\right] > 0$.

The above local tree approximation of the CM allows one to understand the following result.

Proposition 4.2.1. CM is sub-critical *if* $\nu < 1$, *in which case*

$$\mathbf{P}\{\,|C_1| \le a\log(n)\,\} \to 1 \quad \textit{as } n \to \infty$$

for some constant a.
CM is super-critical *if* $\nu > 1$, *in which case the size of the largest component* C_1 *satisfies*

$$\mathbf{P}\left\{\,\left|\frac{|C_1|}{n} - (1 - p_{\text{ext}}(D, \tilde{D}))\right| \le \delta,\ |C_2| \le a'\log n\,\right\} \to 1 \quad \textit{as } n \to \infty,$$

for some a' *and all* $\delta > 0$.

In what follows we will present another approach allowing one to prove the above result.

4.2.2 Phase transition via the law of large numbers

A different approach to the study of the emergence of the giant component in CM is possible via the law of large numbers for the exploration process. In a sense, it shows the creation of the giant component from a macroscopic point of view.

We will consider a variant of the exploration process of CM_n in *continuous time*, whose tempo is dictated by iid Exponential(1) random variables T_i, $i = 1, \cdots, \ell_n$ interpreted as *lifetimes* of the half-edges. All half-edges start their lives at time $t = 0$ and at time T_i the half-edge i *dyes spontaneously*. (Some of the half-edges will be however *killed* before this time ...)

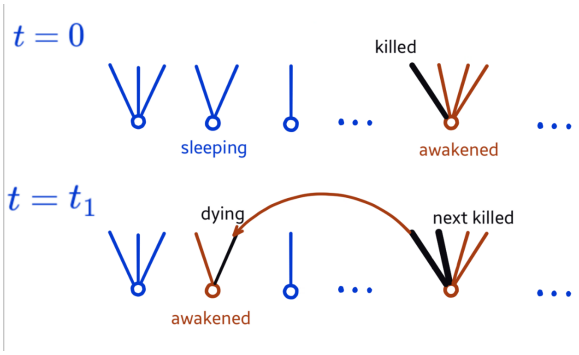
At teach time $t \ge 0$ we will observe the following numbers:

- the number of *sleeping half-edges*, $S(t)$,
- the number of *sleeping vertexes of degree* k, $V_k(t)$,
- the number of *awake half-edges*, $A(t)$,
- the number of *living half-edges*, $L(t)$,

$$L(t) := A(t) + S(t)\,. \tag{4.2.2}$$

All the above stochastic processes are pure jump processes (say, càdlàg [5]) changing their values at the epochs T_i, $i = 1\ldots, \ell_n$ (spontaneous deaths of the half-edges). Here is how these process evolve. Initially all vertexes and edges are sleeping.

1. When there is no awake half-edge (as at the beginning), select a sleeping vertex (say uniformly) and awake it as well as all its half-edges. When there is no sleeping half-edge left, the process stops; the remaining sleeping vertices are all isolated and we have explored all other components.
2. Pick an awake half-edge (say uniformly) and kill it.
3. Wait until the next half-edge dies spontaneously (its exponential life-time T_i expires). This half-edge is joined to the one killed in the previous step 2 to form an edge of the graph. When the vertex incident to it is sleeping, we awake it as well as all other half-edges incident to it. Repeat from step 1.

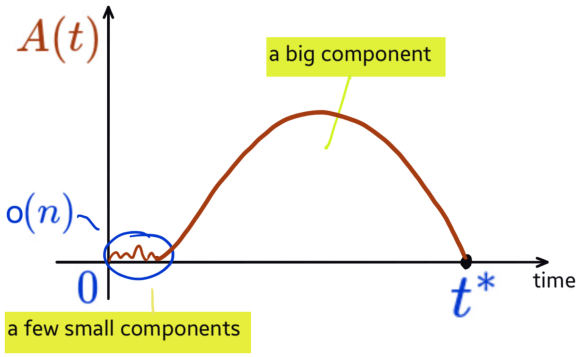


Note that each time step 1 is performed (no active half-edge and, as a result, a new vertex is chosen) we complete the creation of some component (except at time 0) and start creating a new one. When this happens, the number of alive vertexes is close to 0, $A(t) \approx 0$. More precisely, a component is completed at every time T_i when either of the following scenarios occurs:

[5] continuous from the right, having limits from the left

- right before T_i there is only one alive half-edge ($A(T_i-) = 1$) and this half-edges i is dying spontaneously at T_i
- the is no active half-edge right before T_i ($A(T_i-) = 0$) and the dying half-edge i belongs to a sleeping vertex of degree 1.

Macroscopically, for large n, this might happen several times very shortly (in time $o(n)$) after time $t = 0$. The total number of vertexes in all (small) components created during this time can be shown $o(n)$. [6] Then, the process $A(t)$ makes a large excursion in $A(t) > 1$, during which a large component, call it C_1, is discovered. It is completed when $A(t)$ comes back to the neighbourhood of 0. This behaviour is well visible in the limit of the law of large numbers for the process $A(t)$.



Denote $\overline{A}(t) := \lim_{n\to\infty} A(t)/n$ and similarly for $\overline{V}_k(t)$, $\overline{S}(t)$, $\overline{L}(t)$. These limits exist, for some range of $t > 0$. Moreover, it is possible to prove that $\overline{A}(t)$ is strictly positive for $t \in (0, t^*)$ for some $t^* > 0$ iff $\mathbf{E}[D(D-2)] > 0$, i.e., iff $\nu > 1$ (cf (4.2.1)), corresponding the super-critical regime of CM.

In order to characterize the point t^* and calculate the size (and even degree distribution) of the big component created between time 0 and t^* write

$$\begin{aligned}\overline{A}(t) &= \overline{L}(t) - \overline{S}(t)\\ &= \overline{L}(t) - \sum_{k=1}^{\infty} k\overline{V}_k(t)\,.\end{aligned}$$

A crucial observation regarding $\overline{L}(t)$ and $\overline{V}_k(t)$ is the following law or large numbers.

Fact 4.2.2. *For $n \to \infty$, in probability*

$$\frac{1}{n}L(t) \Rightarrow \overline{L}(t) = \lambda e^{-2t} \qquad \textit{uniformly for } t \geq 0,$$

$$\frac{1}{n}V_k(t) \Rightarrow \overline{V}_k(t) = p_k e^{-kt} \qquad \textit{uniformly for } t \in [0, t^*],$$

[6] Note that as $n \to \infty$ the exploration process runs faster at the initial phase. Indeed the exponential time between two successive (spontaneous) deaths of the half-edges has the parameter of order $\ell_n = O(n)$ (and mean $1/\ell_n = O(1/n)$).

where $t^* = -\log s^*$ *and* s^* *is the smallest solution of the equation* $s = \phi_{\tilde{D}-1}(s)$ *in* $s \in [0,1]$

Before saying how the above result can be proved, let us use it and solve the equation

$$\begin{aligned}
\overline{A}(t) = \overline{L}(t) - \sum_{k=1}^{\infty} k\overline{V}_k(t) &= 0 \\
&\Updownarrow \\
\lambda e^{-2t} - \sum_{k=1}^{\infty} k p_k e^{-kt} &= 0 \\
\boxed{\text{substituting } s := e^{-t}} \qquad &\Updownarrow \\
\lambda s^2 - \sum_{k=1}^{\infty} k p_k s^k &= 0 \\
&\Updownarrow \\
s = \sum_{k=1}^{\infty} \frac{k p_k}{\lambda} s^{k-1} &= 0 \\
&\Updownarrow \\
s = \phi_{\tilde{D}-1}(s). &
\end{aligned}$$

Recall, under condition $\nu > 1$ the smallest solution s^* of the above equation [7] satisfies $s^* < 1$ and hence

$$t^* = -\log(s^*) > 0\,.$$

The relative number of all vertexes sleeping at time t^* can be expressed as

$$\begin{aligned}
\sum_{k=0}^{\infty} \overline{V}_k(t^*) &= \sum_{k=0}^{\infty} p_k e^{-kt^*} \\
&= \sum_{k=0}^{\infty} p_k (s^*)^k \\
&= \phi_D(s^*).
\end{aligned}$$

Neglecting the total number $o(n)$ of vertexes in all small components created before the large one C_1 is completed we can conclude that

$$\begin{aligned}
\lim_{n\to\infty} \frac{|C_1|}{n} &= 1 - \sum_{k=0}^{\infty} \overline{V}_k(t^*) \\
&= 1 - \phi_D(s^*) \\
&1 - p_{\text{ext}}(D, \tilde{D})\,,
\end{aligned}$$

[7] which is $s^* = p_{\text{ext}}(\tilde{D} - 1)$ the extinction probability of the "homogeneous" GW$(\tilde{D} - 1)$

which complies with the statement regarding the largest component made in Proposition 4.2.1. Moreover, the vertexes sleeping at time t^* form another CM, which can be proved sub-critical. cf Exercise 5.

Proff (idea) of Fact 4.2.2. Consider first the process $V_k(t)$. Note that any sleeping vertex is awakend when its first half-edge is dying spontanelusly or as a conseqeunce of the step 1 of the exploration algorithm (when there is no active half-edge and a new vertex is awakened uniformly among sleeping vertexes). Ignoring this second possibility (which awakes only $o(n)$ vetrexes before time t^*) we have

$$\begin{aligned}
\frac{V_k(t)}{n} &= \frac{1}{n}\sum_{v=1}^{n_k} \mathbf{1}(\text{vertex } v \text{ of degree } k \text{ is still alive at } t) \\
\boxed{\text{ignoring step 1}} \quad &\approx \frac{1}{n}\sum_{i=1}^{n_k} \mathbf{1}(\text{none of } k \text{ half-edges of } v \text{ dye before } t) \\
&= \frac{1}{n}\sum_{i=1}^{n_k} \mathbf{1}(\min_{i\in\text{edges of } v} T_i > t) \\
\boxed{\text{Exp}_i(k) \text{ idependent in } i} \quad &= \frac{n_k}{n}\frac{1}{n_k}\sum_{i=1}^{n_k} \mathbf{1}(\text{Exp}_i(k) > t) \\
\boxed{\text{(i) of Sec. 4.1.2 and Glivenko-Cantelli}} \quad &\to p_k \mathbf{P}\{\,\text{Exp}(k) > t\,\} \\
&= p_k e^{-kt}\,,
\end{aligned}$$

where in the limit we have used the assumpiton (i) of Section 4.1.2 and Glivenko-Cantelli's theorem regarding the asymptotic behaviour of the empirical distribution functions.

Regarding the process of alive edges $L(t)$, note it is a pure death Markov process, which stays in state $L(t) = x$ for a random $\text{Exp}(x)$ time (until the expiration of the smallest of x unit exponential variables), and then jumps by -2, i.e., goes to the state $L(t) = x - 2$.

It is clear that the process $L'(t) := L(t)/2$ can be seen as a pure death Markov process, which stays in state $L'(t) = x$ for random $\text{Exp}(2x)$ time, and then jumps by -1, i.e., goes to the state $L(t) = x - 1$. Note that $L'(0) = \ell_n/2$. Using the same Glivenko-Cantelli's theorem for $L'(t)$ we obtain

$$\begin{aligned}
\frac{1}{n}L(t) &= \frac{2}{n}L'(t) \\
&= \frac{\ell_n}{n}\frac{1}{\ell_n/2}L'(t) \\
\boxed{\text{Exp}_u(2) \text{ independent in } u} \quad &= \frac{\ell_n}{n}\frac{1}{\ell_n/2}\sum_{u=1}^{\ell_n/2} 1(\text{Exp}_u(2) > t)
\end{aligned}$$

(iii) Sec. 4.1.2, cf Exe. 2 and Glivenko-Cantelli $\rightarrow \lambda e^{-2t}$.

□

For more readings on this approach to the CM, including proof details see ((Van Der Hofstad, 2014, Section 4.1)) or the original article Janson and Luczak ((2009)).

TODO

In relation to the configuration model

- present foundations of diffusion models in the networks.

4.3 Exercises

1. Prove that under assumption (iii) of Section 4.1.2

$$d_{\max}^{(n)} := \max_{i=1,\dots,n} d_i^{(n)} = O(\sqrt{n}) \qquad \text{for large } n.$$

2. Prove that condition (iii) from Section 4.1.2 implies that D_n are uniformly integrable (i.e., for any $\epsilon > 0$ there exists $K < \infty$ such that $\mathbf{E}\,[D_n \mathbf{1} D_n > k] \leq \epsilon$). Conclude that $\mathbf{E}\,[D_n] \to \mathbf{E}\,[D]$.

3. Prove that the degree sequence of Example 4.1.1 satisfies conditions (0)–(iii) of Section 4.1.2.

4. Let D be random variable on $0, 1, \dots$, with the finite mean and $\tilde{D}$ its size-biased version. Show that $\tilde{D} - 1$ is equal in distribution to D iff D is Poisson.

5. Calculate the (asymptotic) degree distribution of the nodes still sleeping at time t^*, when the big component has been completed (cf Section 4.2.2). Observe that it corresponds to a sub-critical CM mode.

6. *Computer exercise.* Consult *KONECT* (Koblenz Network Collection) database at `http://konect.uni-koblenz.de/` collecting large network datasets of all types available to perform research in network science and related fields.

   (a) Chose one network, download the corresponding data and calculate the empirical degree distribution. Double-check with the plots given on `http://konect.uni-koblenz.de/plots/degree_distribution`.

   (b) Generate and visualize a realization of the Configuration model fitting the chosen network (its degree distribution) using `degree.sequence.game` command of the package *igraph* [8]. For simplicity you may need to consider this graph on a smaller number of nodes than in the original data set.

   (c) If the original network does not have multiple edges and self-loops you need to consider the repeated Configuration model using the rejection sampling; i.e., simulate different realizations of the Configuration model until a simple graph is obtained, using `is.simple` command of the *igraph* package to check whether a given realization is simple.

[8] *igraph*, `http://igraph.org/redirect.html` is a library collection for creating and manipulating graphs and analyzing networks. It is written in C and also exists as Python and R packages, with an interface for Mathematica.

Lesson 5

Typical nodes and random unimodular graphs

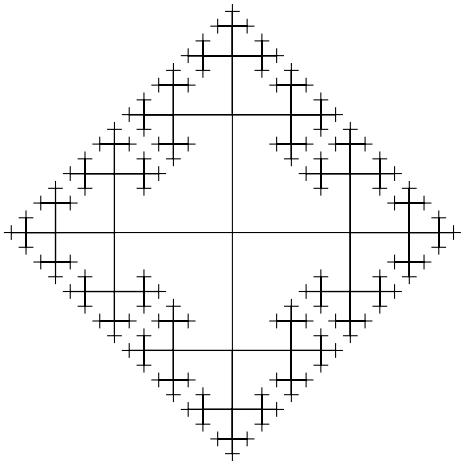


Figure 5.1: Part of the infinite 4-regular tree which is a Cayley graph corresponding to the free group on two generators. This tree is symmetric, meaning all edges have the same relative position in the tree. This makes any node can be considered as the typicla one.

The typical vertex of a finite graph can be naturally defined as the vertex uniformly sampled from all vertexes. This definition cannot be straightforwardly extended to infinite graphs. Unimodular (infinite) graphs can be seen as graphs which exhibit enough regularity allowing one to come up with an equivalent notion of the typical vertex. The term "unimodularity" comes from the study of some class of deterministic graphs, more precisely Cayley graphs which encode the abstract structure of groups. In this lesson we shall briefly present random unimodular random graphs. These are rooted graphs, where the root can be seen as the typical vertex of the graph. Their fundamental, defining, property, called the *mass transport principle*, will appear again when we will study the typical point of the stationary point pro-

cess. For more reading we refer to a very complete treatment of the subject in Aldous and Lyons ((2007)) and also to a related review article Aldous and Steele ((2004)).

5.1 Unimodularity and the typical node

5.1.1 Finite, uniformly rooted graphs

Let us consider a random graph $G = (V, E)$ with vertexes V being a (possibly random) subset of some countable space that, without loss of generality, can be mapped to the natural numbers; $V \subset \mathbb{N}$. We denote non-oriented edges of G by E and write $v_1 \overset{G}{\sim} v_2$ if there is an edge in G between $v_1, v_2 \in V$. We also say then that v_1 and v_2 are *neighbours* in G. If there is no ambiguity, we will omit G writing $v_1 \sim v_2$.

The degree $d_G(v)$ of vertex $v \in V$ in graph G, or simply $d(v)$, is the number of its neighbours

$$d(v) = \sum_{u \in V} \mathbf{1}(v \sim u) = \sum_{x \in \mathbb{N}} \mathbf{1}(v \sim x).$$

The *(graph) distance* between two nodes $u, v \in V$ is the length (number of edges) of a shortest path between u and v in G. The *ball* of radius $\rho \geq 0$ centered at $v \in V$ (we say also the neighbourhood of this vertex) is a sub-graph of G consisting of all vertexes, whose distance to v not larger than ρ.

Our goal is to formalize the notion of the *typical vertex* of the graph G. In this regard, we call a graph G with some selected vertex $o \in V$, called root, *rooted graph* (G, o). Both the graph and its root can be random. We want the selected node o so as that it models the typical vertex of G. A natural way of doing this consists in choosing o uniformly from vertexes of G. This can be done however only if G is a finite graph, i.e., $|V| < \infty$.

Definition 5.1.1 (Uniformly rooted graph)**.** A finite, rooted graph (G, o) is called *uniformly rooted* if given G, o is uniformly distributed on V, i.e., $\mathbf{P}\{\, o = x \,|\, G \,\} = \mathbf{1}(x \in V)/|V|$ for all $x \in \mathbb{N}$.

Equivalently, finite rooted graph (G, o) is uniformly rooted if for any, say non-negative, function $h(G, x)$ of the graph G and $x \in \mathbb{N}$

$$\mathbf{E}\,[h(G, o)] = \mathbf{E}\left[\frac{1}{|V|} \sum_{v \in V} h(G, v)\right]. \tag{5.1.1}$$

Example 5.1.2. The Erdős-Rényi graph ER_n on n nodes with arbitrarily selected vertex $v \in \{1, \ldots, n\}$, (ER_n, v) is uniformly rooted. The Configuration model CM_n on n nodes, with uniformly sampled node $o \in \{1, \ldots, n\}$ is uniformly rooted. (CM_n is not necessarily a simple graph but the framework of uniformly rooted graphs can be straightforwardly extended to multi-graphs.)

Clearly, neither Definition 5.1.1 nor condition (5.1.1) can be straightforwardly extended to infinite graphs, for example to super-critical Galton-Watson trees. In what follows we give another equivalent condition that will offer such possibility.

Proposition 5.1.3. *A finite, rooted graph (G, o) is uniformly rooted iff*

$$\mathbf{E}\left[\sum_{v\in V} f(G,o,v)\right] = \mathbf{E}\left[\sum_{u\in V} f(G,u,o)\right] \tag{5.1.2}$$

for all, say non-negative, functions $h(G, x, y)$ of the graph G and $x, y \in \mathbb{N}$.

Let us make some interpretation before we prove the result. For $u, v \in V$, we interpret $f(G, u, v)$ as the *amount of mass sent from u to v through the graph G*. With this interpretation, the left-hand-side of (5.1.2) corresponds to the *mean total mass sent by the root* while the right-hand-side to the *mean total mass received by the root*. When (5.1.2) holds for all functions $h(G, x, y)$ we say that the finite rooted graph (G, o) satisfies the *mass transport principle*.

Note that the condition (5.1.2) makes sense for an infinite graph. It allows one to extend the notion of the typical node of the graph to infinite graphs, as we shall see in Section 5.1.2.

Proof of Proposition 5.1.3. The proof is merrily a matter of a change of summation.
($\Rightarrow$) For the direct part, assume that (G, o) is uniformly rooted. Then

$$\begin{aligned}
\mathbf{E}\left[\sum_{v\in V} h(G,o,v)\right] &= \mathbf{E}\left[\sum_{y\in\mathbb{N}} \mathbf{1}(y\in V)h(G,o,y)\right] \\
&= \sum_{y\in\mathbb{N}} \mathbf{E}\left[\mathbf{1}(y\in V)h(G,o,y)\right] \\
\text{by (5.1.1)} \quad &= \sum_{y\in\mathbb{N}} \mathbf{E}\left[\frac{1}{|V|}\mathbf{1}(y\in V)\sum_{u\in V} h(G,u,y)\right] \\
&= \mathbf{E}\left[\frac{1}{|V|}\sum_{y\in\mathbb{N}} \mathbf{1}(y\in V)\sum_{x\in\mathbb{N}} \mathbf{1}(x\in V)h(G,x,y)\right] \\
&= \mathbf{E}\left[\frac{1}{|V|}\sum_{x\in\mathbb{N}} \mathbf{1}(x\in V)\sum_{y\in\mathbb{N}} \mathbf{1}(y\in V)h(G,x,y)\right] \\
&= \sum_{x\in\mathbb{N}} \mathbf{E}\left[\frac{1}{|V|}\mathbf{1}(x\in V)\sum_{v\in V} h(G,x,v)\right] \\
\text{by (5.1.1)} \quad &= \sum_{x\in\mathbb{N}} \mathbf{1}(x\in V)\mathbf{E}\left[h(G,x,o)\right]
\end{aligned}$$

$$= \mathbf{E}\left[\sum_{u\in V} h(G,u,o)\right].$$

($\Leftarrow$) For the inverse part, consider a non-negative function $h(G,x)$, define $f(G,x,y) = h(G,x)/|V|$ (does not depend on y). The left-hand side of (5.1.2) gives

$$\mathbf{E}\left[\sum_{v\in V} f(G,o,v)\right] = \mathbf{E}\left[\sum_{v\in V} h(G,o)/|V|\right] = \mathbf{E}\left[h(G,o)\right].$$

The right-hand side of (5.1.2) gives

$$\mathbf{E}\left[\sum_{u\in V} f(G,u,o)\right] = \mathbf{E}\left[1/|V| \sum_{u\in V} h(G,u)\right]$$

proving the relation (5.1.1) and consequently that (G,o) is uniformly rooted. □

5.1.2 Unimodular graphs

Random, locally finite, rooted graphs

We assume now that G is possible infinite, but *locally finite*, that is all nodes have finite degrees; $d(v) < \infty$ for all $v \in V$. We want to extend the notion of the typical node to locally finite graphs. In this regard, first, we have to be more specific regarding the measure-theoretic framework. While there are countably many finite graphs with vertexes in a countable space ($\mathbb{N}$ considered in the previous section) and defining probability measures on such discrete space is straightforward, this is no longer the case for locally-finite graphs and we have to clarify how we define probability measure on such graphs.

First, we are interested only in the "intrinsic" properties of graphs, properties which do not depend on the way the graph vertexes are labeled. That is why we shall "identify" two different graph realizations if there is a way to map the vertexes of one to the other in such a way that the root and edges are preserved. Formally, we say that two (given, deterministic) rooted (possibly multi-) graphs (G,o) and (G',o'), with vertexes possibly in different spaces, are *isomorphic* if there is a bijective function ϕ from the vertexes V of G to the vertexes V' of G' such that $\phi(o) = o'$ and $u \overset{G}{\sim} v$ iff $\phi(u) \overset{G'}{\sim} \phi(v)$. We write then $(G,o) \simeq (G',o')$.

Denote by $\mathcal{G}^*$ the set of all equivalence classes of this isomorphism relation $\simeq$ on *locally finite, connected, rooted (multi-)graphs*. On can define a metric [1] on $\mathcal{G}^*$ making it complete and separable metric space (Polish space). We consider the Borel σ-field on this space. A random rooted graph is defined as a measurable mapping from some probability space to $\mathcal{G}^*$.

For simplicity we will not distinguish between graphs and their equivalence classes and we use the same terminology and notation. That is, we denote random rooted graph by (G,o)

[1] The distance between the classes of equivalence of (G,o) and (G',o') can be defined as $1/(1+a)$, where a is the supremum of those $r > 0$ for which there is some isomorphism of the balls of (graph-distance) radius r around the roots in G and G'.

keeping in mind that is it only a "representative" of all rooted graphs which are isomorphic ($\simeq$ equivalent) to it. A way to bypass this identification is to choose once for all a canonical representative in each class among the graphs on $\mathbb{N}\cup\{0\}$ rooted at 0; see ((Aldous and Lyons, 2007, Section 2)).

Unimodularity

We formulate now the most important property of this lesson, replacing (extending) the notion of uniformly rooted graphs. As in (5.1.2) it involves functions $f(G,u,v)$ of the graph G and its two selected nodes $u,v\in V$. In the setting of locally-finite graphs, formally this will be a measurable function on the space $\mathcal{G}^{**}$ of the equivalence classes of the isomorphisms of graphs with two nodes selected [2], as a Polish space with the metric defined similarly as on $\mathcal{G}^{*}$. Again, we will not distinguish in the notation between (G,u,v) and its equivalence class. We should remember however that $f(G,u,v)$, *as a function on* $\mathcal{G}^{**}$, *is invariant with respect to the isomorphism*:

$$f(G,u,v)=f(G',u',v') \quad \text{if } (G,u,v)\simeq f(G',u',v'), \tag{5.1.3}$$

where $\simeq$ denotes an isomorphism of G to G' preserving edges and mapping u to u' and v to v'.

Definition 5.1.4 (Mass transport principle (MTP))**.** A random rooted graph (G,o) in $\mathcal{G}^{*}$ satisfies the *MTP* if

$$\mathbf{E}\left[\sum_{v\in V} f(G,o,v)\right]=\mathbf{E}\left[\sum_{u\in V} f(G,u,o)\right] \tag{5.1.4}$$

for all, non-negative, measurable functions f an $\mathcal{G}^{**}$ (informally: invariant with respect to the isomorphism; i.e., satisfying (5.1.3)).
Rooted graphs (G,o) in $\mathcal{G}^{*}$ satisfying the *MTP* are called *unimodular graphs*.

Remark 5.1.5. The MTP is a straightforward extension of the condition formulated in Proposition 5.1.3 with one modification: We require the equality (10.2.4) only for functions on $\mathcal{G}^{**}$, in other words functions of (G,u,v) invariant with respect to the graph isomorphism. Without this important limitation, even the most natural examples of rooted infinite graphs would not satisfy the MTP; cf. Exercise 3.

Corollary 5.1.6. *A finite rooted graph* $(G,o)\in\mathcal{G}^{*}$ *is unimodular (satisfies MTP) iff it is uniformly rooted in the sense that it satisfies* (5.1.1) *for measurable functions* h *on* $\mathcal{G}^{*}$ *(invariant with respect to the isomorphism of rooted graphs).*

[2] $(G,x,y)\simeq(G',x',y')$ iff there is a bijective function ϕ from the vertexes V of G to the vertexes V' of G' such that $\phi(x)=x'$, $\phi(y)=y'$ and $u\overset{G}{\sim}v$ iff $\phi(u)\overset{G'}{\sim}\phi(v)$.

Remark 5.1.7. Note that the a finite rooted graph (G, o) considered as an equivalence class in $\mathcal{G}^*$ is uniformly rooted if

$$\mathbf{P}\{\,(G,o) \simeq (G,v)\,|\,G\,\} = \frac{|\{u \in V : (G,u) \simeq (G,v)\}|}{|V|}\,,$$

which can be rephrased as that the conditional probability that the root o is located at some vertex equivalent to v is equal to the proportion of vertexes of V whose "relative position" in G is equivalent to this of v. (We do not have possibility to distinguish between such vertexes by functions on $\mathcal{G}^*$.)

Example 5.1.8. The connected component of the Erdős-Rényi graph ER_n rooted at an arbitrarily selected vertex $v \in \{1, \dots, n\}$ is unimodular.
The connected component of the Configuration model CM_n on n rooted at the uniformly sampled node $o \in \{1, \dots, \}$ is unimodular.

The following result proved in ((Aldous and Lyons, 2007, Proposition 2.2)) allows one to replace equivalently the sums $\sum_{u \in V}$ and $\sum_{v \in V}$ in (10.2.4) by the sums over neighbours only $\sum_{v \sim u}$.

Proposition 5.1.9. *A random rooted graph (G, o) in $\mathcal{G}^*$ satisfies the* MTP *iff it satisfies (10.2.4) for all, non-negative, measurable functions f an $\mathcal{G}^{**}$ supported on the graph edges ($f(G, x, y) = 0$ if $x \not\sim y$). (Such graphs are called involution invariant.)*

5.2 Examples

5.2.1 Deterministic unimodular graphs

The most natural candidates for the unimodular graphs are *deterministic graphs which satisfy some regularity conditions*. The term "unimodular" graphs comes in fact from this class of graphs, cf ((Aldous and Lyons, 2007, Section 1 and 2)).

Consider a deterministic graph G whose all vertexes have the same "relative position" in G. Than we could just fix an arbitrary vertex as o rather than try to choose one uniformly. However, we shall see that the MTP will not be satisfied without some additional regularity assumption on G.

We have the following two definitions.

Definition 5.2.1. We say that a deterministic graph $G = (V, E)$ is

- *transitive* (or vertex symmetric) if $(G, u) \simeq (G, u')$ for all $u, u' \in V$. (All vertexes have the same "relative position" in G.)
- *symmetric* (or edge symmetric) if for $(G, u, v) \simeq (G, u', v')$ for all $u, u', v, v' \in V$ such that $u \sim v$ and $u' \sim v'$. (All edges have the same "relative position" in G.)

Both notions are some kind of regularity conditions and it is easy to see that the symmetry is a stronger condition than the transitivity.

Corollary 5.2.2. *A connected, symmetric graph G is transitive.*

The symmetry is enough for the MTP.

Proposition 5.2.3. *A (deterministic) symmetric connected graph rooted at an arbitrary node is unimodular.*

Proof. We shall verify the version of the MTP supported on the edges, given in Proposition 5.1.9 (the involution invariance). By the symmetry of the graph G, all edges "have the same relative position in G" and thus any function f on $\mathcal{G}^{**}$ (isomorphism invariant), supported on the edges must be constant on all edges

$$f(G,x,y) = f(G,y,x) = \text{Constant}(G)\mathbf{1}(x \sim y)\,.$$

Consequently

$$\sum_{v\sim o} f(G,o,v) = \text{Constant}(G)\, d(o) = \sum_{u\sim o} f(G,u,o)\,.$$

□

Example 5.2.4 (d-regular trees)**.** A d-regular graph is a graph with all vertexes having degree equal to d. d-regular trees are symmetric hence unimodular when rooted at an arbitrary node.

The following example shows that the *transitivity is not enough for the unimodularity.*

Example 5.2.5 (3-regular tree with grandfathers)**.** Consider 3-regular tree (solid edges on Figure 5.3) with some (dotted) edges added. (The resulting graph is no longer a tree). The dotted edges are added according to the following rule: Select (temporarily) an arbitrary node, e.g. the black dot on Figure 5.3, and some path from it to (minus) infinity e.g. the blue curve on the figure. We consider all nodes on this path the *ancestors* of the selected node (that is why we say the path goes to $-\infty$.) This also uniquely defines all ancestors for all other nodes. The additional (dotted) edges connect all nodes to their grandfathers.

Having constructed the graph we can forget about the initially selected node. This graph is *transitive*; all vertexes have the same "relative position" in G as suggested by the following picture with three different positions (orange dots) for the root.

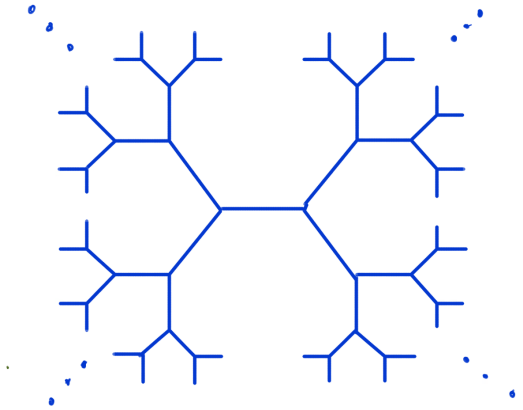


Figure 5.2: 3-regular tree.

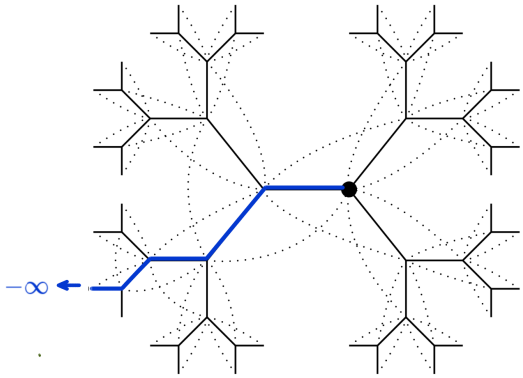
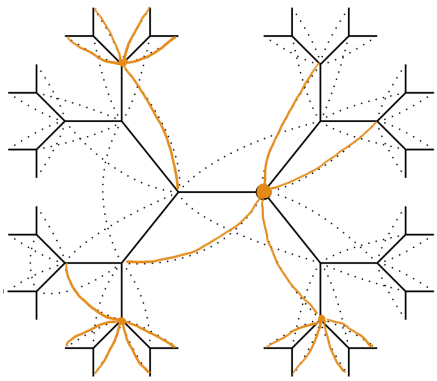


Figure 5.3: 3-regular tree with grandfathers.



However the 3-regular tree with grandfathers is *not symmetric*. Indeed, for any solid edge

$u \sim v$  $(G,u,v) \not\simeq (G,v,u)$; note however, we should not distinguish between types of edges when verifying this property. The following isomorphism invariant function does not satisfy (10.2.4) showing that this graph (rooted at any node) is not unimodular. Take

$$f(G,u,v) = \mathbf{1}(u \text{ is the grandfather of } v) = \begin{cases} 1 & \text{if one observes } \\ 0 & \text{otherwise.} \end{cases}$$ 

Indeed

$$\sum_{v \in V} f(G,o,v) = 4 \quad \boxed{\text{root (as every node) is a grandfather of 4 grandchildren}}$$

$$\sum_{u \in V} f(G,u,o) = 1 \quad \boxed{\text{root (as every node) has one grandfather}}.$$

5.2.2 Random unimodular graphs

To produce examples of random, infinite unimodular graphs we will consider the limits of finite unimodular (equivalently, uniformly rooted) graphs and relay on the following result.

Proposition 5.2.6. *Let (G_n, o_n) be a sequence of unimodular random graphs in $\mathcal{G}^*$ that converges in distribution* [3] *to some random rooted graph (G,o); we denote is $(G_n,o_n) \Rightarrow (G,o)$. Then (G,o) is unimodular.*

The converges in distribution on $\mathcal{G}^*$ can be equivalently related to the following convergence of rooted graphs.

Definition 5.2.7. Consider a sequence of random, locally finite rooted graphs (G_n,o_n). One says that (G_n,o_n) *converges locally, weakly* to a (random) locally finite rooted graph (G,o), denoted by $(G_n,o_n) \overset{\text{l.w.}}{\Rightarrow} (G,o)$ if for any $\rho \geq 0$ the neighbourhood $(G_n|_\rho, o_n)$ of o_n in (G_n,o_n) of radius ρ converges in distribution to the neighborhood of the same radius $(G|_\rho, o)$ of o in (G,o) modulo rooted graph isomorphism, i.e., for any realization (g,o) of $(G|_\rho,o)$

$$\lim_{n\to\infty} \mathbf{P}\{\, (G_n|_\rho, o_n) \simeq (g,o) \,\} = \mathbf{P}\{\, (G|_\rho, o) = (g,o) \,\}$$

for any $\rho \geq 0$.

For more reading on the local week convergence we refer to ((Benjamini, 2013, Section 5)).

Example 5.2.8 (Poisson Galton-Watson tree)**.** We have seen (mentioned) in previous lessons that the connected component of any given node o_n of the Erdős-Rényi graph $\mathrm{ER}_n(\lambda/n)$ with edge probability λ/n, rooted at this node, call this rooted component $(\mathrm{ER}_n(\lambda/n), o_n)$, converges to a Galton-Watson tree with Poisson offspring distribution of mean λ, rooted at its natural root o, $(\mathrm{GW}(\mathrm{Poi}(\lambda)), o)$;

$$(\mathrm{ER}_n(\lambda/n), o_n) \overset{\text{l.w.}}{\Rightarrow} (\mathrm{GW}(\mathrm{Poi}(\lambda)), o)\,.$$

[3] weak convergence on $\mathcal{G}^*$ as a Polish space with the metric specified in Footnote 1

By Proposition 5.2.6 any Galton-Watson tree with Poisson offspring distribution rooted at its natural root is unimodular.

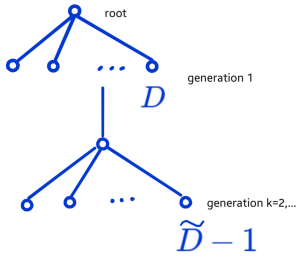
Example 5.2.9 (General unimodular Galton-Watson tree)**.** We have also argued in previous lessons that the connected component of the uniformly selected node o_n of the Configuration model CM_n with asymptotic node degree distribution $\{p_k\}$, rooted at this node $(\mathrm{CM}_n(\{p_k\}), o_n)$, convergences to an *inhomogeneous* Galton-Watson tree $(\mathrm{GW}(\{p_k\},\{\tilde{p}_k\}), o)$ rooted at its natural root o

$$(\mathrm{CM}_n(\{p_k\}), o_n) \stackrel{\text{l.w.}}{\Rightarrow} (\mathrm{GW}(\{p_k\},\{\tilde{p}_k\}), o)\,.$$

The root o of the limiting Galton-Watson tree has the offspring distribution $\{p_k\}$ while all other nodes have degree (offspring plus one) of *size-biased* distribution $\{\tilde{p}_k\}$ with

$$\tilde{p}_k = kp_k/\sum_i p_i\,. \tag{5.2.1}$$

By Proposition 5.2.6 any inhomogeneous Galton-Watson trees $(\mathrm{GW}(\{p_k\},\{\tilde{p}_k\}), o)$ with $\{p_k\}, \{\tilde{p}_k\}$ satisfying relation (5.2.1) is hence unimodular.



In what follows we shall see that the *condition* (5.2.1) *is also necessary* for an inhomogeneous Galton-Watson tree (with the root degree distribution possibly different from the degree distribution of all other nodes) to be unimodular (satisfy the MTP).

Denote by D the degree (offspring number) of the tree root and by $\tilde{D}$ the generic random variable of the degree of all other nodes. Thus $\tilde{D}-1$ is their offspring. Let $\{p_k\}$ be the distribution of D and denote the distribution of $\tilde{D}$ by $\{\tilde{p}_k\}$.

Consider a function

$$f(G,x,y) = \frac{1(d(x)=k)\mathbf{1}(x\sim y)}{k}\,.$$

The left-hand-side of (10.2.4) is equal to

$$\begin{aligned}\mathbf{E}\left[\sum_{v\in V} f(G,o,v)\right] &= \frac{1}{k}\mathbf{E}\left[\sum_{i=1}^{D}\mathbf{1}(D=k)\right]\\ &= \frac{1}{k}\mathbf{E}\left[D\mathbf{1}(D=k)\right]\\ &= \frac{1}{k}kp_k\end{aligned}$$

$$= p_k.$$

The right-hand-side of (10.2.4) is equal to

$$\mathbf{E}\left[\sum_{u\in V} f(G,u,o)\right] = \frac{1}{k}\mathbf{E}\left[\sum_{i=1}^{D}\mathbf{1}(\tilde{D}_i = k)\right]$$

where $\tilde{D}_i$ are the degrees of the children of the origin

$$= \frac{1}{k}\mathbf{E}\left[\sum_{n=1}^{\infty}\mathbf{1}(D=n)\sum_{i=1}^{n}\mathbf{1}(\tilde{D}_i = k)\right]$$

by the independence

$$= \frac{1}{k}\sum_{n=1}^{\infty} p_n \sum_{i=1}^{n}\tilde{p}_k$$

$$= \frac{\tilde{p}_k}{k}\sum_{n=1}^{\infty} n p_n$$

$$= \frac{\tilde{p}_k}{k}\mathbf{E}[D].$$

Consequently, by (10.2.4) $p_k = \tilde{p}_k \mathbf{E}[D]/k$ or, equivalently, $\tilde{p}_k = kp_k/\mathbf{E}[D]$, showing that this is a necessary relation for an inhomogeneous Galton-Watson tree to be unimodular.

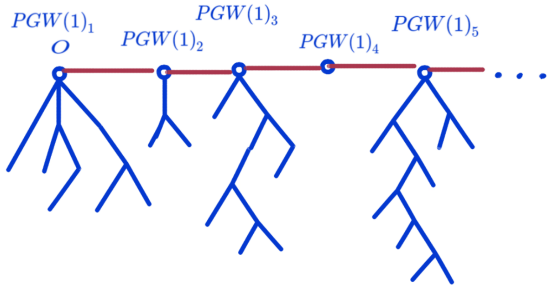
Here is one more interesting example of an infinite tree arising as a local weak limit of uniformly rooted graphs.

Example 5.2.10 (Skeleton tree or $\mathrm{PGW}^{\infty}(1)$)**.** Consider a sequence PGW_i of independent, critical Galton-Watson trees with offspring distribution having Poisson distribution of mean 1. As we know, any of such tree is almost surely finite. Connect by an edge the root of PGW_n and PGW_{n+1} for all $n \geq 1$. The resulting infinite, connected tree is called the skeleton tree or $\mathrm{PGW}^{\infty}(1)$ (with PGW(1) standing for the critical Poisson-Galton-Watson tree).

$\mathrm{PGW}^{\infty}(1)$ rooted at the root o of PGW_1 is a unimodular graph as a local weak limit

$$(T_n, o_n) \overset{\text{l.w.}}{\Rightarrow} (\mathrm{PGW}^{\infty}(1), o)$$

of trees (T_n, o_n) being uniformly distributed on all trees with n nodes, with arbitrary (equivalently, uniformly selected) roots o_n. This convergence result is known as the Grimmett's lemma; cf Grimmett ((1980)).



More examples can be found in Aldous and Steele ((2004)).

5.3 Some properties of the unimodular graphs

We state here two interesting properties of the unimodular graphs. We propose to prove them as exercises.

5.3.1 The typical node sees all configurations

First, we have a result saying that the root o of a unimodular graph (G, o) "sees" all configurations observable form all nodes of this graph. This is compatible with the interpretation of o as the typical node.

Proposition 5.3.1. *Let (G, o) by a unimodular graph. For any non-negative function $h(G, v)$ on $\mathcal{G}^*$, if $\mathbf{E}[h(G, o)] = 0$ then the probability that there exists a vertex v of G such that $h(G, v) > 0$ is equal to 0.*

5.3.2 Random walk on the unimodular graph

We consider now a random walk on the graph. Let (G, o) be unimodular graph with positive and finite mean root degree $0 < \mathbf{E}[d(o)] < \infty$. Consider another rooted graph $(\overline{G}, \overline{o})$ whose distribution is equal to the distribution of (G, o) biased by the degree of the root; i.e.,

$$\mathbf{E}\left[h(\overline{G}, \overline{o})\right] = \frac{\mathbf{E}[d(o)h(G, o)]}{\mathbf{E}[d(o)]} \tag{5.3.1}$$

for any non-negative function on $\mathcal{G}^*$. $(\overline{G}, \overline{o})$ is the called the stationary version of (G, o). Imagine a random walk on the nodes of (G, o), such that the next location is chosen uniformly among the neighbours of the current node. The following result says that the walker in the stationary state "sees"from its position graph $\overline{G}$ observed from $\overline{o}$. Moreover this walk is reversible in time. (The biasing by the degree $d(o)$ of the root in (5.3.1) says that the walker has larger chance to visit nodes with higher degree.)

Proposition 5.3.2. *Let (G, o) be given unimodular graph with $\mathbf{E}[d(o)] < \infty$ and $(\overline{G}, \overline{o})$ its stationary version (5.3.1). Let v_1 be a uniformly sampled neighbor of $\overline{o}$ in $(\overline{G}, \overline{o})$. Then:*

1. stationarity: *$(\overline{G}, \overline{o})$ and $(\overline{G}, v_1)$ have the same distribution on $\mathcal{G}^*$*
2. reversibility: *$(\overline{G}, \overline{o}, v_1)$ and $(\overline{G}, \overline{v}_1, \overline{o})$ have the same distribution on $\mathcal{G}^{**}$.*

5.4 Exercises

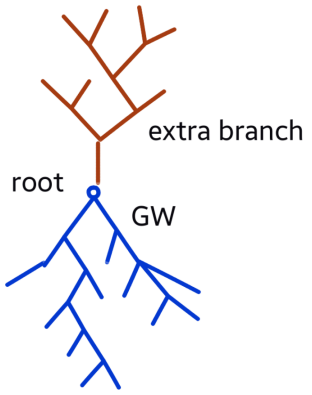
1. Prove that (G, o) is uniformly rooted iff (5.1.1) is satisfied.

2. Prove that a connected, symmetric graph G is transitive (Corollary 5.2.2).

3. Show that the d-regular tree rooted at an arbitrary node does not satisfy condition (10.2.4) for some functions $f(G, x, y)$ which are not isomorphism invariant.

4. An infinite graph G is called *amenable* if there is a sequence of finite subgraphs $(A_n)_{n\geq 1}$ of G such that
$$\lim_{n\to\infty} \frac{|\partial A_n|}{|A_n|} = 0\,,$$
where ∂A_n is the subset of vertexes of A_n which have neighbours in $G \setminus A_n$ and $|\cdot|$ denotes the number of vertexes. Such a sequence $(A_n)_{n\geq 1}$ is called a *Følner sequence*. The notion of uniformly rooted graphs could be in principle extended to amenable graphs with a suitable Følner sequence satisfying $\bigcup_n A_n = G$, by considering the local week limit of uniformly rooted subgraphs (A_n, o), provided this limit exists. This however still does not cover many interesting examples. Indeed the Galton-Watson tree with the offspring distribution $\{p_k\}$ is amenable iff $p_0 + p_1 > 0$ Forghani and Mallahi-Karai ((2017)). Prove the reverse part of this result by showing that *any infinite tree with all nodes of degree at least* 3 *is not amenable*.

5. *Homogeneous, augmented and unimodular Galton-Watson trees.* Consider a homogeneous Galton-Watson tree with a generic offspring variable D'. Assume for simplicity that $D' \geq 1$ almost surely, so that the Galton-Watson tree is infinite. Suppose we want to find in a naive way the typical node of this tree. The root of the this tree cannot be considered as the typical node. It is rather atypical since it is the only node which does not have an edge going "upward". Let us fix this problem by adding to the root one extra edge, going to an extra node, which starts an extra branch with the same offspring distribution D'. This is called an augmented Galton-Watson (AGW) tree, with the offspring variable D'. All nodes in this tree have the same degree distribution $D'+1$ and "look the same".



AGW

Since all nodes look the same, we could think that the original root of the AGW can be considered as its typical vertex. It is not true in general!

(a) Show that the AGW tree is unimodular iff D' is deterministic, i.e., the AGW is a regular tree. Hint: observe that the AGW is an inhomogeneous Galton-Watson tree, with the degree (offspring) of the root $D := D' + 1$ and the degree of other nodes $\tilde{D} := D' + 1$ (offspring D'). From Example 5.2.9 we know that such a Galton-Watson tree is unimodular iff D and $\tilde{D}$ satisfy relation $\mathbf{P}\{\tilde{D} = k\} = k\mathbf{P}\{ D = k \} / \mathbf{E}[D]$; cf (5.2.1). Show that this is satisfied iff D' is a constant.

(b) Show that in order to transform a homogeneous GW tree with offspring D' to a unimodular GW tree by changing its root degree, to say D'', one needs to take

$$\mathbf{P}\{ D'' = k \} := \frac{\mathbf{P}\{ D' = k - 1 \}}{k\mathbf{E}[1/(D' + 1)]} \quad k \geq 1 \,. \tag{5.4.1}$$

This distribution D'' can be interpreted as the distributing of the typical node the original GW tree.

(c) Observe that $\mathbf{E}[D''] = (\mathbf{E}[1/(D' + 1)])^{-1}$. Interpret this fact as that the mean degree of the typical node of a GW tree is equal to the *harmonic mean* of the original node degree.

6. Prove Proposition 5.3.1. Hint: For a given $\epsilon > 0$ and $n \geq 1$ consider a mass transport function $f(G, u, v) = 1(h(G, v) > \epsilon)\mathbf{1}(d(u, v) \leq n)$ where $d(o, v)$ is the graph distance between u and v. The probability that there exists a node v such that $h(G, v) > \epsilon$ within the graph distance n of o is not larger than $1/\epsilon\mathbf{E}\left[\sum_{v:d(v,o)\leq n} h(G, v)\right]$. Use the MTP to prove that this is equal to 0 for all $n \geq 1$ and $\epsilon > 0$.

7. Prove Proposition 5.3.2 using directly (10.2.4).

8. Let (G, o) be an infinite unimodular graph. Prove that the expected root degree satisfies $\mathbf{E}[d(o)] \geq 2$. Hint: Consider a function which sends mass equal to 1 from any vertex u all its neighbours v iff there is *exactly one* edge between u and v and the removal of this edge makes the component of u finite. Show that $D + S - R \geq 2$, where D, S, R is respectively, degree of o, and the mass sent and received by o. Use (10.2.4) to conclude.

Lesson 6

Erdős-Rényi graph — emergence of the connectivity

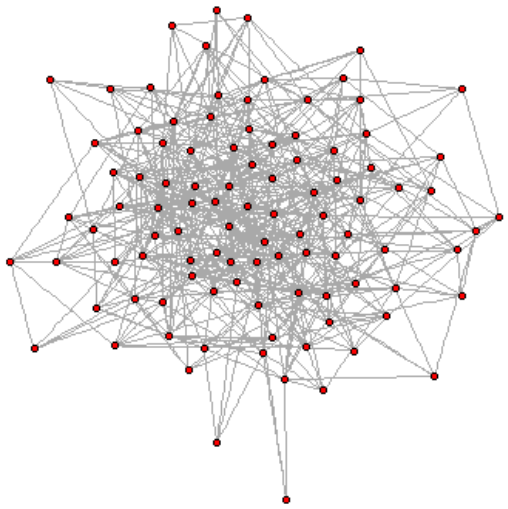


Figure 6.1: A realization of the Erdős-Rényi graph on $n = 100$ vertexes with edge probability $p = 0.0991$; mean vertex degree $np \approx 9.91 \gg 1$.

Recall that the Erdős-Rényi graph is a graph with independent edges. In the previous lesson on this model we have assumed the edge probability $p_n = \lambda/n$. With this parametrization we have observed the phase transition related to the emergence of the giant component when $\lambda > 1$. However, even with very large λ the whole graph is asymptotically connected with probability approaching 0. Indeed, the fraction of nodes in the largest connected component is asymptotically equal to $1 - p_{\text{ext}}(\lambda)$, hence strictly less than 1 even in the super-critical regime $\lambda > 1$. Indeed, recall that $p_{\text{ext}}(\lambda) > 0$ in the Galton-Watson tree with Poisson offspring distribution, since with positive probability zero offspring occurs. In this section we

will examine the regime p_n under which the ER is connected with probability approaching 1 for large n.

A first step consists in finding p_n for which the number of isolated nodes (of degree 0) converges to a finite variable. This latter happens to be a Poisson variable. We will prove this Poisson approximation using Stein-Chen method.

The next step consists in showing that when asymptotically there are no isolated nodes then the ER graph is also fully connected. Note that this is far from being evident, as the graph could be composed of several connected components of size strictly bigger than one (isolated) node. This is not the case. Intuitively we can understand that the giant component in this regime absorbers all other components.

6.1 Poisson approximation using Stein-Chen method

Let P_1, P_2 be two probability measures on some measurable space $(\Omega, \mathcal{F})$. We define *(total) variation distance* between P_1 and P_2 by

$$d_{\text{var}}(P_1, P_2) := 2 \sup_{A \in \mathcal{F}} |P_1(A) - P_2(A)| .$$

d_{var} is a metric on the space of probability measures on $(\Omega, \mathcal{F})$. Convergence in variation distance implies weak convergence:

Fact 6.1.1. *For $P_n, n \geq 1$ and P probability measures on some metric space with its Borel σ-algebra $(\Omega, \mathcal{F})$, if $\lim_{n\to\infty} d_{\text{var}}(P_n, P) \to 0$ then $P_n \Rightarrow P$ weakly as $n \to \infty$.*

Proof. Indeed, by the Portmanteau theorem $P_n \Rightarrow P$ iff $P_n(A) \to P(A)$ for all continuity sets $A \in \mathcal{F}$ of measure P [1]. For a different proof, based directly on the definition of the weak convergence see Exercise 1. □

6.1.1 Stein-Chen approximation

Stein-Chen equation

The Stein-Chen method is a general method in probability theory to obtain bounds on the distance between two probability distributions It was introduced by Charles Stein in 1970 for the study of the approximations by the Gaussian distribution and extended by his Ph.D. student Louis Chen Hsiao Yun in 1975 to obtain approximation results for the Poisson distribution. In what follows we present this extension.

Let N by a random variable taking values in $\{0, 1, \ldots\}$, with $0 < \mathbf{E}[N] < \infty$. Recall from Exercise 2 to the lesson on Configuration Model that N is Poisson random variable iff the distribution of $N + 1$ is equal to the distribution of the size biased version of N. In other words $\mathcal{L}_N = \text{Poi}_\lambda$ iff

$$\lambda \mathbf{E}[f(N+1)] = \mathbf{E}[f(N)N] \tag{6.1.1}$$

[1] $A \in \mathcal{F}$ is a continuity set of measure P if $P(\partial(A)) = 0$, where $\partial(A)$ is the boundary of A.

for some rich enough class of functions, say such that $\mathbf{E}\left[f(N)^2\right] < \infty$. The idea behind Stein-Chen (Chen's variant in this case) approximation can be summarized in the following way: When X is a random variable on $\{0,1,\ldots\}$ for which (6.1.1) holds approximately then the law of X should be close to Poi_λ. Moreover the variation distance between $\mathcal{L}_X$ and Poi_λ can be bounded using the difference between two expressions in (6.1.1).

In order to realize the above idea, for given $A \subset \{0,1,\ldots\}$ denote by $f_A(x)$ a function on $\{0,1,\ldots\}$ which satisfies the following equation called, *Stein-Chen equation*

$$\mathbf{1}(x \in A) - \mathrm{Poi}_\lambda(A) = \lambda f_A(x+1) - f_A(x)x\,, \tag{6.1.2}$$

where $\mathrm{Poi}_\lambda(A) = \sum_{x\in A} e^{-\lambda}\lambda^x/x!$. For all A equation (6.1.2) has a solution, which is given in Exercise 2. We do not need to know its exact form but only that this solution is Lipschitz

$$|f_A(x+1) - f_A(x)| \le \frac{1-e^{-\lambda}}{\lambda}, \tag{6.1.3}$$

for all A, x. Using function $f_A(x)$ we can express the difference between the two probabilities

$$\mathbf{P}\{\,X \in A\,\} - \mathrm{Poi}_\lambda(A) = \mathbf{E}\left[\mathbf{1}(X \in A) - \mathrm{Poi}_\lambda(A)\right] = \mathbf{E}\left[\lambda f_A(X+1) - f_A(X)X\right]$$

and consequently

$$\begin{aligned} d_{\mathrm{var}}(\mathcal{L}_X, \mathrm{Poi}_\lambda) &= 2\sup_{A\in\mathcal{F}} \left|\mathbf{P}\{\,X \in A\,\} - \mathrm{Poi}_\lambda(A)\right| \\ &= 2\sup_{A\in\mathcal{F}} \left|\mathbf{E}\left[\lambda f_A(X+1) - f_A(X)X\right]\right| \\ \boxed{\text{by (6.1.3)}} &\le 2 \sup_{f:\ |f(x+1)-f(x)|\le\frac{1-e^{-\lambda}}{\lambda}} \left|\mathbf{E}\left[\lambda f(X+1) - f(X)X\right]\right|. \end{aligned} \tag{6.1.4}$$

The right-hand side in (6.1.4) can be further bounded for a given variable X thus giving a bound on its variation distance to Poi_λ. Usually one takes $\lambda := \mathbf{E}[X]$.

6.1.2 Case of a sum of Bernoulli variables

We will give now a further, more explicit, bound for (6.1.4) assuming X is a sum of Bernoulli variables.

Let $I_i \in \{0,1\}$, $i \in V$, be a family of, possibly dependent, Bernoulli random variables. Denote $\pi_i := \mathbf{P}\{\,I_i = 1\,\} = \mathbf{E}[I_i]$. Let $X := \sum_{i\in V} I_i$ and denote $\lambda := \mathbf{E}[X] = \sum_{i\in V}\pi_i$.

Suppose there exists (on the same probability space on which we consider $I_i \in \{0,1\}$, $i \in V$) random variables J_{ij}, $i,j \in V$, $i \ne j$, such that the joint law of $\{J_{ij}, j \ne i\}$ is equal to the conditional law of $\{I_j, j \ne i\}$ given $I_i = 1$, for all $i \in V$

$$\mathcal{L}(\{I_j, j \in V, j \ne i\}\,|\,I_i = 1) = \mathcal{L}(\{J_{ij}, j \in V, j \ne i\})\,. \tag{6.1.5}$$

Proposition 6.1.2. *With the above notation*

$$d_{\mathrm{var}}(\mathcal{L}_X, \mathrm{Poi}_\lambda) \le 2\frac{1-e^{-\lambda}}{\lambda} \sum_{i\in V} \pi_i \Big(\pi_i + \sum_{\substack{j\in V\\ j\neq i}} \mathbf{E}\left[|I_j - J_{ij}|\right]\Big). \tag{6.1.6}$$

Proof. Let us evaluate the expression under $|\cdot|$ in the right-hand-side of (6.1.4). We have

$$\begin{aligned}
&\mathbf{E}\left[\lambda f(X+1) - f(X)X\right] \\
&= \sum_{i\in V} \pi_i \mathbf{E}\left[f\big(\sum_{j\in V} I_j + 1\big)\right] - \mathbf{E}\left[\sum_{i\in V} I_i f\big(\sum_{j\in V} I_j\big)\right] \\
&= \sum_{i\in V} \pi_i \mathbf{E}\left[f\big(\sum_{j\in V} I_j + 1\big)\right] - \sum_{i\in V} \pi_i \mathbf{E}\left[f\big(\sum_{\substack{j\in V\\ j\neq i}} I_j + 1\big) \,\middle|\, I_i = 1\right] \\
&\overset{\text{by (6.1.5)}}{=} \sum_{i\in V} \pi_i \mathbf{E}\left[f\big(\sum_{j\in V} I_j + 1\big)\right] - \sum_{i\in V} \pi_i \mathbf{E}\left[f\big(\sum_{\substack{j\in V\\ j\neq i}} J_{ij} + 1\big)\right] \\
&= \sum_{i\in V} \pi_i \mathbf{E}\left[f\big(\sum_{j\in V} I_j + 1\big) - f\big(\sum_{\substack{j\in V\\ j\neq i}} J_{ij} + 1\big)\right].
\end{aligned}$$

Consequently, by (6.1.4) we have

$$\begin{aligned}
&d_{\mathrm{var}}(\mathcal{L}_X, \mathrm{Poi}_\lambda) \\
&\qquad \le 2 \sup_{f:\ |f(x+1)-f(x)|\le \frac{1-e^{-\lambda}}{\lambda}} \sum_{i\in V} \pi_i \mathbf{E}\left[\left|f\big(\sum_{j\in V} I_j + 1\big) - f\big(\sum_{\substack{j\in V\\ j\neq i}} J_{ij} + 1\big)\right|\right] \\
&\boxed{|f(x)-f(y)| \le |x-y|\frac{1-e^{-\lambda}}{\lambda}} \le 2\frac{1-e^{-\lambda}}{\lambda} \sum_{i\in V} \pi_i \mathbf{E}\left[\left|\big(\sum_{j\in V} I_j + 1\big) - \big(\sum_{\substack{j\in V\\ j\neq i}} J_{ij} + 1\big)\right|\right] \\
&\qquad \le 2\frac{1-e^{-\lambda}}{\lambda} \sum_{i\in V} \pi_i \mathbf{E}\left[I_i + \sum_{\substack{j\in V\\ j\neq 1}} |I_j - J_{ij}|\right] \\
&\qquad = 2\frac{1-e^{-\lambda}}{\lambda} \sum_{i\in V} \pi_i \Big(\pi_i + \sum_{\substack{j\in V\\ j\neq i}} \mathbf{E}\left[|I_j - J_{ij}|\right]\Big).
\end{aligned}$$

□

6.2 Emergence of the connectivity in the Erdős-Rényi graph

6.2.1 Controlling isolated nodes

Denote by $I_i = I_i^n$ the indicator that the node i is isolated in ER_n

$$I_i := \prod_{\substack{j \in V_n \\ j \neq i}} (1 - \delta_{\{i,j\}}))\,.$$

and by

$$X = X_n := \sum_{i \in V_n} I_i$$

the number of such nodes in ER_n. Obviously $\pi_i := \mathbf{E}[I_i] = (1-p_n)^{n-1} =: \pi(n)$ and $\mathbf{E}[X_n] = n(1-p_n)^{n-1}$. The first step consists in finding p_n such that the expected number of isolated nodes $\mathbf{E}[X]$ converges to a finite positive constant.

Fact 6.2.1. *Assume*

$$p_n := \frac{\log n + c}{n}\,, \tag{6.2.1}$$

for some constant c, $-\infty < c < \infty$. Then

$$\lim_{n\to\infty} \mathbf{E}[X_n] = e^{-c}\,.$$

Proof. Note $p_n \to 0$ and for large n

$$n(1-p_n)^{n-1} = n((1-p_n)^{p_n^{-1}})^{p_n(n-1)} \approx ne^{-p_n n} = nn^{-1}e^{-c} = e^{-c}\,.$$

□

Next we prove that X_n is asymptotically Poisson.

Theorem 6.2.2. *Assume p_n is as in (6.2.1). Then $X_n \Rightarrow \mathrm{Poi}_\lambda$ in total variation, and hence weakly (in distribution) as $n \to \infty$.*

Proof. The convergence holds in variation metric (which implies weak convergence) and we shall use Proposition 6.1.2 to prove it. It is easy to see that

$$J_{ij} := \prod_{\substack{k \in V_n \\ k \neq i,j}} (1 - \delta_{\{k,j\}})$$

satisfies (6.1.5). Note

$$\mathbf{E}[|I_j - J_{ij}|] = \mathbf{E}\left[\delta_{\{i,j\}} \prod_{\substack{k \in V_n \\ k \neq i,j}} (1 - \delta_{\{k,j\}})\right]$$

$$= p_n(1-p_n)^{n-2}$$

$$= \pi(n)\frac{p_n}{1-p_n}\,.$$

Calculating the right-hand-side of (6.1.6) one gets

$$d_{\mathrm{var}}(\mathcal{L}_X, \mathrm{Poi}_{\mathbf{E}[X]})) \le 2\frac{1-e^{-\lambda}}{\lambda} n\pi(n)\left(\pi(n) + (n-1)\pi(n)\frac{p_n}{1-p_n}\right)$$

$$\boxed{\frac{1-e^{-\lambda}}{\lambda} \le 1 \text{ and } n\pi(n) = \mathbf{E}[X_n]} \le 2\mathbf{E}[X_n]\left(\pi(n) + \mathbf{E}[X_n]\frac{p_n}{1-p_n}\right) \to 0 \quad \text{as } n\to\infty,$$

since $\mathbf{E}[X_n] \to e^{-c}$, $\pi(n) \to 0$ and $p_n \to 0$. Finally, by the triangle inequality for d_{var}

$$d_{\mathrm{var}}(\mathcal{L}_X, \mathrm{Poi}_{e^{-c}}) \le d_{\mathrm{var}}(\mathcal{L}_X, \mathrm{Poi}_{\mathbf{E}[X]}) + d_{\mathrm{var}}(\mathrm{Poi}_{\mathbf{E}[X]}, \mathrm{Poi}_{e^{-c}})$$

and the result follows from the inequality

$$d_{\mathrm{var}}(\mathrm{Poi}_\lambda, \mathrm{Poi}_{\lambda'}) \le 2|\lambda - \lambda'|;$$

cf Exercise 3. □

6.2.2 Connectivity probability

Theorem 6.2.3. *Assume p_n is as in (6.2.1). Then*

$$\lim_{n\to\infty} \mathbf{P}\{\,\mathrm{ER}_n \textit{ is connected}\,\} = e^{-e^{-c}}\,.$$

Remark 6.2.4. Obviously connectivity implies that there are no isolated nodes, hence $\mathbf{P}\{\,\mathrm{ER}_n \text{ is connected}\,\} \le \mathbf{P}\{\,X_n = 0\,\}$ and consequently by Theorem 6.2.2

$$\lim_{n\to\infty} \mathbf{P}\{\,\mathrm{ER}_n \text{ is connected}\,\} \le \lim_{n\to\infty} \mathbf{P}\{\,X_n = 0\,\} = e^{-e^{-c}}\,.$$

Theorem 6.2.3 says that when p_n is as in (6.2.1) then asymptotically lack of isolated nodes is equivalent to full network connectivity.

Remark 6.2.5 (Full connectivity regime)**.** Assume that

$$p_n = \frac{\log n + c_n}{n} \quad \text{with } c_n \to \infty \text{ (any rate, even arbitrarily slow).}$$

Then by Theorem 6.2.3

$$\lim_{n\to\infty} \mathbf{P}\{\,\mathrm{ER}_n \text{ is connected}\,\} = 1\,.$$

Proof of Theorem 6.2.3. Following Remark 6.2.4 it is enough to prove that the probability of observing in ER_n connected components of size strictly larger than 1 goes to 0. We will treat separately components of size 2 and components of size r with $3 \le r \le \lceil n/2 \rceil$. (When there

is no component of size smaller or equal to $\lceil n/2 \rceil$ there is no components of size grater than $\lceil n/2 \rceil$ either.)

$$\begin{aligned}
\mathbf{P}\{\,\text{exists a component of size } 2\,\} &\le \mathbf{E}\,[\text{number of components of size } 2] \\
&= \binom{n}{2} p_n (1-p_n)^{2(n-2)} \\
&= \frac{n(n-1)}{2} \frac{p_n}{(1-p_n)^4} (1-p_n)^{2n} \\
\boxed{\text{using } 1-x \le e^{-x}} &\le \frac{n(n-1)}{2} \frac{p_n}{(1-p_n)^4} e^{-2np_n} \\
&= \frac{n(n-1)}{2} \frac{p_n}{(1-p_n)^4} n^{-2} e^{-2c_n} \\
&\approx \frac{1}{2} \frac{p_n}{(1-p_n)^4} e^{-2c_n} \to 0\,.
\end{aligned}$$

Regarding larger components, if there exists one of size r then (considering some tree spanning this component) ER_n contains a tree disconnected from other nodes (supplementary edges between the nodes of the tree are not excluded). There are $\binom{n}{r}$ ways of choosing r different nodes. There are r^{r-2} different trees on r distinguishable nodes; cf Exercise 4. For a given tree on r nodes, probability that this tree is a subset of ER_n and is disconnected from other nodes is equal to $p_n^{r-1}(1-p_n)^{r(n-r)}$. Putting this together

$$\begin{aligned}
\mathbf{P}\{\,\text{exists a component of size } r,\ 3 \le r \le \lceil n/2 \rceil\,\} &\le \sum_{r=3}^{\lceil n/2 \rceil} \binom{n}{r} r^{r-2} p_n^{r-1} (1-p_n)^{r(n-r)} \\
&\le \sum_{r=3}^{\lceil n/2 \rceil} \frac{n^r}{r!} r^{r-2} p_n^{r-1} (1-p_n)^{r(n-r)} \\
\boxed{r! \approx \sqrt{2\pi r}(r/e)^r \text{ (Stirling fromula) hence for large } n} &\le \sum_{r=3}^{\lceil n/2 \rceil} \frac{n^r}{\sqrt{r}} \Big(\frac{e}{r}\Big)^r r^{r-2} p_n^{r-1} (1-p_n)^{r(n-r)} \\
\boxed{n-r \ge \frac{n}{2}} &\le \frac{1}{p_n} \sum_{r=3}^{\lceil n/2 \rceil} r^{-5/2} e^{r(1+\log np_n - np_n/2)} \\
\boxed{\frac{1}{p_n} \le n, \text{ also } \log np_n - \frac{np_n}{2} \ge -(\frac{1}{2}-\epsilon)np_n \text{ for } \epsilon > 0 \text{ and large } n} &\le n \sum_{r=3}^{\lceil n/2 \rceil} r^{-5/2} e^{r(1/2-\epsilon)np_n} \\
&\le n \sum_{r=3}^{\infty} e^{-r(1/2-\epsilon)np_n}
\end{aligned}$$

$$= n\frac{e^{-3(1/2-\epsilon)np_n}}{1-e^{-(1/2-\epsilon)np_n}}$$

$$= O(n^{-1/2+3\epsilon}) \to 0.$$

□

TODO

In relation to the full connectivity:

- Discuss *small world* and *scale-free* networks.
- Sketch the *Barabási-Albert (BA) preferential attachment model.*

6.3 Exercises

1. Let P_1, P_2 be probability measures on some metric space $(\Omega, \mathcal{F})$, both absolutely continuous with respect to a (not necessarily probability) measure P ([2]). Prove that

$$d_{\text{var}}(P_1, P_2) = \int_\Omega \left| \frac{\mathrm{d}P_1}{\mathrm{d}P}(\omega) - \frac{\mathrm{d}P_2}{\mathrm{d}P}(\omega) \right| \mathcal{P}(\mathrm{d}\omega)\,. \tag{6.3.1}$$

   In case of a discrete space Ω, taking P to be the counting measure one obtains

$$d_{\text{var}}(P_1, P_2) = \sum_{\omega\in\Omega} \left| P_1(\omega) - P_2(\omega) \right|.$$

   Using (6.3.1) prove Fact 6.1.1.

2. Prove that

$$f_A(x+1) = \frac{\mathrm{Poi}_\lambda(A \cap \{0,\ldots,x\}) - \mathrm{Poi}_\lambda(A)\mathrm{Poi}_\lambda(\{0,1,\ldots,x\})}{\lambda \mathrm{Poi}_\lambda(\{x\})}$$

   is a solution of the Stein-Chen equation (6.1.2). Prove that $f_A(x)$ satisfies (6.1.3). *Hint:* Remark that

$$f_A(i+1) - f_A(i) = \sum_{j\in A} (f_{\{j\}}(i+1) - f_{\{j\}}(i)) \le (f_{\{i\}}(i+1) - f_{\{i\}}(i))\,,$$

   with the expression on the right-hand-side being the only positive term of the sum. Next, observe that

$$f_{\{i\}(i+1)} - f_{\{i\}}(i) \le \frac{1-e^{-\lambda}}{\lambda}$$

   and $f_A(i) = -f_{\bar{A}}(i)$ where $\bar{A}$ is the complement of A.

3. Show that $d_{\text{var}}(\mathrm{Poi}_\lambda, \mathrm{Poi}_{\lambda'}) \le 2(1-e^{-2|\lambda-\lambda'|}) \le 2|\lambda-\lambda'|$. *Hint*: For $\lambda > \lambda'$ represent Poi_λ random variable as a sum of two independent variables, having distributions $\mathrm{Poi}_{\lambda'}$ and $\mathrm{Poi}_{\lambda-\lambda'}$.

4. *Cayley's theorem:* There are exactly r^{r-2} different trees on r distinguishable nodes. Prove this result showing that the following function ϕ (called *Prüfer code*) is a bijection between the set of such trees and the set of words consisting of $r-2$ letters from r-element alphabet. For a tree T on nodes $\{0,1,\ldots,r-1\}$ let $\phi(T) = (w_1,\ldots,w_{r-2})$ be constructed in the following way:

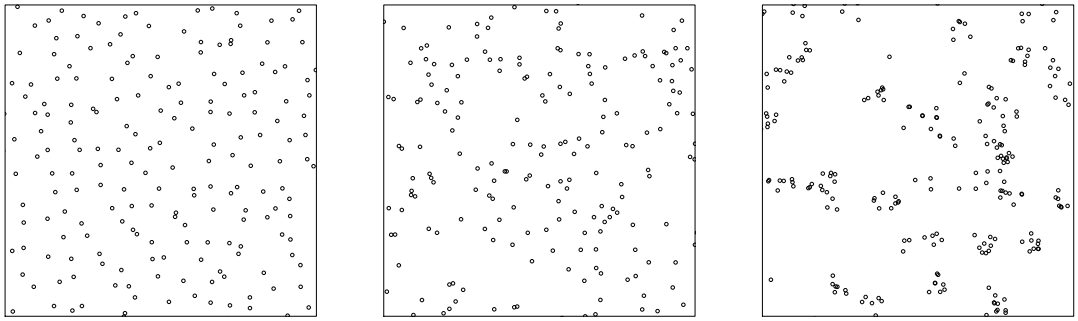
   Root the tree at the node 0. Leaves of this rooted tree $(T,0)$ are all nodes of degree 1 different than the root 0. Find the leaf with the smallest index: $v_1 := \min\{v > 0 : v \text{ is a leaf of } T\}$. Let w_1 be the (unique) node to which v_1 is connected.

   Remove v_1 from the tree T and repeat the previous step for $T \setminus \{v_1\}$. Continue until only one leaf remains, i.e., after having extracted $r-2$ nodes.

[2] I.e., $P_1(A) = P_2(A) = 0$ whenever $P(A) = 0$. In this case there exit *Radon-Nikodym derivatives* $\frac{\mathrm{d}P_i}{\mathrm{d}P}(\omega)$, $i = 1, 2$ i.e., measurable functions on $(\Omega, \mathcal{F})$ such that for all $A \in \mathcal{F}$, $P_i(A) = \int_A \frac{\mathrm{d}P_i}{\mathrm{d}P}(\omega)\,\mathcal{P}(\mathrm{d}\omega)$.

Lesson 7

Poisson point process



more regular pattern of repelling points | Poisson point process; independent points | more clustering pattern of attracting points

Figure 7.1: Realizations of three point processes with the same mean density of points and different dependence structures.

Informally, point process represents locations of a countable family of particles in some space. The best known example is the *Poisson point process*. Its points exhibit *complete independence* (to be defined formally), which greatly simplifies the study and makes it a good model for particles with no observed or expected interaction. This complete independence is tightly related to the *Poisson distribution of the number of points* in any given region, which is usually also postulated in the definition, and which explains the name of the process.

Both properties make the Poisson point process can be seen as a point process counterpart of the Erdős-Rényi graph and its (local-weak) limit — Galton-Watson tree with Poisson offspring distribution. We shall be motioning some analogies between them in the subsequent lessons.

For a supplementary material to this lesson see a working book project ((Baccelli et al., 2017, Chapters 2-3)). We refer also to the recent, very complete monograph Last and Penrose ((2017)) on Point processes, or more brief ((Daley and Vere-Jones, 2003, Chapter 2)),

((Baccelli and Błaszczyszyn, 2009, Chapter 1)).

7.1 Point processes

7.1.1 Framework

In the usual framework point process are represented using random counting measures.

Space of points Let $\mathbb{E}$ be a topological space, locally compact (i.e., every point has a compact neighbourhood), second countable (i.e., its topology has a countable base), Hausdorff (or separable; i.e., distinct points have disjoint neighbourhoods) abbreviated by LCSCH space. In particular, $\mathbb{E}$ is a *Polish* space, which means it is separable (i.e., there exits in $\mathbb{E}$ a countable, dense subset) and it admits a complete metric. [1] Also, $\mathbb{E}$ is σ-compact (i.e., it can be covered by countably many compact sets).

A standard example of LCSCH space is d-dimensional Euclidean space $\mathbb{R}^d$, with $1 \leq d < \infty$.

We consider the Borel σ-algebra $\mathcal{B}$ on $\mathbb{E}$, that is, the σ-algebra generated by the open sets of the topology. A set $B \in \mathcal{B}$ is called *(topologically) bounded* if it is relatively compact, i.e.; if its closure is compact. Denote by $\mathcal{B}_c$ all bounded (i.e., relatively compact) Borel subsets of $\mathbb{E}$.

Space of configurations of points In the theory of point processes, at most countable subset (called also configuration) of points $\{x_1, \ldots, x_J\} \subset \mathbb{E}$ is identified with the measure

$$\mu = \sum_{i=1}^{J} \delta_{x_i} \qquad J \in \{1, \ldots, \infty\}, \tag{7.1.1}$$

on $(\mathbb{E}, \mathcal{B})$, where δ_x is the *Dirac measure* at x taking values $\delta_x(B) = 1$ if $x \in B$ and 0 otherwise. Usually one requires the configuration of points to be *locally finite*, i.e; to have finite number of points in any bounded set $B \in \mathcal{B}_c$. Measures on $(\mathbb{E}, \mathcal{B})$ which are finite on all bounded Borel sets $\mathcal{B}_c$ are called *Radon measures* (or locally finite measures). A *counting measure* on $(\mathbb{E}, \mathcal{B})$ is a Radon measure that takes only non-negative integer values or infinity (this latter value possible only on unbounded sets). Denote by $\mathbb{M}$ be the set of all counting measures on $(\mathbb{E}, \mathcal{B})$. Any non-null measure $\mu \in \mathbb{M}$ can be expressed as in (7.1.1), where the $(x_i)_{i=1,\ldots,J}$ is a sequence of points of $\mathbb{E}$ without accumulation points (otherwise μ is not Radon), see Kallenberg ((1983)) for the proof of this and many other results stated in this document without proofs. Sometimes, less formally, we shall write $x \in \mu$ to say $\mu(\{x\}) \geq 1$.

Since we want to consider random configurations of points in $\mathbb{M}$ we need to define a σ-field on this space. Let $\mathcal{M}$ be the σ-algebra on $\mathbb{M}$ generated by the mappings $\mu \mapsto \mu(B), B \in \mathcal{B}$ (equivalently for all $B \in \mathcal{B}_c$), i.e.; the smallest σ-algebra making these mappings measurable.

[1] Such metric is not unique, each Polish space admits many complete metrics giving rise to the same topology. The presented point process framework is not based on any particular one.

A counting measure μ as in (7.1.1) is called *simple* if all x_i are distinct. A non-simple measure μ (corresponding to a configuration with multiple points) can be represented as $\mu = \sum_{i=1}^{J'} k_i \delta_{x'_i}$, with $k_i \in \{1, 2, \ldots\}$, where atoms x'_i are distinct. Moreover, we can choose the enumeration of the atoms x'_k in such a way that the mappings $\mu \mapsto x'_i$ and $\mu \mapsto k_i$ are measurable [2].

Example 7.1.1 (Canonical enumeration of points on $\mathbb{R}$)**.** In case $\mathbb{E} = \mathbb{R}$ we shall consider the following canonical enumeration of points of a simple counting measure μ:

$$\ldots < x_{-2} < x_{-1} < x_0 \leq 0 < x_1 < x_2 < \ldots .$$

As we shall see later, this particular enumeration of points has some important properties in Palm and ergodic theory of point processes. Lack of an analogue enumeration of points in $\mathbb{R}^d$, for $d > 1$, makes a difference between point processes on the line and in higher dimensions.

For μ as in (7.1.1) and a real function f consider the integral $\int f \mathrm{d}\mu := \int_{\mathbb{E}} f(x)\, \mu(\mathrm{d}x) = \sum_{i=1}^{J} f(x_i)$ (in case $J = \infty$ the sum needs to be well define). The mapping $\mu \mapsto \int f \mathrm{d}\mu$ is also measurable.

Point process A *point process* Φ is a measurable mapping form some probability space $(\Omega, \mathcal{A}, \mathbf{P})$ to the space of configurations of points $(\mathbb{M}, \mathcal{M})$. The distribution of Φ, is the probability measure $\mathbf{P}_\Phi$ on $(\mathbb{M}, \mathcal{M})$ being the image of $\mathbf{P}$ by Φ, i.e.; $\mathbf{P}_\Phi(\Gamma) = \mathbf{P}\{\, \Phi \in \Gamma \,\}$, for $\Gamma \in \mathcal{M}$.

Note that a point process Φ can be seen as a stochastic process $\Phi = \{\Phi(B)\}_{B \in \mathcal{B}_c}$ with the state space $\{0, 1, \ldots\} \ni \Phi(B)$ and where the index B runs over bounded Borel subsets $\mathcal{B}_c$ of $\mathbb{E}$. Hence, by the Kolmogorov's extension theorem the distribution of a point process is entirely characterized by the family of finite dimensional distributions $(\Phi(B_1), \ldots, \Phi(B_k))$, where $k \geq 1$ and $B_1, \ldots, B_k$ run over $\mathcal{B}_c$.

Point process Φ is *simple* if $\mathbf{P}\{\, \forall x \in \mathbb{E}, \Phi(\{x\}) \leq 1 \,\} = 1$; for the measurably cf Exercise 2. We say Φ has a *fixed atom* at x_0 if $\mathbf{P}\{\, \Phi(\{x_0\}) > 0 \,\} > 0$.

7.1.2 A few characteristics of point process

Mean measure $M = M_\Phi$ is a measure defined on $(\mathbb{E}, \mathcal{B})$ as $M(B) := \mathbf{E}\,[\Phi(B)]$. Note $M(B)$ is well defined for all $B \in \mathcal{B}$ but can be infinite even for $B \in \mathcal{B}_c$.

Void probability $\nu = \nu_\Phi$ is a set function defined on $(\mathbb{E}, \mathcal{B})$ as $\nu(B) := \mathbf{P}\{\, \Phi(B) = 0 \,\}$.

Fact 7.1.2 (Rényi's theorem)**.** *The probability distribution of a simple point process* Φ *is characterized by the family of its void probabilities* $\nu_\Phi(B)$ *for all* $B \in \mathcal{B}_c$.

[2] There exists a partition of $\mathbb{E}$ into sets $B_0, B_1, \ldots \in \mathcal{B}_c$. We may associate to each atom $x \in \mu \in \mathbb{M}$ a first index k such that $x \in B_k$. Regarding the finite number of atoms μ belonging to the same set B_k we enumerate them lexicographically according to their distances a fixed set $\{a_0, a_1, \ldots, \}$ dense in $\mathbb{E}$.

Proof. Sketch: one shows first that void probabilities allow one to express probabilities of the form $\pi_k(B_1, \ldots, B_k; B) := \mathbf{P}\{\, \Phi(B_1) > 0, \ldots, \Phi(B_k) > 0, \Phi(B) = 0 \,\}$ for all $k \geq 1$ and $B_i \in \mathcal{B}_c$. Next, finite dimensional distributions $\mathbf{P}\{\, \Phi(A_1) = n_1, \ldots, \Phi(A_l) = n_l \,\}$ are limits of some expressions involving $\pi_k(\cdot;\cdot)$ with $k \to \infty$ and the sets B_i's dissecting A_s's more and more precisely in a nested way. Since a realization of the point process is a locally finite measure at some level of precision in each set B_i there is at most one point of the point process; cf ((Kallenberg, 1983, Theorem 3.3)). □

Laplace functional (transform) $\mathcal{L} = \mathcal{L}_\Phi$ is a functional on the space of non-negative, measurable functions $f : \mathbb{E} \mapsto \mathbb{R}^+$ as $\mathcal{L}(f) := \mathbf{E}\left[e^{-\int f \mathrm{d}\Phi}\right]$. Its domain can be extended to functions f for which the expectation is well defined.

Fact 7.1.3. *The Laplace functional completely characterizes the distribution of the point process.*

Proof. For $f(x) = \sum_{i=1}^k t_i \mathbf{1}(x \in B_i)$, $\mathcal{L}_\Phi(f) = \mathbf{E}\left[e^{-\sum_i t_i \Phi(B_i)}\right]$, seen as a function of the vector $(t_1, \ldots, t_k)$, is the joint Laplace transform of the random vector $(\Phi(B_1), \ldots, \Phi(B_k))$, whose distribution is characterized by this transform. When $B_1, \ldots, B_k$ run over all bounded subsets of the space, one obtains a characterization of all finite-dimensional distributions of the point process. □

Remark 7.1.4. Comparison of void probabilities and (higher order) moment measures allows one to compare clustering properties of point processes. Smaller void probabilities and smaller moment measures indicate more regular point processes, cf Figure 7.1 and Błaszczyszyn and Yogeshwaran ((2014)).

7.1.3 Campbell's averaging formula

Here is very basic formula allowing one to evaluate expected values of integrals of deterministic functions with respect to point process.

Theorem 7.1.5 (Campbell's averaging formula)**.** *Let Φ be a point process on $\mathbb{E}$ with intensity measure M. Then for any measurable function $f : \mathbb{E} \to \mathbb{R}$ which is either non-negative or integrable with respect to M, the integral $\int_{\mathbb{E}} f \mathrm{d}\Phi$ is almost surely well defined and*

$$\mathbf{E}\left[\int_{\mathbb{E}} f(x)\, \Phi(\mathrm{d}x)\right] = \int_{\mathbb{E}} f(x)\, M(\mathrm{d}x)\,. \tag{7.1.2}$$

Proof. Consider first a simple function $f = \sum_{j=1}^k a_j \mathbf{1}_{B_j}$, where $a_j \geq 0$ and $B_j \in \mathcal{B}$. Then

$$\mathbf{E}\left[\int f \mathrm{d}\Phi\right] = \mathbf{E}\left[\sum_{j=1}^k a_j \Phi(B_j)\right]$$

$$= \sum_{j=1}^{k} a_j M(B_j) = \int f \mathrm{d}M$$

For a general non-negative function f consider an increasing sequence of simple functions converging to f and use the monotone convergence theorem. For f integrable with respect to M consider $f^+ := f\mathbf{1}(f \geq 0)$ and $f^- := -f\mathbf{1}(f < 0)$. □

Later, Palm theory will offer us an extension of the Campbell's averaging formula allowing one to consider expectations of the integrals of stochastic processes, i.e.; expressions $\mathbf{E}\left[\int f(x, \Phi)\, \Phi(\mathrm{d}x)\right]$.

7.2 Poisson point process

7.2.1 Definition and first properties

Let Λ be a (deterministic) Radon measure on $(\mathbb{E}, \mathcal{B})$.

Definition 7.2.1 (Poisson point process)**.** A point process Φ on $\mathbb{E}$ is a *Poisson point process of intensity (measure)* Λ if the following two conditions are satisfied:

1. For any $B \in \mathcal{B}_c$, $\Phi(B)$ is a Poisson random variable of intensity $\Lambda(B)$, i.e.; $\mathbf{P}\{\, \Phi(B) = n \,\} = e^{-\Lambda(B)}(\Lambda(B))^n/n!$.

2. For every $k = 1, 2, \ldots$ and all sets $B_i \in \mathcal{B}_c$, $i = 1, \ldots, k$, pairwise disjoint, random variables $(\Phi(B_1), \ldots, \Phi(B_k))$ are independent.

Clearly the above two conditions characterize finite dimensional distributions of a point process, provided it exists. Suppose for the moment that it does exist (we show construct it later) Here are a few observations made directly from the above definition.

Mean measure of Poisson point process is equal to its intensity measure. Indeed, $M(B) = \mathbf{E}\left[\Phi(B)\right] = \Lambda(B)$ by the first property of the above definition.

Void probability of Poisson point process is equal to $\nu(B) = \mathbf{P}\{\, \Phi(B) = 0 \,\} = e^{-\Lambda(B)}$, also by the first property of the above definition.

Laplace functional

Fact 7.2.2. *Laplace functional of Poisson point process on $\mathbb{E}$ of intensity Λ is equal to*

$$\mathcal{L}(f) = \mathbf{E}\left[e^{-\int_{\mathbb{E}} f \mathrm{d}\Phi}\right] = e^{-\int_{\mathbb{E}} (1 - e^{-f(x)})\, \Lambda(\mathrm{d}x)}. \tag{7.2.1}$$

Proof. Consider first a simple function $f = \sum_{j=1}^{k} a_j \mathbf{1}_{B_j}$, where $a_j \geq 0$ and $B_j \in \mathcal{B}_c$, which, without loss of generality, can be assumed pairwise disjoint. Then

$$\begin{aligned}
\mathcal{L}(f) &= \mathbf{E}\left[e^{-\int_{\mathbb{E}} f \mathrm{d}\Phi}\right] \\
&= \mathbf{E}\left[\prod_{j=1}^{k} e^{-a_j \Phi(B_j)}\right] \\
\boxed{\text{by the independence of } \Phi(B_j), j=1,\ldots,k} &= \prod_{j=1}^{k} \mathbf{E}\left[e^{-a_j \Phi(B_j)}\right] \\
\boxed{\text{by Poisson distribution of } \Phi(B_j)} &= \prod_{j=1}^{k} e^{-\Lambda(B_j)(1-e^{-a_j})} \\
&= e^{-\sum_{j=1}^{k} \Lambda(B_j)(1-e^{-a_j})} \\
& e^{-\int (1-e^{-f}) \mathrm{d}\Lambda}.
\end{aligned}$$

For a general function f consider an increasing sequence of simple functions converging to f and use the monotone convergence theorem. □

Conditional distribution of points given the number The following property can be proved directly from the definition of Poisson process.

Fact 7.2.3. *Consider $B_1, \ldots, B_k \in \mathcal{B}_c$ pariwise disjoint and denote $B := \sum_{i=1}^{n} B_k$. For all $n, n_1, \ldots, n_k \in \{0, 1, \ldots\}$ with $\sum_i n_i = n$,*

$$\mathbf{P}\{\, \Phi(B_1) = n_1, \ldots, \Phi(B_k) = n_k \mid \Phi(W) = n \,\} = \frac{n!}{n_1! \ldots n_k!} \frac{1}{\Lambda(B)^n} \prod_{i=1}^{k} \Lambda(B_i)^{n_i}. \tag{7.2.2}$$

We recognize in the above conditional distribution is a *multinomial distribution.*

Remark 7.2.4. *We can conclude form Fact 7.2.3 that given there are n points of the Poisson process in the window B, these points are independently and identically distributed (i.i.d.) in B according to the law $\frac{\Lambda(\cdot)}{\Lambda(B)}$.*

Construction of the Poisson point process Given a Radon measure Λ on $\mathbb{E}$ and $W \in \mathcal{B}_c$. Consider the following *independent* random objects $\{N, X_1, X_2, \ldots\}$, where

- N is a Poisson random variable with parameter $\Lambda(W)$,
- $X_1, X_2, \ldots$ are identically distributed random vectors (points) taking values in W with

$$\mathbf{P}\{\, X_1 \in \cdot \,\} = \Lambda(\cdot)/\Lambda(W). \tag{7.2.3}$$

Consider point process $\Phi = \sum_{k=1}^{N} \delta_{X_i}$. Using Laplace functional one can show that Φ is Poisson process of intensity $\Lambda|_W(\cdot) = \Lambda(\cdot \cap W)$, i.e.; Λ truncated to W. The same idea can be used to construct Poisson process on the whole space $\mathbb{E}$ provided $\Lambda(\mathbb{E}) < \infty$. The extension to the case of infinite total intensity can be done by considering a countable partition of $\mathbb{E}$ into bounded windows and an independent generation of the Poisson processes in each window; cf the paragraph on superposition of point processes in Section 7.2.4 and Exercise 10.

Simple Poisson point process Recall from Section 7.1.2, a point process is simple if with probability one it does not have multiple points. A fixed atom of a point process is a (deterministic) location in the space where it has an atom with positive probability. Being simple and not having a fixed atom are two different properties in general, except in case of a Poisson process.

Fact 7.2.5. *Let Φ be a Poisson process on $\mathbb{E}$ with intensity measure Λ.*

1. *Φ has a fixed atom at $x_0 \in \mathbb{E}$ iff Λ has an atom at x_0 (i.e., $\Lambda(\{x_0\}) > 0$).*
2. *Φ is simple iff Λ is non-atomic, i.e.; $\Lambda(\{x\}) = 0$ for all $x \in \mathbb{E}$.*

Proof. The prove of the first statement is straightforward from the definition. The second part can be proved using the conditional distribution of Poisson points. Consider a bounded subset $B \subset \mathcal{B}_c$.

$$\begin{aligned}
&\mathbf{P}\{\,\Phi \text{ has multiple points in } B\,\} \\
&= \sum_{n=2}^{\infty} \mathbf{P}\{\,\Phi(B) = n\,\}\, \mathbf{P}\{\,n \text{ points of } \Phi \text{ in } B \text{ are not all distinct} \mid \Phi(B) = n\,\} \\
&\overset{\text{by Remark 7.2.4}}{=} \sum_{n=2}^{\infty} \mathbf{P}\{\,\Phi(B) = n\,\}\, \frac{1}{(\Lambda(B))^n} \int_{B^n} \mathbf{1}(\exists_{i \neq j} x_i = x_j)\, \Lambda(\mathrm{d}x_1) \ldots \Lambda(\mathrm{d}x_n)\,.
\end{aligned}$$

Now

$$\begin{aligned}
\int_{B^n} \mathbf{1}(\exists_{i \neq j} x_i = x_j)\, \Lambda(\mathrm{d}x_1) \ldots \Lambda(\mathrm{d}x_n) &\leq \sum_{i<j=1}^{n} \int_{B^n} \mathbf{1}(x_i = x_j)\, \Lambda(\mathrm{d}x_1) \ldots \Lambda(\mathrm{d}x_n) \\
&\leq \sum_{i<j=1}^{n} (\Lambda(B))^{n-2} \int_B \Lambda(\{x_j\}) \Lambda(\mathrm{d}x_j) \\
&\overset{\Lambda(\{x_j\}) = 0 \text{ since } \Lambda \text{ is non-atomic}}{=} 0
\end{aligned}$$

We conclude the proof that $\mathbf{P}\{\,\Phi \text{ is not simple}\,\} = 0$ by considering an increasing sequence of bounded sets $B \nearrow \mathbb{E}$. □

Homogeneous Poisson process on $\mathbb{R}^d$ Poisson process of intensity $\Lambda(\mathrm{d}x) = \lambda \mathrm{d}x$ on $\mathbb{R}^d$, where λ $(0 < \lambda < \infty)$ is a constant, is called *homogeneous Poisson process of intensity* λ. From Fact 7.2.5 we know it is simple.

7.2.2 Strong Markov property of Poisson process

The strong Markov property of Poisson process extends the following property, which is a simple consequence of the independence property from its definition. Let Φ is a Poisson point process and consider a real measurable function f on $(\mathbb{M}, \mathcal{M})$. For any $B \in \mathcal{B}_c$

$$\mathbf{E}\left[f(\Phi)\right] = \mathbf{E}\left[f\Big(\Phi|_B + \Phi'|_{\mathbb{E}\setminus B}\Big)\right], \tag{7.2.4}$$

where Φ' is an *independent copy* of Φ and $\mu|_B$ denote the truncation of the measure $\mu|_B(\cdot) = \mu(\cdot \cap B)$. The strong Markov property says that the above statement hold when B is not necessarily constant but a random stopping set with respect to Φ. This latter notion can be formalized as follows.

Consider a general point process Φ on $\mathbb{E}$. We call $S \in \mathcal{B}_c$ a *random compact set* (with respect to Φ) when $S = S(\Phi)$ is a compact set that is a function of the realization of Φ. We give an example in Example 7.2.6. A random compact set $S(\Phi)$ is called a *stopping set* (with respect to Φ) if the event $\{\, S(\Phi) \subset K \,\}$ is $\Phi|_K$-measurable, i.e.; belongs to the σ-field generated by $\Phi|_K(B)$ for all $B \in \mathcal{B}_c$. In more simple words, $S(\Phi)$ is a stopping set if one can say whether the event $\{\, S(\Phi) \subset K \,\}$ holds or not knowing only the points of Φ in K.

Here is a very typical example of a stopping set.

Example 7.2.6 (k th smallest random ball)**.** For a point process Φ on $\mathbb{R}^d$, consider the random (closed) ball $B_0(R_k^*)$ centered at the origin, with the random radius equal to the k th smallest norm of $x_i \in \Phi$; i.e., $R_k^* = R_k^*(\Phi) = \min\{r \geq 0 : \Phi(B_0(r)) = k\}$. In order to prove that $B_0(R_k^*)$ is a stopping set let us perform the following mental experiment. Given a realization of Φ and a compact set K, let us start 'growing' a ball $B_0(r)$ centered at the origin, increasing its radius r from 0 until the moment when either (1) it accumulates k or more points or (2) it hits the complement K^c of K. If (1) happens, then $B_0(R_k^*) \subset K$. If (2) happens, then $B_0(R_k^*) \not\subset K$. In either case, we have not used any information about points of Φ in K^c; so $B_0(R_k^*) = B_0(R_k^*(\Phi))$ is a stopping set with respect to Φ.

The above example shows a very useful way to establish the stopping property: if there is a one-parameter sequence of growing compact sets which eventually leads to the construction of a random compact, then this compact is a stopping set.

The following result extends (7.2.4) to the case when B is a stopping set.

Fact 7.2.7 (Strong Markov property of Poisson point process)**.** *Let Φ be a Poisson point process and $S = S(\Phi)$ a random stopping set relative to Φ. Then the following holds*

$$\mathbf{E}\left[f(\Phi)\right] = \mathbf{E}\left[f\Big(\Phi|_{S(\Phi)} + \Phi'|_{\mathbb{E}\setminus S(\Phi)}\Big)\right], \tag{7.2.5}$$

Proof. The Poisson process is a Markov field indexed by $B \in \mathcal{B}_c$. By ((Rozanov, 1982, Theorem 4)) it has the strong Markov property with respect to all compact stopping sets. □

Construction of the homogeneous Poisson process on the line Consider homogeneous Poisson point process Φ of intensity λ on the real line $\mathbb{R}$ ($0 < \lambda < \infty$). Let us enumerate its points $\Phi = \sum_k \delta_{X_k}$ in the canonical way (cf Example 7.1.1). In particular, $X_1 = \sup\{x > 0 : \Phi((0,x)) = 0\}$ is the first atom of Φ in the open positive half-line $(0,\infty)$. We will show that $\{X_k\}$ can be constructed as a *renewal process with exponential holding times*, i.e., $X_k = \sum_{i=1}^{k} F_i$ for $k \geq 1$ and $X_k = -\sum_{i=k}^{0} F_i$ for $k \leq 0$, where $\{F_k : k = \ldots, -1, 0, 1 \ldots\}$ is a sequence of independent, identically distributed exponential random variables. Indeed,

$$\mathbf{P}\{\, F_1 > t \,\} = \mathbf{P}\{\, X_1 > t \,\} = \mathbf{P}\{\, \Phi((0,t]) = 0 \,\} = e^{-\lambda t}$$

so $X_1 = F_1$ is exponential random variable with parameter λ. By the strong Markov property for $k \geq 2$,

$$\begin{aligned} \mathbf{P}\{\, F_k > t \mid F_1, \ldots, F_{k-1} \,\} &= \mathbf{P}\{\, X_k - X_{k-1} > t \mid X_1, \ldots, X_{k-1} \,\} \\ \boxed{\text{By (7.2.5) with } S(\Phi) = [0, X_{k-1}]} &= \mathbf{P}\{\, \Phi((X_{k-1}, X_{k-1} + t]) = 0 \mid X_{k-1} \,\} \\ &= e^{-\lambda t} \end{aligned}$$

and similarly for $k \leq 0$, with $\{F_k\}_{k\leq 0}$ and $\{F_k\}_{k\geq 1}$ being independent. The above construction is specific for the dimension 1 and cannot be directly extended to a higher dimension. However, the Markov structure related to the complete independence property, which appeared in this example, can be observed in a general case.

Ordering of Poisson points according to their distance Let Φ be a Poisson point process of intensity Λ on $\mathbb{E}$. Consider some metric on $\mathbb{E}$ (cf Section 7.1.1) and let $\{R_k^* = R_k^*(\Phi)\}_{k\geq 1}$ be the sequence of the distances of the points of Φ from a fixed selected $x_0 \in \mathbb{E}$ arranged in increasing order (i.e. R_k^* is the distance of the k-th nearest point of Φ to x_0). We tacitly assume that these points are defined uniquely [3]. One can conclude from the strong Markov property of the Poisson point process that this sequence is a Markov chain with transition probability

$$\mathbf{P}\{\; R_k^* > t \mid R_{k-1}^* = s \,\} = \begin{cases} e^{-\Lambda(B_0(t)) - \Lambda(B_0(s))} & \text{if } t > s \\ 1 & \text{if } t \leq s\,. \end{cases} \tag{7.2.6}$$

7.2.3 Equivalent characterizations of Poisson process

We shall show under what mild assumptions each of the two conditions of the definition of Poisson point process (Poisson distribution of the number of points and independence) alone characterizes the Poisson process.

[3] This is the case e.g. when the intensity measure Λ of Φ is null on every sphere $\{x \in \mathbb{E} : d(x_0 - x) = r\}$ $r > 0$, where d is the metric on $\mathbb{E}$.

Characterizations by the form of the distribution We shall give now without proof the following result.

Proposition 7.2.8. *Suppose that Φ is a simple point process Then Φ is a Poisson point process iff there exists a Radon non-atomic measure Λ such that for any subset $B \in \mathcal{B}_c$, $\mathbf{P}\{\,\Phi(B) = 0\,\} = e^{-\Lambda(B)}$.*

This is a consequence of the Rényi's theorem (cf Fact 7.1.2). In particular:

Corollary 7.2.9. *Φ is a Poisson process provided it is simple and all marginal distributions $\Phi(B)$ for $B \in \mathcal{B}_c$ are Poisson.*

The assumption that Φ is simple cannot be relaxed since one can construct two Poisson random variables N_1 and N_2, of parameters μ_1, μ_2, respectively, and such that $N_1 + N_2$ is Poisson of parameter $\mu_1 + \mu_2$, with N_1 and N_2 *not* being independent, cf ((Stoyanov, 2013, Section 12.3)).

Characterization by complete independence One says that the point process Φ has the property of *complete independence* if for any finite family of subsets $B_1, \ldots, B_k \in \mathcal{B}_c$ that are mutually disjoint, the random variables $\Phi(B_1), \ldots, \Phi(B_k)$ are independent.

Proposition 7.2.10. *Suppose that Φ is a point process without fixed atoms. Then Φ is a Poisson point process iff Φ is simple and has the property of complete independence.*

The necessity follows from Fact 7.2.5. For sufficiency, one shows that the measure $\Lambda(A) = -\log(\mathbf{P}\{\,\Phi(A) = 0\,\})$ satisfies the assumptions of Proposition 7.2.8. (cf. ((Kallenberg, 1983, Section 2.1))).

Again, the assumption that Φ has no fixed atoms cannot be relaxed as adding a deterministic atom to a simple Poisson process preserves the complete independence but destroys Poisson distribution of the number of points in any B covering x_0.

7.2.4 Operations preserving the Poisson law

Superposition The superposition of point processes Φ_k is defined as the sum $\Phi = \sum_k \Phi_k$. Note that the summation in the above definition is understood as the summation of (point) measures. It always defines a point measure, which however, in general, might not be locally finite (we do not assume the last sum to have finitely many terms). A crude, but useful sufficient condition for this to not happen is that the sum of mean measures $\sum_k \mathbf{E}\,[\Phi_k(\cdot)]$ is a locally finite measure. A refined sufficient condition may be found by the Borel–Cantelli lemma.

Fact 7.2.11. *The superposition of independent Poisson point processes with intensities Λ_k is a Poisson point process with intensity measure $\sum_k \Lambda_k$ iff the latter is a locally finite measure.*

Thinning A thinning of a point process Φ consists in independent removing of some points of Φ. Formally, consider a measurable function $p : \mathbb{E} \mapsto [0,1]$ and a point process Φ on $\mathbb{E}$. The *thinning* of $\Phi = \sum_k \delta_{X_k}$ with the *retention function* p is a point process given by

$$\Phi^p = \sum_k I_k \delta_{X_k} , \tag{7.2.7}$$

where the random variables $\{I_k\}_k$ are independent given Φ, and $\mathbf{P}\{\, I_k = 1 \mid \Phi \,\} = 1 - \mathbf{P}\{\, I_k = 0 \mid \Phi \,\} = p(X_k)$. It is not difficult to verify that the above construction transforms a Poisson point process into another Poisson point process.

Fact 7.2.12. *The thinning of the Poisson point process of intensity measure Λ with the retention probability p yields a Poisson point process of intensity measure $p\Lambda$ with $(p\Lambda)(B) = \int_B p(x)\,\Lambda(\mathrm{d}x)$.*

In particular, the truncation $\Phi|_W$ of a Poisson point process of intensity measure Λ to some given set W is a Poisson point process with intensity measure $\Lambda(\cdot \cap W) = \Lambda|_W(\cdots)$.

Random Transformation of points It consists in random, independent displacing each point of a point process to some new location, possibly in a different space, according to some probability kernel p.

Consider a probability kernel $p(x, B)$ from $(\mathbb{E}, \mathcal{B})$ to some LCSCH space $(\mathbb{E}', \mathcal{B}')$, i.e. for all $x \in \mathbb{E}$, $p(x, \cdot)$ is a probability measure on $\mathbb{E}'$ and for all $B' \in \mathcal{B}'$ the function $p(\cdot, B')$ is measurable. The transformation Φ^p of a point process $\Phi = \sum_k \delta_{X_k}$ by a probability kernel $p(\cdot, \cdot)$ is a point process on $\mathbb{E}'$ given by

$$\Phi^p = \sum_k \delta_{Y_k} , \tag{7.2.8}$$

where the $\mathbb{E}'$-valued random vectors $\{Y_k\}_k$ are independent given Φ, with $\mathbf{P}\{\, Y_k \in B' \mid \Phi \,\} = p(X_k, B')$.[4] W tacitly assume that Φ^p is a Radon measure.

This operation preserves the Poisson point process property as stated in the following theorem.

Fact 7.2.13 (Displacement Theorem)**.** *The transformation of the Poisson point process of intensity measure Λ by a probability kernel p is the Poisson point process with intensity measure $\Lambda'(B') = \int_{\mathbb{E}} p(x, B)\,\Lambda(\mathrm{d}x)$, $B' \subset \mathbb{E}'$, provided Λ' is a Radon measure.*

In particular, consider a Poisson process at time $t = 0$ whose points move in time following independent Markov processes. Then the configuration of points at any time t has Poisson distribution.

Another consequence is that an deterministic mapping of Poisson points (e.g. translation and/or rescaling in $\mathbb{R}^d$) remains Poisson process.

[4] We use the same notation Φ^p for the p-thinning and the transformation by kernel p. The context indicates what is meant.

TODO

- Introduce higher-order momgent measures.

7.3 Exercises

1. For the canonical enumeration of points on the line given in Example 7.1.1 show that the mapping $\mu \mapsto x_i$ is $(\mathbb{M}, \mathcal{M}) \to (\mathbb{R}, \mathcal{B})$ measurable.

2. Show that $\{\forall x \in \mathbb{E}, \Phi(\{x\}) \leq 1\} \in \mathcal{M}$.

3. Prove Fact 7.2.3,

4. Prove that Φ given by (7.2.3) is Poisson process.

5. Construct a point process which is simple and has a fixed atom. Construct also a point process which is not simple but does not have fixed atoms.

6. Prove the first statement of Fact 7.2.5.

7. Show that if $S = S(\Phi)$ is a stopping set with respect to Φ, then for any $\mu \in \mathbb{M}$,

$$S(\Phi) = S(\Phi|_{S(\Phi)} + \mu|_{\mathbb{E}\setminus S(\Phi)}) .$$

   In other words, all modifications of Φ outside the set $S(\Phi)$ have no effect on $S(\Phi)$.

8. Describe the stopping sets allowing to conclude (7.2.6) from Fact 7.2.7.

9. Prove that the superposition of independent Poisson processes is a Poisson process as stated in Fact 7.2.11. Construct a counterexample showing that independence is required.

10. Using Fact 7.2.11 construct Poisson point process on $\mathbb{E}$ with intensity Λ, such that $\Lambda(\mathbb{E}) = \infty$. Use the fact that every Radon measure on $\mathbb{E}$ is σ-finite.

11. Prove that any independent thinning of a Poisson process is some Poisson process as stated in Fact 7.2.12. *Hint:* Calculate the Laplace functional.

12. Prove the Displacement Theorem for Poisson process; cf Fact 7.2.13.

13. *Construction of an inhomogeneous Poisson process on the line.* Let Ψ be a unit-intensity homogeneous Poisson point process on $\mathbb{R}^2$. For a given non-negative function $\lambda(x)$ on $\mathbb{R}$ such that $\int_{\mathbb{R}} f(x)\,\mathrm{d}x < \infty$, one constructs a point process on the line $\mathbb{R}$ by projecting the atoms of Ψ which are below the curve $\lambda(x)$ on the x axis. Using Displacement Theorem (Fact 7.2.13) show that Φ is a Poisson process of intensity $\Lambda(\mathrm{d}x) = \lambda(x)\,\mathrm{d}x$.

14. Show that the Laplace functional of the independent thinning Φ^p of a general point process Φ with Laplace functional $\mathcal{L}_\Phi$ is equal to $\mathcal{L}_{\Phi^p}(f) = \mathcal{L}_\Phi(\tilde{f})$ where $\tilde{f}(x) = -\log\left[1 - p(x)\left(1 - e^{-f(x)}\right)\right]$. This relation of $\mathcal{L}_{\Phi^p}$ to $\mathcal{L}_\Phi$ can be used to define the independent thinning as an operation on the probability distributions of point processes.

15. Show that the Laplace functional of the transformation Φ^p of a general point process Φ with Laplace functional $\mathcal{L}_\Phi$, by the probability kernel p is equal to $\mathcal{L}_{\Phi^p}(f) = \mathcal{L}_\Phi(\tilde{f})$ where
$$\tilde{f}(x) = -\log\left[\int_{\mathbb{E}'} e^{-f(x)} p(x, \mathrm{d}y)\right].$$
This relation of $\mathcal{L}_{\Phi^p}$ to $\mathcal{L}_\Phi$ can be used to define the transformation by the kernel as an operation on the probability distributions of point processes.

16. *Computer exercise with "spatstat".* *Spatstat*, `http://spatstat.org/`, an R package for spatial statistics with a strong focus on the analysis of spatial point patterns. For the package reference guide see Package ((2017)).

    (a) Generate a homogeneous Poisson point pattern with intensity $\lambda = 1$ in a square observation window $[0, 10]^2$ using
    `pp <- rpoispp(1, win=owin(c(0,10),c(0,10)))`

    (b) Plot your point process calling `plot(pp)`

    (c) Count the number of points in the window with `npoints(pp)` What is the expected number of points in the window?

    (d) Estimate and plot the density of the generated point pattern with
    ```
    L <- density.ppp(pp, 0.05)
    plot(L)
    ```
    This is the density of the intensity measure of your process, and should be close to 1 in your example.

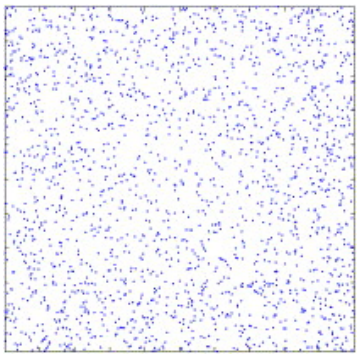
    (e) Repeat the same with non-homogeneous Poisson process of the intensity $\Lambda(d(x, y)) = 100e^{-3x}\mathrm{d}x\mathrm{d}y$ generated with
    `npp <- rpoispp(function(x,y) 100 * exp(-3*x), 100)`

    (f) As two examples of non-Poisson patterns, generate, and plot *Strauss process* `rStrauss`, exhibiting point repulsion (see ((Daley and Vere-Jones, 2003, Example 7.1(c)))) and the *Neyman-Scott* process `rNeymanScott`, which is a cluster process (see ((Daley and Vere-Jones, 2003, Example 6.3(a)))).

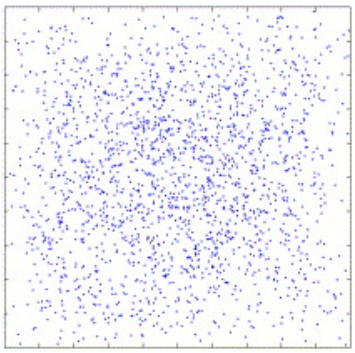
    (g) Choose some data set of point patterns available in *spatstat* (cf Baddeley et al.). Plot the data and estimate the density of points.

Lesson 8

Point conditioning and Palm theory for point processes



conditional given a waypoint



unconditional

Figure 8.1: Illustration of the difference between conditional and unconditional distribution of particles in a random *waypoint* mobility model. In this (space-time) particle mobility model a particle independently chooses its successive locations (called *waypoints*) uniformly in the given window and moves straight to them with constant speed. The locations of the particle observed when it was at some waypoint form hence a homogeneous Poisson process (left image). The unconditional distribution of particles, observed at an arbitrary time, concentrates more in the middle of the window (right image), where the particles are "most of the time". Pictures borrowed from Le Boudec ((2007)). See also Exercise 12.

Palm distributions of simple point process Φ are supposed to play the role of the conditional distributions of Φ, given it has an atom at x

$$\mathbf{P}\{\, \Phi \in \Gamma \,|\, x \in \Phi \,\} \,. \tag{8.0.1}$$

They were named after a Swedish electrical engineer and statistician, Conny Palm, who studied in telecommunication the conditional probabilities given observation of a specific event

and their relations to unconditional probabilities [1]. Phenomena of the same nature can be nicely observed e.g. in some random particle mobility models, see Figure 8.1.

However, when x is not a fixed atom of Φ, i.e., $\mathbf{P}\{\, x \in \Phi \,\} = 0$, the simple definition (8.0.1) of the conditional distribution makes no sense. For this reason, a modern, general approach to Palm distributions, proposed by a Polish mathematician Czesław Ryll-Nardzewski, is based on Radon-Nikodym's theorem. For a supplementary material to this lesson see a working book project ((Baccelli et al., 2017, Chapters 4)).

As the main result of this lesson we shall prove that conditional distributions (8.0.1), or more formally the Palm distributions, of a Poisson point process correspond to the distribution of the original process Φ with the conditioning point x just added to Φ. This is not a very surprising fact in view of the complete independence property of the Poisson process and, as this property, it also characterizes Poisson process. This new characterization is named after a Russian and German mathematicians I.M. Slivnyak and Joseph Mecke.

8.1 Palm distributions

8.1.1 Campbell's measure and Palm distributions

Let Φ be a point process on LCSCH space $\mathbb{E}$ with Radon (locally finite) mean measure $M = M_\Phi$. Define the following measure on $(\mathbb{E} \times \mathcal{B}, \mathcal{B} \otimes \mathcal{M})$ called *Campbell's measure* [2] of Φ, by specifying its values on rectangles

$$C(B \times \Gamma) := \mathbf{E}\left[\int_{\mathbb{E}} \mathbf{1}(x \in B)\mathbf{1}(\Phi \in \Gamma)\, \Phi(\mathrm{d}x)\right] \qquad B \in \mathcal{B}, \Gamma \in \mathcal{M}\,.$$

It is easy to see that for each $\Gamma \in \mathcal{M}$, the projection of the Campbell's measure C on $(\mathbb{E}, \mathcal{B})$ is absolutely continuous with respect to the mean measure M of Φ. Indeed,

$$C(B \times \Gamma) \le \mathbf{E}\left[\int_{\mathbb{E}} \mathbf{1}(x \in B)\, \Phi(\mathrm{d}x)\right] = M(B) \quad \text{for all } B \in \mathcal{B}.$$

Hence, by the Radon-Nikodym's theorem, still for any fixed $\Gamma \in \mathcal{M}$, there exist a non-negative, measurable function $P_x(\Gamma)$ on $(\mathbb{E}, \mathcal{B})$ (remark, x is the argument and Γ is a parameter) such that for all $B \in \mathcal{B}$

$$C(B \times \Gamma) = \int_{\mathbb{E}} \mathbf{1}(x \in B) P_x(\Gamma)\, M(\mathrm{d}x)\,. \tag{8.1.1}$$

The function $P_x(\Gamma)$ (of x, for a given Γ) is simply a version of the Radon-Nikodym's (R-N) derivative of the projection of the Campbell's measure $C(\cdot \times \Gamma)$ with respect to the mean

[1] Conny Palm was among the pioneers of the theory of point processes, closely related at that time to queueing theory. The term "point process" was used for the first time in 1940 in a paper by William Feller, and the PhD dissertations of Ove Lundberg and Conny Palm, who were using German word "Punktprozesse".

[2] Norman Robert Campbell (1880–1949) was an English physicist and philosopher of science, who introduced in 1909 the averaging formula called after his name (see the Campbell's averaging formula in the previous lesson).

measure $M(\cdot)$. As usually in the case of version of the R-N derivative, it is not uniquely defined. Different versions of $P_x(\Gamma)$ may differ on a set of x's of M-measure null. Topological properties of the space $\mathbb{E}$ allow one to choose these versions for different $\Gamma \in \mathcal{M}$ in such a way that, when $x \in \mathbb{E}$ is fixed, $P_x(\cdot)$ is a probability distribution on $(\mathbb{M}, \mathcal{M})$. In this case $P_\cdot(\cdot)$ is a probability kernel from $(\mathbb{E}, \mathcal{B})$ to $(\mathbb{M}, \mathcal{M})$ and distributions $P_x(\cdot)$, $x \in \mathbb{E}$ are called *Palm distributions* of Φ. Note, that we may still have different versions of Palm distributions. In particular one can arbitrarily modify given distributions $P_x(\cdot)$ on a sets of M-measure null.

Remark 8.1.1. Since for M-almost all $x \in \mathbb{E}$ $P_x(\cdot)$ is a probability measure on $\mathbb{M}$, it can be identified with a probability distribution P_{Φ_x} of some point process Φ_x often called *Palm version* of Φ at x. It can be assumed to be defined on a suitable extension of the probability space $(\Omega, \mathcal{A})$ on which Φ is defined. [3] For this process we have $P_{\Phi_x} = P_x$, that is $\mathbf{P}\{\, \Phi_x \in \Gamma \,\} = P_x(\Gamma)$, with $\mathbf{P}\{\, \cdot \,\}$ being the probability on the aforementioned extension of the probability space $(\Omega, \mathcal{A})$. In most cases Φ_x are merely used to simply the notation, as for example in $P_x(\{\mu : x \in \mu\}) = \mathbf{P}\{\, x \in \Phi_x \,\}$ or $\int_{\mathbb{M}} f(\mu)\, P_x(\mathrm{d}\mu) = \mathbf{E}\,[f(\Phi_x)]$. We shall try not to use Φ_x unless really hard to avoid.

Remark 8.1.2. We have said in the introduction to this lesson that Palm distributions are supposed to play the role of the conditional distributions of Φ as in (8.0.1). The reality is slightly more subtle. For x being a fixed atom of Φ (i.e., $\mathbf{P}\{\, \Phi(\{x\}) > 0 \,\} > 0$)

$$P_x(\Gamma) = \frac{\mathbf{E}\,[\mathbf{1}(\Phi \in \Gamma)\Phi(\{x\})]}{\mathbf{E}\,[\Phi(\{x\})]} \qquad \Gamma \in \mathcal{M} \tag{8.1.2}$$

(prove it). This latter is the distribution of Φ biased by the mass of Φ at $\{x\}$; i.e., $\Phi(\{x\})$ (yet another size biasing!). In case x is a *simple fixed atom*, i.e., $\mathbf{P}\{\, \Phi(\{x\}) = 1 \,\} = 1 - \mathbf{P}\{\, \Phi(\{x\}) = 0 \,\}$ the right hand side of (8.1.2) further reduces to the conditional probability (8.0.1), given Φ has a point at x. More generally, the conditional probability interpretation of Palm distributions is valid for simple point processes and should be replaced by the above size biasing interpretation in case of non-simple point processes.

Palm distributions are defined so as to satisfy the following key result of Palm theory, which can serve as en their equivalent definition. This is also an extension of Campbell's averaging formula (cf Theorem 1.4 in the Lesson on Poisson Point Process) to the case when f is a stochastic process (i.e., it depends on Φ).

Theorem 8.1.3 (Campbell-Little-Mecke (CLM))**.** *Let Φ be a Point process on $\mathbb{E}$ with Radon mean measure M and Palm distributions P_x, $x \in \mathbb{E}$. For all non-negative measurable functions f on $\mathbb{E} \times \mathbb{M}$ we have*

$$\mathbf{E}\left[\int_{\mathbb{E}} f(x, \Phi)\, \Phi(\mathrm{d}x)\right] = \int_{\mathbb{E}} \int_{\mathbb{M}} f(x, \mu) P_x(\mathrm{d}\mu) M(\mathrm{d}x). \tag{8.1.3}$$

[3] In contrast to the canonical space $(\mathbb{M}, \mathcal{M})$ we do not have any topological properties of $(\Omega, \mathcal{A})$. In some cases, all Φ_x can be constructed directly on $(\Omega, \mathcal{A})$. This is the case e.g. for any Poisson point process, cf. Slivnyak-Mecke Theorom 8.1.8 or a general point process defined in the stationary framework, as we shall explain in the Lesson on Stationary Point Processes.

The result extends to all functions f for which either of the two sides of the equality (8.1.3) is finite when f is replaced by $|f|$.

Note that the right hand side of (8.1.3) can be rewritten using the Palm versions Φ_x of Φ as $\int_{\mathbb{E}} \mathbf{E}\left[f(x,\Phi_x)\right] M(\mathrm{d}x)$ with all the precautions on the usage of Φ_x and $\mathbf{E}\left[\cdot\right]$ discussed in Remark 8.1.1.

Proof of Theorem 8.1.3. Directly from the definition of P_x as a version of the R-N derivative one easily shows the desired equality for $f(x,\mu) = \mathbf{1}(x \in B, \mu \in \Gamma)$ where $B \in \mathcal{B}, \Gamma \in \mathcal{M}$

$$\mathbf{E}\left[\int_{\mathbb{E}} f(x,\Phi)\Phi(\mathrm{d}x)\right] = C(B\times\Gamma) = \int_{\mathbb{E}}\int_{\mathbb{M}} f(x,\mu)P_x(\mathrm{d}\mu)M(\mathrm{d}x)\,.$$

By (rather standard) measure theoretic approximation arguments the result can be extended first to the case of $f(x,\mu) = \mathbf{1}((x,\mu)\in L)$, for arbitrary $L \in \mathcal{B}\otimes\mathcal{M}$, next to any simple function on $\mathbb{E}\times\mathbb{M}$ and finally to a general function whose integrals of the positive and negative part are finite. □

Corollary 8.1.4. *Under assumptions of Theorem 8.1.3, for M-almost all $x \in \mathbb{E}$, $P_x(\{\mu : x \in \mu\}) = 1$, i.e., x is almost surely an atom of the point process Φ_x whose distribution is P_x.*

Proof. Indeed, for any $B \in \mathcal{B}_c$ we have

$$\begin{aligned} 0 &\le \int_{\mathbb{E}} \mathbf{1}(x\in B)(1 - P_x(\{\mu : x\in\mu\})\, M(\mathrm{d}x) \\ &= \int_{\mathbb{E}}\int_{\mathbb{M}} \mathbf{1}(x\in B)(1-\mathbf{1}(x\in\mu))\, P_x(\mathrm{d}\mu)M(\mathrm{d}x) \\ &\boxed{\text{CLM with } f(x,\mu)=\mathbf{1}(x\in B)(1-\mathbf{1}(x\in\mu))} = \mathbf{E}\left[\int_{\mathbb{E}} \mathbf{1}(x\in B)(1-\mathbf{1}(x\in\Phi))\,\Phi(\mathrm{d}x)\right] \\ &\boxed{\mathbf{1}(x\in\Phi)=1 \text{ for } \Phi(\mathrm{d}x)\text{-all } x} = 0\,. \end{aligned}$$

Hence $P_x(\{\mu : x\in\mu\}) = 1$ for M-almost all $x \in \mathcal{B}_c$. □

Remark 8.1.5. Sometimes one considers *reduced Palm distributions* $P_x^!$ of point process Φ defined as regular versions of the R-N derivatives of the projection of the *reduced Campbell's measure*

$$C^!(B\times\Gamma) := \mathbf{E}\left[\int_{\mathbb{E}} \mathbf{1}(x\in B)\mathbf{1}((\Phi-\delta_x)\in\Gamma)\,\Phi(\mathrm{d}x)\right] \qquad B\in\mathcal{B},\Gamma\in\mathcal{M}$$

on $\mathbb{E}$ with respect to the mean measure M. (Remark the subtraction of the counted atom δ_x form Φ in $\mathbf{1}((\Phi-\delta_x\in\Gamma)$.) Under the assumptions of Theorem 8.1.3, the analog reduced Campbell-Little-Mecke's formula takes the following form

$$\mathbf{E}\left[\int_{\mathbb{E}} f(x,\Phi-\delta_x)\,\Phi(\mathrm{d}x)\right] = \int_{\mathbb{M}}\int_{\mathbb{M}} f(x,\mu)P_x^!(\mathrm{d}\mu)M(\mathrm{d}x). \tag{8.1.4}$$

Using both version of the Campbell-Little-Mecke formula it is easy to show that $P_x^!(\Gamma) = P_x(\{\mu : \mu - \delta_x \in \Gamma\})$ (equivalently $P_x(\Gamma') = P_x^!(\{\mu : \mu + \delta_x \in \Gamma'\})$) for M-almost $x \in \mathbb{E}$. Note by Corollary 8.1.4 that $\mu - \delta_x \in \mathbb{M}$ for P_x almost all μ.

The following useful relation relates the Laplace functional $\mathcal{L} = \mathcal{L}_\Phi$ of a given Point process (recall $\mathcal{L}(f) = \mathbf{E}\left[e^{-\int f \,\mathrm{d}\Phi}\right]$) with the Laplace functionals $\mathcal{L}_x := \mathcal{L}_{P_x}$ of its Palm distributions P_x.

Proposition 8.1.6. *Let Φ be a point process with Radon mean measure M on $\mathbb{E}$ and Laplace functional $\mathcal{L}$.*

1. *For all non-negative, measurable functions f, g on $\mathbb{E}$, bounded, with bounded support*

$$\frac{\mathrm{d}\mathcal{L}(f+tg)}{\mathrm{d}t}\Big|_{t=0} := \lim_{t\to 0} \frac{\mathcal{L}(f+tg) - \mathcal{L}(f)}{t} = -\int_{\mathbb{E}} g(x) L_x(f)\, M(\mathrm{d}x)\,, \tag{8.1.5}$$

   where

$$L_x(f) = \mathcal{L}_{P_x}(f) := \int_{\mathbb{M}} e^{-\int f \,\mathrm{d}\mu}\, P_x(\mathrm{d}\mu)\,. \tag{8.1.6}$$

2. *If (8.1.5) is satisfied for some family of functionals $L_x(\cdot)$ and all non-negative, measurable functions f, g on $\mathbb{E}$, bounded, with bounded support, then L_x are Laplace functionals of P_x, i.e. (8.1.6) is satisfied, for M-almost all $x \in \mathbb{E}$.*

Remark 8.1.7. Note by Proposition 8.1.6 that the family of Palm distributions $P_x(\cdot)$, $x \in \mathbb{E}$ characterizes the distribution of the Point process.

Proof of Proposition 8.1.6. Statement in 1. Note that

$$\begin{aligned}
\Big|\frac{e^{-\int (f+tg)\,\mathrm{d}\Phi} - e^{-\int f\,\mathrm{d}\Phi}}{t}\Big| &= |e^{-\int f\,\mathrm{d}\Phi}|\Big|\frac{e^{-\int tg\,\mathrm{d}\Phi} - 1}{t}\Big| \\
&\le \Big|\frac{e^{-t\int g\,\mathrm{d}\Phi} - 1}{t}\Big| \\
\boxed{|e^{-u} - 1| \le |u|} &\le \int_{\mathbb{E}} g\,\mathrm{d}\Phi
\end{aligned}$$

and by the Campbell's averaging formula $\mathbf{E}\left[\int_{\mathbb{E}} g\,\mathrm{d}\Phi\right] = \int_{\mathbb{E}} g\,\mathrm{d}M < \infty$. Hence, by the Lebesgue dominated convergence theorem

$$\begin{aligned}
\lim_{t\to 0} \frac{\mathcal{L}(f+tg) - \mathcal{L}(f)}{t} &= \mathbf{E}\left[\lim_{t\to 0} \frac{e^{-\int (f+tg)\,\mathrm{d}\Phi} - e^{-\int f\,\mathrm{d}\Phi}}{t}\right] \\
&= \mathbf{E}\left[\int_{\mathbb{E}} g(x)\,\mathrm{d}\Phi \times e^{-\int f\,\mathrm{d}\Phi}\right] \\
\boxed{\text{CLM formula}} &= \int_{\mathbb{E}} g(x) \int_{\mathbb{M}} e^{-\int f\,\mathrm{d}\mu}\, P_x(\mathrm{d}\mu)\, M(\mathrm{d}x)\,,
\end{aligned}$$

which completes the proof of the first statement.
Statement 2. For a given f, since (8.1.5) is satisfied for $L_x(f)$ and $\mathcal{L}_x(f)$ and all g, we have $L_x(f) = \mathcal{L}_x(f)$ for M-almost all $x \in \mathbb{E}$. Topological properties of $\mathbb{E}$ allow one to conclude that $L_x = \mathcal{L}_x$ for M-almost all $x \in \mathbb{E}$. □

8.1.2 Palm distributions of Poisson process — Slivnyak's Theorem

We shall show now another key result of Palm theory regarding Poisson point process.

Theorem 8.1.8 (Slivnyak-Mecke)**.** *Let Φ be a point process of Radon mean measure M on $\mathbb{E}$. Φ is Poisson point process of intensity $\Lambda = M$ iff for M-almost all $x \in \mathbb{E}$*

$$P_x = P_{\Phi+\delta_x}\,,$$

that is Palm distribution P_x of Φ is equal to the distribution $P_{\Phi+\delta_x}$ of Φ with an extra *atom added at x.*

Proof. Assume that Φ is Poisson process of intensity measure $\Lambda = M$. It's Laplace functional is equal to $\mathcal{L}(f) = e^{-\int(1-e^{-f})\,\mathrm{d}\Lambda}$; cf. Fact 2.2 in the lesson on Poisson Point Process. For f, g as in Proposition 8.1.6

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{L}(f+tg)}{\mathrm{d}t}\Big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t} e^{-\int(1-e^{-f-tg})\,\mathrm{d}\Lambda}\Big|_{t=0} \\
&= -e^{-\int(1-e^{-f})\,\mathrm{d}\Lambda}\,\frac{\mathrm{d}}{\mathrm{d}t}\int(1-e^{-f-tg})\,\mathrm{d}\Lambda\Big|_{t=0} \\
&= -e^{-\int(1-e^{-f})\,\mathrm{d}\Lambda}\,\lim_{t\to 0}\int\frac{e^{-f}(1-e^{-tg})}{t}\,\mathrm{d}\Lambda \\
\boxed{|1-e^{-u}| \le |u|,\ \text{Dominated Convergence}} &= -e^{-\int(1-e^{-f})\,\mathrm{d}\Lambda}\int\lim_{t\to 0}\frac{e^{-f}-e^{-f-tg}}{t}\,\mathrm{d}\Lambda \\
&= -\mathcal{L}(f)\int_{\mathbb{E}} g(x)e^{-f(x)}\,\Lambda(\mathrm{d}x)\,.
\end{aligned}$$

Using the second statement of Proposition 8.1.6 we have $\mathcal{L}_x(f) = \mathcal{L}(f)e^{-f(x)}$ for M-almost all $x \in \mathbb{E}$. Clearly this is the Laplace functional of $\Phi + \delta_x$. Indeed:

$$\mathbf{E}\left[e^{-\int f\mathrm{d}(\Phi+\delta_x)}\right] = \mathbf{E}\left[e^{-\int f\mathrm{d}\Phi - f(x)}\right] = \mathcal{L}(f)e^{-f(x)}\,.$$

This completes the proof of the direct part of the result.

For the reverse part, for f as in Proposition 8.1.6 define a function $F(t) := \mathcal{L}(tf)$ where $\mathcal{L}$ is the Laplace functional of Φ. We have

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = \lim_{\epsilon\to 0}\frac{\mathcal{L}(tf+\epsilon f)-\mathcal{L}(tf)}{\epsilon}$$

$$\begin{aligned}
\boxed{\text{Proposition 8.1.6, with } f = tf \text{ and } g = f} &= -\int_{\mathbb{E}} f(x)\mathcal{L}_x(tf)\, M(\mathrm{d}x) \\
\boxed{P_x = P_{\Phi+\delta_x}, \text{ hence } \mathcal{L}_x(f) = \mathcal{L}(f)e^{-f(x)}} &= -\mathcal{L}(tf)\int_{\mathbb{E}} f(x)e^{-tf(x)}\, M(\mathrm{d}x) \\
&= F(t)a(t)\,,
\end{aligned}$$

where $a(t) := -\int_{\mathbb{E}} f(x)e^{-tf(x)}\, M(\mathrm{d}x)$. The differential equation $\mathrm{d}F(t)/\mathrm{d}t = F(f)a(t)$ with initial condition $F(0) = 1$ admits a unique solution of the form

$$\begin{aligned}
F(t) &= \exp\Big[\int_0^t a(s)\,\mathrm{d}s\Big] \\
&= \exp\Big[-\int_0^t \int_{\mathbb{E}} f(x)e^{-sf(x)}\, M(\mathrm{d}x)\mathrm{d}s\Big] \\
&= \exp\Big[-\int_0^t \int_{\mathbb{E}} \mathbf{1}(f(x) > 0)f(x)e^{-sf(x)}\, M(\mathrm{d}x)\mathrm{d}s\Big] \\
\boxed{\text{Fubini's Theorem}} &= \exp\Big[-\int_{\mathbb{E}} \mathbf{1}(f(x) > 0)f(x)\int_0^t e^{-sf(x)}\,\mathrm{d}s M(\mathrm{d}x)\Big] \\
&= \exp\Big[-\int_{\mathbb{E}} \mathbf{1}(f(x) > 0)f(x)\frac{1}{f(x)}(1 - e^{tf(x)})\, M(\mathrm{d}x)\Big] \\
&= \exp\Big[-\int_{\mathbb{E}} (1 - e^{tf(x)})\, M(\mathrm{d}x)\Big]
\end{aligned}$$

and we recognize the expression of the Laplace functional of Poisson point process of intensity $\Lambda = M$. □

Remark 8.1.9. We conclude from Slivnyak-Mecke's theorem that the reduced Palm distribution of Poisson point process is equal to the original distribution: $P_x^! = P_\Phi$. Moreover this property characterizes Poisson point process distribution.

Corollary 8.1.10 (Cambell-Little-Mecke's formula for Poisson process)**.** *Let Φ be a Poisson point process of intensity Λ. For f as in Proposition 8.1.3 we have*

$$\mathbf{E}\left[\int_{\mathbb{E}} f(x,\Phi)\,\Phi(\mathrm{d}x)\right] = \int_{\mathbb{E}} \mathbf{E}\left[f(x,\Phi+\delta_x)\right]\Lambda(\mathrm{d}x) \tag{8.1.7}$$

and the reduced version

$$\mathbf{E}\left[\int_{\mathbb{E}} f(x,\Phi-\delta_x)\,\Phi(\mathrm{d}x)\right] = \int_{\mathbb{E}} \mathbf{E}\left[f(x,\Phi)\right]\Lambda(\mathrm{d}x). \tag{8.1.8}$$

8.2 Exercises

1. Let X be a non-negative, integer-valued random variable with finite mean, $\mu \in \mathbb{M}$ a Radon counting measure on $\mathbb{E}$ and define a point process $\Phi := X\mu$. Show that its mean measure is equal to $M = \mathbf{E}[X]\,\mu$ the and Palm distribution P_x is the distribution of $\Phi_x := Y\mu$, where Y has the size biased distribution of X

$$\mathbf{P}\{\, Y = k \,\} = \frac{k\mathbf{P}\{\, X = k \,\}}{\mathbf{E}[X]} \qquad k = 0, 1, \ldots \tag{8.2.1}$$

(*Hint:* Verify for Campbell's measures $C(B \times \{k\mu\}) = \mu(B)\mathbf{P}\{\, X = k \,\}/\mathbf{E}[X]$.) In particular, if X is a Poisson random variable of mean λ, then $Y \stackrel{\mathcal{L}}{=} X + 1$.

2. For Φ as in Exercise 1 and x such that $\mu(\{x\}) > 0$ show that $\mathbf{P}\{\, \Phi = k\mu \,|\, x \in \Phi \,\} = \mathbf{P}\{\, X = k \,|\, X > 0 \,\}$. Observe that in general this last conditional law of X given $X > 0$ is not equal to the size biased law of X in (8.2.1).

3. Prove the statement of Remark 8.1.2.

4. For $\Gamma \in \mathcal{M}$, prove that if $\mathbf{P}\{\, \Phi \in \Gamma \,\} = 1$ then $P_x(\Gamma) = 1$ for M-almost all $x \in \mathbb{E}$.

5. Prove the relations between P_x and $P_x^!$ given in Remark 8.1.5.

6. Let Φ be a point process with Radon mean measure M. For a non-negative, measurable function f on $\mathbb{E}$, bounded, with bounded support, such that $0 < \int_{\mathbb{E}} f \,\mathrm{d}M$, let us define a probability measure

$$P_f(\cdot) = \frac{\int_{\mathbb{E}} P_x(\cdot) f(x)\, M(\mathrm{d}x)}{\int_{\mathbb{E}} f \,\mathrm{d}M}$$

on $(\mathbb{M}, \mathcal{M})$. Show that

$$\mathcal{L}_\Phi(f) = 1 - \int_{\mathbb{E}} f \,\mathrm{d}M \times \int_{\mathbb{M}} g(\mu)\, P_f(\mathrm{d}\mu)\,,$$

where

$$g(\mu) = \frac{1 - e^{-\int_{\mathbb{E}} f \,\mathrm{d}\mu}}{\int_{\mathbb{E}} f \,\mathrm{d}\mu}$$

if $\int_{\mathbb{E}} f \,\mathrm{d}\mu > 0$ and 1 otherwise. Conclude that the law of Φ is characterized by the family of laws P_f, for f as assumed above.

7. *Palm distribution at two points.* Define the *moment measure of order* 2 of Φ as a measure on $(\mathbb{E}^2, \mathcal{B}^{\otimes 2})$ satisfying $M^2(B_1 \times B_2) := \mathbf{E}[\Phi(B_1)\Phi(B_2)]$. Assume that M^2 is a Radon measure. For each $x, y \in \mathbb{E}$, consider a probability distribution $P_{x,y}$ on $\mathbb{M}$ as the Palm distribution at y of the point process Φ_x being the Palm version of Φ at x: $P_{x,y} := (P_x)_y$. We call $P_{x,y}$, for $(x, y) \in \mathbb{E}^2$ the family of *two-fold Palm distributions* of

Φ. Prove that for M^2-almost all $(x,y) \in \mathbb{E}^2$ $P_{x,y} = P_{y,x}$ and they satisfy the following *Campbell-Little-Mecke's* formula of the second order: for $f \geq 0$ measurable on $\mathbb{E}^2 \times \mathbb{M}$

$$\mathbf{E}\left[\int_{\mathbb{E}^2} f((x,y),\Phi)\,\Phi(\mathrm{d}x)\Phi(\mathrm{d}y)\right] = \int_{\mathbb{E}^2}\int_{\mathbb{M}} f((x,y),\mu)P_{x,y}(\mathrm{d}\mu)M^2(\mathrm{d}(x,u))\,.$$

8. Observe that for Poisson process of intensity Λ, $M^2(B_1\times B_2) = \Lambda(B_1)\Lambda(B_2)+\Lambda(B_1\cap B_2)$ and $P_{x,y} = P_{\Phi+\delta_x+\delta_y}$.

9. Define *reduced Palm distributions of order 2* via $P^!_{x,y}(\Gamma) = P_{x,y}(\{\mu : \mu - \delta_x - \delta_y \in \Gamma\})$. For Poisson point process of intensity Λ prove

$$\mathbf{E}\left[\int_{\mathbb{E}^2} f((x,y),\Phi)\,\Phi(\mathrm{d}x)(\Phi-\delta_x)(\mathrm{d}y)\right] = \int_{\mathbb{E}^2} \mathbf{E}\left[f((x,y),\Phi)\right]\,\Lambda(\mathrm{d}x)\Lambda(\mathrm{d}y)\,.$$

10. *Gibbs point process.* Let Φ be Poisson point process of intensity Λ on $\mathbb{E}$. For a given non-negative, measurable function ϕ on $\mathbb{M}$ such that $\mathbf{E}\left[\phi(\Phi)\right] = 1$ define a probability distribution P_ϕ on $(\mathbb{M},\mathcal{M})$, $P_\phi(\Gamma) := \mathbf{E}\left[\phi(\Phi)\mathbf{1}(\Phi\in\Gamma)\right]$; i.e., ϕ is the density of P_ϕ with respect to the distribution P_Φ of Poisson point process Φ. The distribution P_ϕ is called *Gibbs distribution* with density ϕ with respect to Poisson distribution of intensity Λ. Denote by Φ_ϕ a point process (on a suitable extension of the given probability space) having Gibbs distribution P_ϕ.

    Show that the mean measure of Φ_ϕ is equal to

$$M_{\Phi_\phi}(B) = \mathbf{E}\left[\Phi_\phi(B)\right] = \mathbf{E}\left[\phi(\Phi+\delta_x)\right]\Lambda(\mathrm{d}x)\,,$$

    i.e., $\mathbf{E}\left[\phi(\Phi+\delta_x)\right]$ is the density of M_{Φ_ϕ} with respect to the intensity of Poisson process.

11. Show that the Palm distributions of the Gibbs point process Φ_ϕ as in Exercise 10 are Gibbs distribution with the density $\phi_x(\mu) := \frac{\phi(\mu)}{\mathbf{E}[\phi(\Phi)+\delta_x]}$ with respect to the Palm distribution P_x of the Poisson process Φ at x. Consequently, the reduced Palm distribution of Φ_ϕ have the densities

$$\phi^!_x(\mu) := \frac{\phi(\mu+\delta_x)}{\mathbf{E}\left[\phi(\Phi)+\delta_x\right]}$$

    with respect to the original Poisson process Φ.

12. *Computer exercise.* Using R and *spatstat* consider the waypoint mobilliy model described in Figure 8.1, cf also Le Boudec ((2007)). Consider successive i.i.d. waypoints of the particle and choose the unit speed of the particle. Construct the following two point processes: Φ_{wp} consisting of the first 500 generated waypoints and Φ consisting of the locations of the particle at 500 equally spaced time instants.

    (a) Plot both point processes and observe the difference between them.

(b) Using the complete independence property explain why Φ_{wp} is a homogeneous Poisson process (more precisely, Poisson process conditioned to have 100 points in the window).

(c) Estimate and plot the density of both point processes using

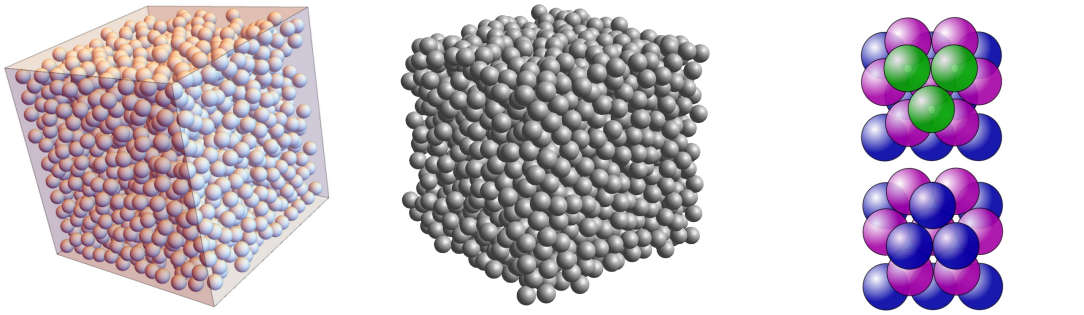
```
L <- density.ppp(pp, 0.05)
plot(L)
```

(d) Compare the density of both point processes in the middle of the window and close to the edges.

(e) Explain in what sense Φ_{wp} can be called a Palm version.

Lesson 9

Hard-core point processes



RSA, 0.38% volume | jamming, 63% volume | two optimal, 74% volume

Figure 9.1: Four sphere packing models in 3D with their volume fractions: saturated random sequential adsorption (RSA), random addition and jamming https://www.digitalrocksportal.org/projects/41 and two optimal packings, which are regular cubic and hexagonal arrangements.

In geometry, a sphere packing is an arrangement of non-overlapping spheres, usually considered of the same radius, within a given space. The problem has a long history and is still an active research area.

A *regular* or *lattice* packing is the one in which the centers of the spheres form a symmetric, periodic pattern. The problem is trivial in 1D (one dimension), relatively easy in 2D, where the placement of the disks on the hexagonal lattice achieves the highest fraction $1/6\pi\sqrt{3} \approx 0.90\%$ of the covered area. For 3D, it was conjectured by Kepler in 1611 that the cubic or hexagonal packings (the rightmost configurations on Figure 9.1), which both have volume fraction $\pi/(3\sqrt{2}) \approx 74\%$, are the densest possible amongst all (regular and irregular) arrangements — this became known as the Kepler conjecture. Carl Friedrich Gauss proved in 1831 that indeed, these packings have the highest density amongst all possible lattice packings. The full Kepler conjecture was proved only recently, in 2014, by using a formal

automated proof checking of some complex computer comparison of the densities for many individual cases, to which the prove was reduced earlier in 1998 by Thomas Callister Hales, thus removing any doubt [1].

Motivated by applications in materials science, a lot of research was devoted to finding good models for *irregular* sphere packing, where the sphere centers are randomly located, with possibly high fraction of the covered space. Point processes, which can model locations of such spheres are called *hard-core* point processes, with the *exclusion distance* R equal to the double of the sphere radius $R/2$.

Among the densest random packing of equal spheres are the *rigid* or *jammed* models. In these models the spheres are randomly added to a container and then displaced (compressed) in some way, until no more sphere can be added. In 1D and 2D this leads to a regular packing but in 3D simulations and theoretical predictions suggest that a volume fraction of only about 64% can be achieved, cf Song et al. ((2008)).

In this lesson we shall consider three so called Matérn models, much simpler to analyze, which arise from the homogeneous Poisson point process by some *hard-core thinning* of points, i.e. selecting a subset of its points respecting the exclusion distance. They are appealing by the simplicity of the analysis and simulations, but can achieve only a moderately high volume fraction of about 0.38% in 3D, for the most efficient among them, the *random sequential adsorption (RSA)* model. Matérn-type thinnings give only some specific examples of thinning operations, see Hirsch and Last ((2017)) for more considerations.

Another possibility to construct hard-core models from Poisson distribution consists in conditioning on the observation of a hard-core configuration. This leads to *Gibbs hard-core models*. Gibbs hard-core processes can achieve higher volume fractions but are notoriously difficult to simulate. For more on Gibbs and Gibbs hard-core models see ((Chiu et al., 2013, Section 5.5.2, 5.5.3))

9.1 Hard-core thinning of Poisson process

Let Φ be a homogeneous Poisson process with intensity λ on the d-dimensional Euclidean space $\mathbb{R}^d$. Note it is a simple point process. Let $R > 0$ be a fixed parameter. Any point $X_i \in \Phi$ is called R-isolated (or isolated for short) if there is no any other point of Φ within the distance R from it. Denote by $\Phi_1 \subseteq \Phi$ all isolated points of Poisson process Φ. This is clearly a hard-core point process, the most natural one can construct from Poisson process. In the next section we shall calculate the mean measure of this process and see that it does not scale good with increasing intensity λ of the original Poisson process. We shall also see that $\mathbf{P}\{\,\Phi_1 = \Phi\,\} = 0$, that is the probability of having all points isolated in a Poisson configuration of points is equal to zero. This rules out the possibility to construct more dense hard-core point processes by conditioning on Poisson point process having all isolated points.

[1] cf "Google Code Archive - Long-term storage for Google Code Project Hosting." https://code.google.com/p/flyspeck/wiki/AnnouncingCompletion.

More subtle, local conditioning is required in this regard leading to Gibbs hard-core models; cf Exercise 3. Another possibility is to select a hard-core subset of each realization of Poisson process in a more efficient way than the isolated points Φ_1. This operation can be seen as a dependent thinning of the Poisson process (Recall that independent thinning leads again to a Poisson process, of smaller intensity.)

In the next sections we shall consider the isolated points and two other, more efficient dependent thinning of Poisson process leading to hard-core models. This will be also a good occasion to see applications of some fundamental tools of Palm theory for Poisson process: Campbell-Little-Mecke (CLM) and Slivnyak-Mecke theorems.

9.1.1 Isolated points of Poisson process

Consider the isolated points of the Poisson process

$$\begin{aligned}\Phi_1 &:= \sum_{X_i\in\Phi} \delta_{X_i}\mathbf{1}(X_i \text{ is } R\text{-isolated})\\ &= \sum_{X_i\in\Phi} \delta_{X_i}\mathbf{1}\Big(\Phi\big(B_{X_i}(R)\big)=1\Big),\end{aligned}$$

where $B_x(r)$ is the ball centered at x of radius r. Denote by M_1 the mean measure of Φ_1; $M_1(B)=\mathbf{E}\left[\Phi_1(B)\right]$ for any $B\in\mathcal{B}$, where $\mathcal{B}$ are Borel subsets of $\mathbb{R}^d$.

Proposition 9.1.1. *The mean measure M_1 of the isolated points of Poisson process is the constant multiple of the Lebesgue measure $M_1(\mathrm{d}x)=\lambda_1\,\mathrm{d}x$, where*

$$\lambda_1=\lambda_1(\lambda,R)=\lambda e^{-\nu_d\lambda R^d}$$

and $\nu_d=|B_0(1)|$ is the volume of the unit-radius ball in $\mathbb{R}^d$.

Proof. Consider a bounded, Borel set $B\subset\mathbb{R}^d$. We have

$$\begin{aligned}M_1(B) &= \mathbf{E}\left[\int_{\mathbb{R}^d}\mathbf{1}(x\in B)\mathbf{1}\Big(\Phi\big(B_x(R)\big)=1\Big)\,\Phi(\mathrm{d}x)\right]\\ \text{CLM formula}\quad &= \lambda\int_{\mathbb{R}^d}1(x\in B)\int_{\mathbb{M}}\mathbf{1}\Big(\mu\big(B_x(R)\big)=1\Big)\,P_x(\mathrm{d}\mu)\mathrm{d}x\\ \text{Slivnyak-Mecke's formula}\quad &= \lambda\int_{\mathbb{R}^d}1(x\in B)\,\mathbf{P}\big\{\,(\Phi+\delta_x)\big(B_x(R)\big)=1\,\big\}\,\mathrm{d}x\\ &= \lambda\int_{\mathbb{R}^d}1(x\in B)\,\mathbf{P}\big\{\,\Phi\big(B_x(R)\big)=0\,\big\}\,\mathrm{d}x\\ \text{Poisson assumption}\quad &= \lambda|B|e^{-\lambda|B_0(R)|}\\ &= \lambda|B|e^{-\nu_d\lambda R^d},\end{aligned}$$

which completes the proof. □

Remark 9.1.2. We can call the constant λ_1 the intensity of the isolated points of the homogeneous Poisson process of intensity λ. Note that Φ_1 can be seen as a thinning of Φ with probability $p_1 := \lambda_1/\lambda = e^{-\nu_d \lambda R^d}$. It is of course *not* independent thinning that would lead to a Poisson (and not a hard-core) process of same mean measure $\lambda_1 \, \mathrm{d}x$.

Remark 9.1.3 (Maximal density of isolated points)**.** Note that, for given $R > 0$, the intensity $\lambda_1 = \lambda e^{-\nu_d \lambda R^d}$ of isolated Poisson points, as a function of λ, first increases to attain its maximum value $1/(e\nu_d R^d)$ for $\lambda = 1/(\nu_d R^d)$ and then decreases to 0 as $\lambda \to \infty$. In other words, the densest hard-core point process with exclusion distance R one can obtain as the subset of isolated points of a Poisson process has the intensity $1/(e\nu_d R^d)$.

Let us now give a look at the non-isolated points of Poisson process $\Phi' := \Phi - \Phi_1$. The mean measure M' of this process is obviously equal to $M'(\mathrm{d}x) = (\lambda - \lambda_1) \, \mathrm{d}x = \lambda(1 - p_1) \, \mathrm{d}x$. Since $p_1 = e^{-\nu_d \lambda R^d} < 1$ for any $\lambda > 0$, $R > 0$, the expected number of non-isolated points of Poisson process in a set given set B is strictly positive $M'(B) > 0$ provided $|B| > 0$. Moreover, $M'(\mathbb{R}^d) = \infty$. In what follows we shall prove more, namely that $\mathbf{P}\{ \Phi'(\mathbb{R}^d) = \infty \} = 1$. Note, it does not follow from $\mathbf{E}\left[\Phi'(\mathbb{R}^d)\right] = \infty$, but is a consequence of the ergodicity of Poisson process, that we will considered in a future lesson. Here we will prove it directly from the complete independence of Poisson process using the classical law of large numbers.

Proposition 9.1.4. *For Poisson point process of intensity λ, $0 < \lambda < 0$, the number of non-isolated points is infinite almost surely; $\mathbf{P}\{ \Phi'(\mathbb{R}^d) = \infty \} = 1$.*

Proof. Consider some windows $W_i \subset \mathbb{R}^d$ $i = 0, 1, \ldots$, which are translations of $W_1 := [0, 1]^d$ such that the distance between W_i and W_j is bigger than $2R$, for any $i \neq j$. Obviously one can find infinite sequence of such windows. By the complete independence and translation invariance of the distribution of Poisson process $\Phi'(W_i)$, $i = 1, 2, \ldots$ are iid rv's. By the strong law of large numbers, with probability 1

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\Phi'(W_i) > 0) = \mathbf{P}\{ \Phi'(W_1) > 0 \} > 0 \,,$$

where the last inequality results from $\mathbf{E}\left[\Phi'(W_1)\right] = \lambda(1 - p_1) > 0$. Consequently, with probability 1, there is a sub-sequence i_k, $k = 1, 2, \ldots$ of windows with at least one isolated point in each window $\Phi'(W_{i_k}) > 0$, which completes the proof. □

As we said in the introduction to this section, the result of Proposition 9.1.4 rules out the possibility of obtaining infinite, homogeneous hard-core configurations of points by conditioning on the entire Poisson point process satisfying this condition.

9.1.2 Matérn hard-core model

The point process Φ_1 of isolated points of Poisson process is sometimes called Matérn I hard-core model. In this section we present a more efficient strategy of dependent thinning of

Poisson process, which leads to a model called Matérn (or Matérn II) hard-core model. In order to realize this hard-core thinning of the homogeneous Poisson process Φ of intensity λ, we first consider its independent marking

$$\tilde{\Phi} = \sum_{X_i \in \Phi} \delta_{(X_i, U_i)}\,, \tag{9.1.1}$$

where, given Φ, U_i are iid rv's, uniform on $[0,1]$. Recall from the lesson on Poisson point process that $\tilde{\Phi}$ is a Poisson point process on $\mathbb{R}^d \times [0,1]$ with intensity measure $\tilde{\Lambda}(\mathrm{d}(x,u)) = \lambda\,\mathrm{d}x\mathrm{d}u$, $x \in \mathbb{R}^d$, $u \in [0,1]$ (cf Exercise 1). We interpret U_i as the *birth time* of the point X_i. Denote by Φ^a the points of Φ which are born up to time a, $a \in [0,1]$

$$\Phi^a := \sum_{(X_i,U_i)\in\tilde{\Phi}} \delta_{X_i} \mathbf{1}(U_i \le a)\,.$$

Observe that, as an independent thinning of Φ, Φ^a is a homogeneous Poisson process on $\mathbb{R}^d$ of intensity $\lambda \mathbf{P}\{\, U \le a \,\} = \lambda a$, where U is a uniform random variable on $[0,1]$.

We define the Matérn hard-core point process as consisting of all points of Φ, which do not have in their neighbourhood of radius R any point of Φ born earlier (with smaller mark U)

$$\begin{aligned}\Phi_2 &:= \sum_{(X_i,U_i)\in\tilde{\Phi}} \delta_{X_i} \mathbf{1}\Big(\tilde{\Phi}\big(B_{X_i}(R)\times[0,U_i]\big) = 1\Big) \\ &= \sum_{(X_i,U_i)\in\tilde{\Phi}} \delta_{X_i} \mathbf{1}\Big(\Phi^{U_i}\big(B_{X_i}(R)\big) = 1\Big)\,,\end{aligned}$$

where we recognize in Φ^{U_i} points of Φ born up to time U_i, at which X_i was born. Obviously $X_i \in \Phi^{U_i}$.

Observe that the isolated points form a subset of the Matérn point process, $\Phi_1 \subset \Phi_2$.

Proposition 9.1.5. *The mean measure M_2 of the Matérn hard-core model is the constant multiple of the Lebesgue measure $M_2(\mathrm{d}x) = \lambda_2\,\mathrm{d}x$, where*

$$\lambda_2 = \lambda_2(\lambda, R) = \frac{1 - e^{-\nu_d \lambda R^d}}{\nu_d R^d}\,.$$

Proof. Consider a bounded, Borel set $B \subset \mathbb{R}^d$. Denote by $\tilde{\mathbb{M}}$ counting measures on $\mathbb{R}^d \times [0,1]$ We have

$$M_2(B) = \mathbf{E}\left[\int_{\mathbb{R}^d}\int_0^1 \mathbf{1}(x\in B)\mathbf{1}\Big(\tilde{\Phi}\big(B_x(R)\times[0,u]\big)=1\Big)\,\tilde{\Phi}(\mathrm{d}(x,u))\right]$$

$$\overset{\text{CLM formula for } \tilde{\Phi}}{=} \lambda \int_{\mathbb{R}^d} 1(x\in B)\int_0^1\int_{\tilde{\mathbb{M}}} \mathbf{1}\Big(\tilde{\mu}\big(B_x(R)\times[0,u]\big)=1\Big)\,P_{(x,u)}(\mathrm{d}\tilde{\mu})\,\mathrm{d}u\mathrm{d}x$$

$$
\begin{aligned}
\text{Slivnyak-Mecke's formula for } \tilde{\Phi} \quad &= \lambda \int_{\mathbb{R}^d} 1(x \in B) \int_0^1 \mathbf{P}\Big\{ (\tilde{\Phi} + \delta_{(x,u)})\big(B_x(R) \times [0,u]\big) = 1 \Big\} \, \mathrm{d}u\mathrm{d}x \\
&= \lambda \int_{\mathbb{R}^d} 1(x \in B) \int_0^1 \mathbf{P}\Big\{ \tilde{\Phi}\big(B_x(R) \times [0,u]\big) = 0 \Big\} \, \mathrm{d}u\mathrm{d}x \\
&= \lambda \int_{\mathbb{R}^d} 1(x \in B) \int_0^1 \mathbf{P}\big\{ \Phi^u\big(B_x(R)\big) = 0 \big\} \, \mathrm{d}u\mathrm{d}x \\
\text{Poisson assumption} \quad &= \lambda |B| \int_0^1 e^{-\lambda u |B_0(R)|} \, \mathrm{d}u \\
&= |B| \frac{1 - e^{-\nu_d \lambda R^d}}{\nu_d R^d} \, .
\end{aligned}
$$

which completes the proof. □

Remark 9.1.6 (Saturated Matérn model)**.** Note that, for given $R > 0$, λ_1 is an increasing function of λ and

$$\lim_{\lambda \to \infty} \lambda_2(\lambda, R) = \frac{1}{\nu_d R^d} \, .$$

This limiting value is called intensity of the *saturated* Matérn model. It is bigger than the optimal density of isolated points by the factor $e \approx 2.718$; cf Remark 9.1.3.

9.1.3 Random sequential adsorption (RSA) model

One can still improve upon the intensity of the Matérn model. Observe that a point X_i born at time U_i, $(X_i, U_i) \in \tilde{\Phi}$, which is not retained in Φ_2 because of some $(X_j, U_j) \in \tilde{\Phi}$, with $|X_j - X_i| \le R$ and $U_j < U_i$, may prevent another point (X_k, U_k) with $|X_k - X_i| \le R$ with $U_k \ge U_i$ from being accepted for Φ_2. However the hard-core constraint is not violated if the distance of this latter point is bigger from X_j (and all other *retained* points) than R. Improving upon this fact, leads to the following *axiomatic description the RSA model* as a point process Φ_3 satisfying the following property: a point X_i, with birth time U_i is retained in Φ_3 if there is no any other point in Φ_3 within the distance R from it and born before it

$$\Phi_3 = \sum_{(X_i, U_i) \in \tilde{\Phi}} \delta_{X_i} \mathbf{1}\Big(\big(\Phi^{U_i} \cap \Phi_3)\big)\big(B_{X_i}(R)\big) = 1 \Big) \, . \tag{9.1.2}$$

Note that (9.1.2) cannot be straightforwardly considered as the definition of Φ_3. It is rather a postulate, an equation on Φ_3, which appears in the right-hands side of (9.1.2). We have to prove that this equation has a unique solution, which then defines Φ_3.

Sequential adsorption for finite Φ. Suppose for a moment that $\Phi(\mathbb{R}^d)$ is finite. It is relatively easy to see that (9.1.2) has a unique solution, which can be constructed as the following *sequential adsorption* of points: Starting from the point of Φ with the smallest

birth-time, keep adding points of Φ to Φ_3 in order of their birth-times, provided the hard-core constraint is not violated with respect to the previous taken points.

The above construction of the solution to (9.1.2) fails for infinite point patterns Φ, and hence almost surely for any homogeneous Poisson configuration, since $\inf_{X_i\in\Phi} U_i = 0$ and one cannot find the point in Φ with the minimum birth-time. (In fact one cannot find the point born next after any given time instant.) A different approach is required in this case.

Proposition 9.1.7. *For any independently, uniformly marked Poisson point process of intensity $\lambda \in (0,\infty)$, equation (9.1.2) has almost surely a unique solution Φ_3, called the* RSA model.

Proof. Our proof relies on the following non-percolation results that will be proved in a separate lesson on the connectivity of the Boolean model: For any $R > 0$, there exists a constant $\lambda_c = \lambda_c(d) > 0$ such that the graph on the homogeneous Poisson point process of intensity $\lambda' < \lambda_c$ in $\mathbb{R}^d$, $d \geq 2$, with edges connecting any two points closer than R to each other, does not percolate (does not have any infinite component) with probability one. We call the aforementioned graph the *Gilbert graph*.

For the given independently, uniformly marked Poisson point process $\tilde{\Phi}$ of intensity λ, consider Poisson point processes $\Phi'^k = \Phi^{ka} - \Phi^{(k-1)a}$ consisting of all points of the original process born strictly after $(k-1)a$ and not later than ka, $k = 1,\dots,\lceil 1/a\rceil$. As an independent thinning Φ'^k are Poisson processes with intensities equal (or smaller for $k = \lceil 1/a\rceil$) to $a\lambda$. Let us fix $a > 0$ so that $a\lambda < \lambda_c$. Obviously $\sum_{k=1}^{\lceil 1/a\rceil} \Phi'^k = \Phi$.

We construct the solution of (9.1.2) using the following exploration in space and backward in time around any given point of Φ. For any given point $X_i \in \Phi'^k$, we consider all points which are in the same connected component of the Gilbert graph on Φ'^k. Since the intensity of this process is smaller than the critical intensity λ_c, almost surely there are only finitely many such points, denote them by C_k. Consider all connected components of the Gilbert graph on Φ'^{k-1}, which are closer than R to any point of C_k. By the non-percolation of the Gilbert graph on Φ'^{k-1} all these components are finite. By the finiteness of C_k and the local finiteness of (any) point process there can be at most finitely many such components. Consequently, the union of these components is a finite subset of points of Φ, call it C_{k-1}. By the recursion we construct $C_{k-2},\dots,C_1$. All sets $C_1,\dots,C_k$ are finite. Observe also that the condition

$$\mathbf{1}\Big(\big(\Phi^{U_i}\cap\Phi_3)\big)\big(B_{X_i}(R)\big) = 1\Big)$$

in (9.1.2) on whether X_i is taken to Φ_3 or not, does not involve points outside $C := C_1\cup\ldots\cup C_k$. Since this set is finite, we can apply the sequential adsorption strategy to decide which points of C are in Φ_3. This implies a decision for $X_i \in C_k \subset C$, thus completing the proof. □

Remark 9.1.8 (Saturated RSA model)**.** There is no closed form expression for the intensity λ_3 of the RSA model Φ_3, except for $d = 1$, known as the *Rényi car parking problem*. However, one can prove for any dimension d that, for given R, λ_3 is increasing with respect to λ; cf Exercise 2. The limit $\lim_{\lambda\to\infty}\lambda_3$ is called intensity of the *saturated* RSA model. It can

been estimated from simulations. We shall compare the saturated RSA to the saturated Matérn model and the isolated points in therms of the respective volume fractions considered in the next section; cf Table 9.1.

9.2 Volume fractions of hard-core models

Let $\psi = \{x_i\}$ by a deterministic hard-core configuration of points in $\mathbb{R}^d$ exhibiting the exclusion distance R. Let us center a ball of radius $R/2$ at any point of ψ and consider the disjoint (since the hard-core condition) union

$$\Xi = \Xi(\psi, R) := \bigcup_{x_i \in \psi} B_{x_i}(R/2)\,.$$

We define the *volume fraction* of Ξ as the asymptotic fraction of the volume of a large window occupied by Ξ

$$v_\Xi := \liminf_{n\to\infty} \frac{|\Xi \cap W_n|}{|W_n|}\,, \tag{9.2.1}$$

where the $W_n := [-n^{1/d}/2, n^{1/d}/2]^d$ is the window of volume n.

A configuration ψ of points is called a *saturated hard-core configuration* (with exclusion distance R), when no ball can be added to ψ without violating the hard-core condition.

Fact 9.2.1. *The volume fraction v_Ξ of a saturated configuration ψ of points in $\mathbb{R}^d$ is not smaller than $1/2^d$.*

Proof. Ignoring boundary the effects, for large n

$$\begin{aligned}
\frac{|\Xi \cap W_n|}{n} &= \frac{1}{n}\Big|\bigcup_{x_i\in\psi} B_{x_i}(R/2) \cap W_n\Big| \\
\boxed{\text{disjoint union}} &= \frac{1}{n}\sum_{x_i\in\psi}\Big|B_{x_i}(R/2) \cap W_n\Big| \\
\boxed{\text{border effects}} &\approx \frac{1}{n}\sum_{x_i\in\psi\cap W_n} |B_{x_i}(R/2)| \\
&= \frac{1}{2^d}\frac{1}{n}\sum_{x_i\in\psi\cap W_n} |B_{x_i}(R)| \\
\boxed{\text{border effects}} &\approx \frac{1}{2^d}\frac{1}{n}\sum_{x_i\in\psi} |B_{x_i}(R) \cap W_n| \\
&\geq \frac{1}{2^d}\frac{1}{n}\Big|\bigcup_{x_i\in\psi} B_{x_i}(R) \cap W_n\Big|
\end{aligned}$$

$$\boxed{(*)} = \frac{1}{2^d}\frac{|W_n|}{n}$$
$$= \frac{1}{2^d},$$

where (*) follows from the assumption that ψ is saturated: indeed any location $y \in \mathbb{R}^d$ in the space is within the distance at most R from some point of ψ; consequently, $\bigcup_{x_i \in \psi} B_{x_i}(R) = \mathbb{R}^d$. □

For a stationary and ergodic [2] hard-core point process $\Psi = \{X_i\}$ (as our three models Φ_i, $i = 1, 2, 3$ considered previously) one can prove that lim inf = lim in (9.2.1) and is equal to

$$v_\Xi = \frac{1}{|W|}\mathbf{E}\left[\Big|\bigcup_{X_i \in \Psi \cap W}^{\cdot} B_{x_i}(R/2)\Big|\right]$$
$$\boxed{\text{disjoint sum}} = \frac{1}{|W|}\mathbf{E}\left[\sum_{X_i \in \Psi \cap W} |B_{x_i}(R/2)|\right]$$
$$= \frac{1}{|W|}\nu_d (R/2)^d \mathbf{E}\left[\Psi(W)\right]$$
$$= \frac{1}{|W|}\nu_d (R/2)^d \lambda_\Psi |W|$$
$$= \frac{\lambda}{\nu_d}(R/2)^d,$$

where λ_Ψ is the intensity of the hard-core point process Ψ.

In particular the maximal volume fraction obtained with isolated points of Poisson process is equal to $v_1 = 1/(e2^d)$ (cf Remark 9.1.3). The volume fraction related to the saturated Matérn model is equal to $v_2 = 1/2^d$ (cf Remark 9.1.6) thus attaining the lower bound of all saturated configurations in Fact 9.2.1. The volume fraction related to the saturated RSA model (cf Remark 9.1.8), calculated $d = 1$ and estimated for dimension $d = 2, 3$ are presented in Table 9.1.

[2] We shall devote one lesson tho this subject.

dimension d	isolated points	Matérn	RSA	densest packing
1	0.183939...	0.5	0.747598...	1.
2	0.091967...	0.25	0.54700	0.90689...
3	0.045984...	0.125	0.38278	0.74048...

Table 9.1: Volume fractions of some classical hard-sphere models: the densest isolated points of Poisson process, the saturated Matérn and RSA models (for the latter model we took estimated values from Torquato et al. ((2006))) and the densest packing (cf the introduction to this lesson on page 101).

9.3 Exercises

1. Using the Displacement Theorem (Lesson on Poisson process) prove that $\tilde{\Phi}$ is a Poisson process on $\mathbb{R}^d \times [0,1]$ of intensity $\lambda \,\mathrm{d}x\mathrm{d}u$, $x \in \mathbb{R}^d$, $u \in [0,1]$.

2. Prove that the intensity of the RSA model is increasing in λ. Hint: Consider the Poisson point process of intensity $\lambda = \lambda' + \lambda''$, i.i.d. marked by the uniform marks on $[0,1]$ and the corresponding RSA model Φ_3. Consider the points Φ^a of Φ born in the time interval $[0,a]$ for a such that $a\lambda = \lambda'$. Observe that the RSA model constructed with respect to points Φ^a (using their original marks multiplied by $1/a$, which preserves their independence and brings back the distribution to the uniform on $[0,1]$) is equal to the corresponding subset $\Phi_3 \cap \Phi^a$ of the RSA Φ_3.

3. *Finite Gibbs hard-core point process.* Let Φ be a homogeneous Poisson point process of intensity λ in $\mathbb{R}^d$. Let $W \subset \mathbb{R}^d$ be some bounded window. The Gibbs hard-core point process Ψ on W, with exclusion distance R, with respect to Φ is defined in terms of the conditional distribution of Φ having all isolated points in W

$$P_\Psi(\Gamma) := \frac{\mathbf{E}\left[\mathbf{1}(\Phi \in \Gamma) 1(\text{all points of } \Phi \cap W \text{ are } R\text{-isolated})\right]}{\mathbf{P}\{\,\text{all points of } \Phi \cap W \text{ are } R\text{-isolated}\,\}}\,.$$

(Recall form Proposition 9.1.4 that this definition does not make sense for $W = \mathbb{R}^d$.) Prove that mean measure M_Ψ of Ψ is equal to $M_\Psi(\mathrm{d}x) = p(x)\lambda\,\mathrm{d}x$ for $x \in W$, where

$$p(x) = \frac{\mathbf{P}\{\,\text{all points of } (\Phi + \delta_x) \cap W \text{ are } R\text{-isolated}\,\}}{\mathbf{P}\{\,\text{all points of } \Phi \cap W \text{ are } R\text{-isolated}\,\}}\,.$$

For more on finite Gibbs point process, including hard-core ones, see((Chiu et al., 2013, Section 5.5.2)). (The infinite Gibbs hard-core model can be defined as the week limit of P_Ψ, when $W \nearrow \mathbb{R}^d$, provided such a limit exists. In fact there is a phase transition in this regard: the limit exists for λ small enough and does not when λ is too large.)

4. *Gibbs point processes with square-well pair potential (Strauss hard-core).* Consider the following generalization of the Gibbs hard-core point process from Exercise 3. Let ρ, R, b

be constants such that $0 < R < \rho$, $b \in \mathbb{R}$. R is the *exclusion distance*, ρ is called the *interaction distance* and b *interaction parameter*. For a configuration of points $\mu = \{x_i\}$ on W, denote by $S(\mu, t)$ the number of inter-point distances in μ which are less than or equal to t. Let

$$P_\Psi(\Gamma) := \frac{\mathbf{E}\left[\mathbf{1}(\Phi \in \Gamma) f(\Phi \cap W)\right]}{\mathbf{E}\left[f(\Phi \cap W)\right]},$$

where f is the following function on the space of configurations of points μ (counting measures) on W:

$$f(\mu) := \begin{cases} 0 & \text{if } S(\mu, R) > 0 \text{ (hard-core constraint with exclusion distance } R) \\ e^{bS(\mu,\rho)} & \text{if } S(\mu, R) = 0 \text{ (point interaction within distance } \rho). \end{cases}$$

Observe that $b > 0$ gives bias towards configurations with more inter-point distances between R and ρ (as if points were attracting each other on the distance ρ, without however getting closer than R). By increasing b one can obtain more dense hard-core models. On the other hand, $b < 0$ privileges configurations of points with less inter-point distances between R and ρ (as if points were repelling each other even beyond the distance R and up to ρ). Obviously $b = 0$ is the previously considered Gibbs hard-core model from Exercise 3. The case $R = 0$ is called *Strauss model*.

5. *Computer exercise.* The following hard-core model generators are available in *spatstat* R package:

   - `rMaternI` generates the hard-core point process of isolated points (Matérn I) The initial intensity of the Poisson process, exclusion distance and a simulation window can be specified.
   - `rMaternII` generates the Matérn (Matérn II) hard core model.
   - `rSSI` generates the following variant of the RSA hard-core model in a finite window: starting with an empty window, the algorithm adds points one-by-one. Each new point is generated uniformly in the window and independently of preceding points. If the new point lies closer than R units from an existing point, then it is rejected and another random point is generated. The algorithm terminates either when the desired number of points is reached, or when the current point configuration has not changed for some number of iterations.
   - `rHardcore` generates Gibbs hard-core process.
   - `rStraussHard` generates Strauss hard-core process.

   (a) Using *spatstat* simulate a few selected hard-core point processes and estimate their intensities.

   (b) Estimate Ripley's $L(r)$ function for hard-core models and explain observations.

   (c) Estimate Ripley's $L(r)$ function for Strauss hard-core model with $b > 0$ and $b < 0$ and explain observations.

Lesson 10

Stationary point processes and mass transport principle

Stationarity can be defined for a random field, point process, random graph, or random set in $\mathbb{R}^d$ as the invariance property of the distribution of this structure with respect to any translation in the space. This assumption implies several interesting intrinsic properties of these structures often referred to as conservations laws. In the case of a stationary point process these properties often involve the *Palm probability* of the point process.

While in the general context of (non-stationary) point processes the Palm distributions P_x were interpreted as the conditional distributions given a point located at x, the (unique) Palm probability $\mathbf{P}^0$ of a stationary point process will be interpreted as the probability under which we observe the *typical point* of this process located at the origin 0. (The two notions are related, as we shall see.) For this Palm probability we shall prove a *unimodularity* property analogue to the property we have introduced for random abstract graphs.

For a supplementary material to this lesson see the working book project ((Baccelli et al., 2017, Section 7.1)).

10.1 Stationary point processes

10.1.1 Stationary framework

We consider point processes on d-dimensional Euclidean space $\mathbb{E} = \mathbb{R}^d$, with Borel σ-algebra $\mathcal{B}$ and the corresponding space of counting measures $(\mathbb{M}, \mathcal{M})$ on it. Point process $\Phi : (\Omega, \mathcal{A}, \mathbf{P}) \longrightarrow (\mathbb{M}, \mathcal{M})$ is called *stationary* if its distribution $\mathbf{P}_\Phi$ (which is the image of the probability $\mathbf{P}$ by Φ on $(\mathbb{M}, \mathcal{M})$; $\mathbf{P}_\Phi(\Gamma) = \mathbf{P}\{\, \Phi \in \Gamma \,\}$, $\Gamma \in \mathcal{M}$) is invariant with respect to the translation by any vector $t \in \mathbb{R}^d$

$$\mathbf{P}_\Phi = \mathbf{P}_{S_t\Phi} \qquad \text{for all } t \in \mathbb{R}^d, \tag{10.1.1}$$

where $S_t\Phi$ is a point process resulting from the translation of all atoms of Φ by $-t$. More generally, stochastic process $X = \{X(x)\}_{x\in\mathbb{R}}$ on $(\Omega, \mathcal{A}, \mathbf{P})$ with values in some measurable

space $(\mathbb{K}, \mathcal{K})$ is called *stationary* if its distribution is invariant with respect to the translation of its argument x by any vector $t \in \mathbb{R}^d$

$$\mathbf{P}_{\{X(x)\}_{x\in\mathbb{R}^d}} = \mathbf{P}_{\{X(x+t)\}_{x\in\mathbb{R}^d}} \qquad \text{for all } t \in \mathbb{R}^d. \tag{10.1.2}$$

A family of point processes Φ_i and stochastic processes X_i, $i = 1, 2, \ldots$ defined on the same probability space is called *jointly stationary* if the *joint* distribution of all these random objects is invariant with respect to the respective translation by any vector $t \in \mathbb{R}^d$.

In order to facilitate the analysis of such jointly stationary objects we shall introduce some *stationary framework* directly on the probability space $(\Omega, \mathcal{A}, \mathbf{P})$. In particular, it allows one to define the probabilities of different point processes directly on $(\Omega, \mathcal{A})$ and study the relations between these probabilities as well as consider distributions of all random object defined on $(\Omega, \mathcal{A})$ under these Palm probabilities. As we shall see, this gives rise to several interesting conservation laws including *mass transport principle* for point processes.

Shift operator on measures and functions

For any $t \in \mathbb{R}^d$, let S_t be the *shift operator* on the space of measures: for any measure μ on $(\mathbb{R}^d, \mathcal{B})$ $S_t\mu$ is a measure on $(\mathbb{R}^d, \mathcal{B})$ such that

$$S_t\mu(B) = \mu(B + t),$$

where $B + t = \{x + t \in \mathbb{R}^d : x \in B\}$. Note, in case of a counting measure $\mu \in \mathbb{M}$ with $\mu = \sum_i \delta_{x_i}$ we have

$$S_t\mu = \sum_i \delta_{x_i - t}\,. \tag{10.1.3}$$

We extend the shift operator to all functions $X(\cdot)$ defined on $\mathbb{R}^d$ with values is some arbitrary space, by putting

$$S_t X(x) = X(x + t)\,.$$

The following immediate relation will be often used, for $B \in \mathcal{B}$, $t \in \mathbb{R}^d$

$$\int_B X(x)\, S_t\mu(\mathrm{d}x) = \int_{B+t} X(x - t)\, \mu(\mathrm{d}x)\,. \tag{10.1.4}$$

Flow on the probability space

Consider a measurable space $(\Omega, \mathcal{A})$ that will serve as the probability space. We assume there exists a family of measurable mappings $\theta_t : \Omega \to \Omega$, $t \in \mathbb{R}^d$ satisfying the following conditions:

1. For each $t \in \mathbb{R}^d$, the mapping θ_t is a bijection from Ω to Ω.
2. For all $t, s \in \mathbb{R}^d$, $\theta_t \circ \theta_s = \theta_{s+t}$, with $\circ$ denoting the composition of mappings on Ω.
3. The mapping $(\mathbb{R}^d, \Omega) \ni (t, \omega) \longmapsto \theta_t(\omega)$ is $\mathcal{B} \otimes \mathcal{A}$ measurable.

Observe that for any given $t \in \mathbb{R}^d$, the inverse of θ_t is equal to $\theta_t^{-1} = \theta_{-t}$.

The family $\{\theta_t\}_t$ of mappings satisfying conditions $1, 2, 3$ above will be called *(measurable) flow* on $(\Omega, \mathcal{A})$. We shall denote the space equipped with such flow by $(\Omega, \mathcal{A}, \{\theta_t\})$.

Processes compatible with the flow

We shall say that a point process $\Phi : (\Omega, \mathcal{A}, \{\theta_t\}) \to (\mathbb{M}, \mathcal{M})$ is *compatible with the flow* if for all $t \in \mathbb{R}^d$

$$\Phi \circ \theta_t = S_t \Phi \,,$$

where $S_t\Phi$ is the shift of the counting measure Φ; in other words

$$\Phi(\theta_t(\omega))(B) = S_t\Phi(\omega)(B) = \Phi(\omega)(B + t)$$

for all $\omega \in \Omega$, $B \in \mathcal{B}$.

Example 10.1.1 (Canonical probability space with the flow)**.** The space $(\Omega, \mathcal{A}, \{\theta_t\}) = (\mathbb{M}, \mathcal{M}, \{S_t\})$ is the canonical space supporting point process $\Phi(\mu) = \mu$, $\mu \in \mathbb{M}$, compatible with the flow.

Similarly, stochastic process $X = \{X(x)\}_{x\in\mathbb{R}^d}$ defined on $(\Omega, \mathcal{A}, \{\theta_t\})$, with values some measurable space, will be said *compatible with the flow* if for all $t \in \mathbb{R}^d$

$$X \circ \theta_t = S_t X \,;$$

that is

$$\{X(\theta_t(\omega))(x)\}_{x\in\mathbb{R}^d} = \{S_t X(\omega)(x)\}_{x\in\mathbb{R}^d} = \{X(\omega)(x + t)\}_{x\in\mathbb{R}^d} \,.$$

Remark 10.1.2. For any stochastic process X compatible with the flow on $(\Omega, \mathcal{A}, \{\theta_t\})$ we have

$$X(x) = X \circ \theta_x(0) \qquad \text{for all } x \in \mathbb{R}^d;$$

that is $X(\omega)(x) = X(\theta_x(\omega))(0)$. In words: the value of the stochastic process X at x is equal to its value at the origin $x = 0$, but taken at some different ω, namely $\theta_x(\omega)$. This means that any stochastic process compatible with the flow is generated by one random variable $f(\omega)$, namely its value at the origin $f(\omega) = X(\omega)(0)$, with values at all other points $X(x)$ obtained by the appropriate flow-shift in ω. It this context, we say that the *stochastic process X is generated by random variable f* if $X(x) = f \circ \theta_x$. Clearly $f = X(0)$.

Most of our stochastic processes compatible with the flow will be fonctionals of some point processes compatible with the flow, as in the following example.

Example 10.1.3. Let Φ be a point process compatible with the flow in $(\Omega, \mathcal{A}, \{\theta_t\})$. Consider stochastic process

$$X(x) := \min_{y\in\Phi} |y - x|$$

describing the distance from the argument x to the nearest point of Φ. Observe that X is generated by the random variable

$$R^* = \min_{y\in\Phi} |y|$$

that is the distance from the origin to the nearest point of Φ. Indeed,

$$R^* \circ \theta_x = \min_{y \in \Phi \circ \theta_x} |y|$$

$$\boxed{\text{compatibility of } \Phi} = \min_{y \in S_x \Phi} |y|$$

$$\boxed{\text{by (10.1.3)}} = \min_{y' \in \Phi} |y - x| \,.$$

Marked point process compatible with the flow

Let Φ be a simple point process compatible with the flow on the probability space $(\Omega, \mathcal{A}, \{\theta_t\})$. We assume Φ is simple (has no multiple points); we will speak about marked point process only in such case. [1] Usually when considering marking of points, one thinks of points of $\Phi = \sum_i \delta_{x_i}$ being numbered in some measurable way (i.e., x_i are random variables, for all i). In this context, the *marked version* of the point process Φ, with the *sequence of marks* generated by a random variable K, is the point process

$$\tilde{\Phi} = \sum_i \delta_{(x_i, k_i)} \,,$$

on $(\mathbb{R}^d \times \mathbb{K}, \mathcal{B} \otimes \mathcal{K})$ where

$$k_i = K \circ \theta_{x_i}. \tag{10.1.5}$$

Note that k_i are random variables depending on $\omega \in \Omega$ in a twofold way $k_i(\omega) = K(\theta_{X_i(\omega)}(\omega))$. These marks are values of a stochastic process $K(x) := K \circ \theta_x$ taken at the points of Φ. In this way any stochastic process $K(x)$ compatible with the flow can be used to generate marks for a given point process Φ by taking $k_i := K(x_i)$. [2]

A marked version of a simple point process Φ compatible with the flow on $(\Omega, \mathcal{A}, \{\theta_t\})$, with marks generated by a random variable K can be also defined regardless of any numbering of points; cf Exercise 3.

Here is an example of some particular way of marking of points. We shall see later its theoretical utility.

Example 10.1.4 (Universal marks of a point process)**.** Let Φ be a simple point process compatible with the flow on the probability space $(\Omega, \mathcal{A}, \{\theta_t\})$ and define a marked version $\tilde{\Phi}$ of Φ with marks in $(\mathbb{K}, \mathcal{K}) = (\Omega, \mathcal{A})$ equal to $\theta_i := \theta_{x_i}$. We call $\{\theta_i\}$ the *universal sequence of marks* of the point process Φ; θ_i represents a flow-shilf on $(\Omega, \mathcal{A})$ by the random vector $-x_i \in \Phi$ thus making $0 \in (\Phi \circ \theta_i)(\omega)$ for all $\omega \in \Omega$; cf Exercise 4.

[1] A non-simple marked point process can be represented as a point process having simple atoms, with marks representing the multiplicity of points as well as their other characteristics.

[2] Recall that for any $x_i \in \Phi$ we have $0 \in \Phi \circ \theta_{x_i}$ and thus the marks k_i in (10.1.5) generated by the random variable K are invariant with respect to any modification of K outside $\Omega^0 := \{\omega \in \Omega : 0 \in \Phi(\omega)\}$. In other words, marks of a point process are generated by a random variable defined on Ω^0. However the stochastic process $K(x) = K \circ \theta_x$ requires K to be defined on Ω.

For a more thorough treatment of stationary marked point processes see the book project ((Baccelli et al., 2017, Chapter 8)).

Stationary probability

Let $(\Omega, \mathcal{A}, \{\theta_t\})$ be a measurable space with the flow. Let $\mathbf{P}$ be a probability measure on $(\Omega, \mathcal{A})$ invariant with respect to all elements of the flow

$$\mathbf{P}\theta_t^{-t} = \mathbf{P} \quad \text{for all } t \in \mathbb{R}^d; \tag{10.1.6}$$

that is $\mathbf{P}\{\,\omega : \theta_t(\omega) \in A\,\} = \mathbf{P}\{\,A\,\}$ for all $A \in \mathcal{A}$. We call $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$ a *stationary framework*.

The following result follows immediate from the above definition.

Fact 10.1.5. *Let Φ_i and X_i, $i = 1, 2, \ldots$ be a family of point processes and stochastic processes,respectively, defined on a stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$ compatible with the flow. Then Φ_i and X_i are* jointly stationary*, i.e.; the* joint *distribution of all these random objects is invariant with respect to any shift S_t*

$$\mathbf{P}_{(S_t\Phi_1, S_t\Phi_2, \ldots, S_tX_1, S_tX_2, \ldots)} = \mathbf{P}_{(\Phi_1, \Phi_2, \ldots, X_1, X_2, \ldots)} \quad \textit{for all } t \in \mathbb{R}^d\,.$$

We will call $\mathbf{P}$ the *stationary probability* on $(\Omega, \mathcal{A}, \{\theta_t\})$ to distinguish it form the Palm probabilities of different point processes to be defined on the same space.

Example 10.1.6. The distribution of a homogeneous Poisson process (having intensity $\Lambda(\mathrm{d}x) = \lambda \mathrm{d}x$) is invariant with respect to any shift S_t, $t \in \mathbb{R}^d$. The canonical probability space can serve as a stationary framework for it.

10.2 Palm probabilities in the stationary framework

From now on Φ will be a point process defined on the stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$ and compatible with the flow.

Fact 10.2.1. *The mean measure $M(\mathrm{d}x) = M_\Phi(\mathrm{d}x)$ of Φ is equal to the Lebesgue measure multiplied by a constant $M(\mathrm{d}x) = \lambda\,\mathrm{d}x$, with $0 \le \lambda \le \infty$. We call the constant $\lambda = \lambda_\Phi$ the* intensity *of the point process Φ.*

Proof. By the invariance of the probability $\mathbf{P}$ and the compatibility of Φ with the flow, M is invariant with respect to any shift S_t, $t \in \mathbb{R}^d$: for all $B \in \mathcal{B}$

$$\begin{aligned} S_t M(B) &= M(B + t) \\ &= \mathbf{E}\,[\Phi(B + t)] \\ &= \mathbf{E}\,[S_t\Phi(B)] \end{aligned}$$

$$\boxed{\text{compatibility of } \Phi} = \mathbf{E}\left[\Phi \circ \theta_t(B)\right]$$

$$\boxed{\text{invariance of } \mathbf{P}} = \mathbf{E}\left[\Phi(B)\right]$$

$$= M(B)\,.$$

The only measure on $(\mathbb{R}^d, \mathcal{B})$ that is invariant with respect to all shifts is a constant-multiple of the Lebesque measure. □

Following a similar line of thought as when defining the Palm distributions for a Point process on a general space (cf. Lesson on the Palm Theory for Point Processes) we define the following variant of the Campbell's measure. Let $\mathcal{C}$ be a measure on $(\mathbb{R}^d \times \Omega, \mathcal{B} \otimes \mathcal{A})$ defined on rectangles by

$$\mathcal{C}(B \times A) := \mathbf{E}\left[\int_{\mathbb{R}^d} \mathbf{1}(x \in B)\mathbf{1}(\theta_x \in A)\,\Phi(\mathrm{d}x)\right] \qquad B \in \mathcal{B}, A \in \mathcal{A}\,. \tag{10.2.1}$$

Remark 10.2.2. The measure $\mathcal{C}$ is called *Campbell-Matthes'* measure. Note the following two differences with respect to Campbell's measure defined in a general (non-stationary) context:

- Cambell-Matthes' measure is define on $\mathbb{R}^d \times \Omega$, not on $(\mathbb{E}, \mathcal{M})$. While $\mathbb{R}^d = \mathbb{E}$ for stationary processes, the difference consist in having the probability space Ω as the second component, not the space of realizations. This allows one to define the Palm probabilities (on the probability space) instead of the Palm distributions (on the space of realizations).

- There is a flow-shift θ_x in $\mathbf{1}(\theta_x \in A)$ in the definition of Cambell-Matthes' measure while for Campbell's measure we had just $\mathbf{1}(\Phi \in \Gamma)$, not $\mathbf{1}(S_x\Phi \in \Gamma)$. As a result we shall have just one Palm probability instead of a family indexed by x, as in the general case. Recall also from Example 10.1.4 for $x \in \Phi$, θ_x is a random shift, being the universal mark of x. See Exercise 11 for the consequences.

Very much as in the general context, the Campbell-Matthes' measure is the refinement of the mean measure $\mathcal{C}(\cdot \times \Omega) = M(\cdot)$ and for any $A \in \mathcal{A}$ we have $\mathcal{C}(\cdot \times A) \le M(\cdot)$. Thus, there exists a R-N derivative $\frac{\mathrm{d}\mathcal{C}(\cdot \times A)}{\mathrm{d}M(\cdot)}(x)$ of $\mathcal{C}(\cdot \times A)$ with respect to $M(\cdot)$. This time however this derivative (which in general case would be a measurable function on $\mathbb{R}^d$) is in fact a constant. Indeed, observe that similarly to $M(\cdot)$, $\mathcal{C}(\cdot \times A)$ is invariant with respect to all shifts on $\mathbb{R}^d$, cf Exercise 8. Hence $\mathcal{C}(\mathrm{d}x \times A) = Const_A \mathrm{d}x$, where $0 < Const_A < \infty$ is some constant. Consequently,

$$\frac{\mathrm{d}\mathcal{C}(\cdot \times A)}{\mathrm{d}M(\cdot)}(x) = \frac{Const_A\,\mathrm{d}x}{\lambda\,\mathrm{d}x}(x) = \frac{Const_A}{\lambda}\,,$$

which is well defined provided $0 < \lambda < \infty$. It is easy to see in this case that $Const_A/\lambda$, as a function of A, is a probability measure on $(\Omega, \mathcal{A})$, called the *Palm probablity* on Φ.

We can summarize the above construction in the following definition of the Palm probability, sometimes called *the Palm-Matthes definition*: Let Φ be a point process compatible

with the flow on the probability space $(\Omega, \mathcal{A}, \{\theta_t\})$, with finite, non-null intensity $0 < \lambda < \infty$. The Palm probability of Φ (or related to Φ) is the unique probability measure $\mathbf{P}^0$ on $(\Omega, \mathcal{A})$ given by

$$\mathbf{P}^0(A) = \frac{1}{\lambda|B|}\mathbf{E}\left[\int_{\mathbb{R}^d} \mathbf{1}(x \in B)\mathbf{1}(\theta_x \in A)\,\Phi(\mathrm{d}x)\right] \qquad A \in \mathcal{A}\,; \tag{10.2.2}$$

with any set $B \in \mathcal{B}$ of finite, non-null Lebesgue measure $|B|$. One can verify that $\mathbf{P}^0$ is indeed a probability measure and its value does not depend on the choice of the set B. In what follows we shall denote by $\mathbf{E}^0$ the expectation with respect to $\mathbf{P}^0$.

The following result is the variant of the Campbell-Little-Mecke result in our stationary setting.

Theorem 10.2.3 (Campbell-Little-Mecke-Matthes (CLMM))**.** *Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, compatible with the flow, and having finite, non-null intensity $0 < \lambda < \infty$. Denote by $\mathbf{P}^0$ the Palm probability of Φ. For any non-negative measurable functions f on $\mathbb{R}^d \times \Omega$ ([3]), we have*

$$\mathbf{E}\left[\int_{\mathbb{R}^d} f(x, \theta_x)\,\Phi(\mathrm{d}x)\right] = \lambda \int_{\mathbb{R}^d} \mathbf{E}^0\left[f(x, \omega)\right]\,\mathrm{d}x. \tag{10.2.3}$$

The result extends to all functions f for which either of the two sides of the equality (10.2.3) is finite when f is replaced by $|f|$.

Proof. Directly from the definition of $\mathbf{P}^0$ one easily shows the desired equality for $f(x, \omega) = \mathbf{1}(x \in B, \omega \in A)$ where $B \in \mathcal{B}$, $A \in \mathcal{A}$. By usual measure theoretic approximation arguments the result can be extended first to the case of $f(x, \omega) = \mathbf{1}((x, \omega) \in L)$, for arbitrary $L \in \mathcal{B} \otimes \mathcal{A}$, next to any simple function on $\mathbb{R}^d \times \Omega$ and finally to general functions whose integrals of the positive and negative part are finite. □

Using the above result one can easily show the following properties of $\mathbf{P}^0$.

Corollary 10.2.4. *Under the assumptions of Theorem 10.2.3, $\mathbf{P}^0$-almost surely $0 \in \Phi$.*

The point 0 of Φ under $\mathbf{P}^0$ is called the *typical point* of Φ. Several arguments will be given for this terminology including unimodularity discussed in Remark 10.2.9 and ergodicity discussed in a separate lesson.

Usually we assign to the typical point the index 0; thus $x_0 = 0$ under $\mathbf{P}^0$.

One can prove the following relation between the Palm probability $\mathbf{P}^0$ of Φ and its Palm distributions P_x on $\mathbb{M}$ defined in the general context.

Corollary 10.2.5. *Under the assumptions of Theorem 10.2.3, let $\mathbf{P}^0_\Phi$ be the distribution of Φ under its Palm probability $\mathbf{P}^0$ and P_x Palm distributions of Φ. Then*

$$\mathbf{P}^0_\Phi = P_x S_x^{-1} \qquad \text{for Lebesgue almost all } x \in \mathbb{R}^d\,.$$

[3] f is in fact a stochastic process but not necessarily compatible with the flow.

See Exercise 13 for the proof idea.

Using Corollary 10.2.5 one can conclude a version of the Slivnyak-Mecke's result characterizing Poisson point process.

Corollary 10.2.6. *A stationary point process with finite intensity is a Poisson point process iff its distribution under the Palm probability (considered in some stationary framework e.g. the canonical one) is equal to the distribution of $\Phi + \delta_0$ under the original stationary distribution.*

10.2.1 Mass transport formula and unimodularity for point processes

As we have said in the introduction, the true benefit form defining the Palm probability of a point process in the stationary framework is the possibility to study the relations between these probabilities corresponding to different point processes living on the same probability space and being *jointly* stationarity.

Here is one such result.

Theorem 10.2.7 (Mass transport formula for two point processes)**.** *Consider two point processes Φ and Φ' defined on a common stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$ and compatible with the flow, having non-null and finite intensities λ, λ', respectively. (We do not assume any particular dependence between these two point processes. In particular they might be independent or one can be a subset of the other, etc.) We denote by $\mathbf{P}^0$ and $\mathbf{P}^{0'}$ the respective Palm probabilities with respect to Φ and Φ'. For any (say non-negative) measurable functions g on $\mathbb{R}^d \times \Omega$ (not necessarily compatible with the flow) we have*

$$\lambda \mathbf{E}^0 \left[\int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y) \right] = \lambda' \mathbf{E}^{0'} \left[\int_{\mathbb{R}^d} g(-x, \theta_x)\, \Phi(\mathrm{d}x) \right] . \tag{10.2.4}$$

Proof. Let $B \in \mathcal{B}$ of unit Lebesgue measure, $|B| = 1$. We have

$$\begin{aligned}
\lambda \mathbf{E}^0 \left[\int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y) \right] &= \lambda \int_{\mathbb{R}^d} \mathbf{1}(x \in B) \mathbf{E}^0 \left[\int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y) \right] \mathrm{d}x \\
\boxed{\text{CLMM for } \Phi,\ f(x,\omega) = \mathbf{1}(x \in B) \textstyle\int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y)} &= \mathbf{E} \left[\int_{\mathbb{R}^d} \mathbf{1}(x \in B) \int_{\mathbb{R}^d} g(y, \theta_x)\, \Phi' \circ \theta_x(\mathrm{d}y)\, \Phi(\mathrm{d}x) \right] \\
\boxed{\text{by compatibility of } \Phi' \text{ and (10.1.3)}} &= \mathbf{E} \left[\int_{\mathbb{R}^d} \mathbf{1}(x \in B) \int_{\mathbb{R}^d} g(y - x, \theta_x)\, \Phi'(\mathrm{d}y)\, \Phi(\mathrm{d}x) \right] \\
\boxed{\text{Fubini's theorem}} &= \mathbf{E} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}(x \in B) g(y - x, \theta_x)\, \Phi(\mathrm{d}x)\, \Phi'(\mathrm{d}y) \right] \\
\boxed{\theta_{-y} \circ \theta_y = \theta_0 \text{ — indentity function}} &= \mathbf{E} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}(x \in B) g(y - x, \theta_{x-y} \circ \theta_y) \right. \\
&\qquad \left. \Phi \circ \theta_{-y} \circ \theta_y(\mathrm{d}x)\, \Phi'(\mathrm{d}y) \right]
\end{aligned}$$

$$\boxed{\text{CLMM for } \Phi' \\ f(x,\omega)=\int_{\mathbb{R}^d}\mathbf{1}(x\in B)g(y-x,\theta_{x-y})\,\Phi\circ\theta_{-y}(\mathrm{d}x)} = \lambda'\int_{\mathbb{R}^d}\mathbf{E}^{0'}\left[\int_{\mathbb{R}^d}\mathbf{1}(x\in B)g(y-x,\theta_{x-y})\Phi\circ\theta_{-y}(\mathrm{d}x)\right]\mathrm{d}y$$

$$\boxed{\text{by compatibility of } \Phi \text{ and (10.1.3)}} = \lambda'\int_{\mathbb{R}^d}\mathbf{E}^{0'}\left[\int_{\mathbb{R}^d}\mathbf{1}(x+y\in B)g(-x,\theta_x)\Phi(\mathrm{d}x)\right]\mathrm{d}y$$

$$\boxed{\text{Foubini's theorem}} = \lambda'\mathbf{E}^{0'}\left[\int_{\mathbb{R}^d}g(-x,\theta_x)\int_{\mathbb{R}^d}\mathbf{1}(x+y\in B)\,\mathrm{d}y\,\Phi(\mathrm{d}x)\right]$$

$$\boxed{\int_{\mathbb{R}^d}\mathbf{1}(x+y\in B)\,\mathrm{d}y=1 \text{ since } |B-y|=|B|=1} = \lambda'\mathbf{E}^{0'}\left[\int_{\mathbb{R}^d}g(-x,\theta_x)\,\Phi(\mathrm{d}x)\right].$$

□

We explain now why Theorem 10.2.7 is called mass transport principle.

Remark 10.2.8. *Equivalent form of the mass transport formula.* Let $m(x,y,\omega)$ be a measurable function on $\mathbb{R}^d\times\mathbb{R}^d\times\Omega$ interpreted as the amount of mass sent from x to y on the event ω. We assume that m is compatible with the flow in the following sense $m(x,y,\omega)=m(x-t,y-t,\theta_t)$ for all $x,y,t\in\mathbb{R}^d$. Define $g(y,\omega):=m(0,y,\omega)$ as the amount of mass sent form the origin 0 to y on the event ω. Then by the compatibility of m we have $m(0,-x,\theta_x)=m(x,0,\omega)$ and (10.2.4) is equivalent to

$$\lambda\mathbf{E}^0\left[\int_{\mathbb{R}^d}m(0,y,\omega)\,\Phi'(\mathrm{d}y)\right]=\lambda'\mathbf{E}^{0'}\left[\int_{\mathbb{R}^d}m(x,0,\omega)\,\Phi(\mathrm{d}x)\right], \tag{10.2.5}$$

which can be interpreted by saying that the proportion between the expected total mass sent from the typical point of Φ (located at the origin under $\mathbf{E}^0$) to all points of Φ' and the expected total mass received by the typical point of Φ' (located at the origin under $\mathbf{E}^{0'}$) from all points of Φ is equal to the proportion of the point processes intensities λ' to λ. In particular, if $\lambda=\lambda'$, on average, total mass sent out of the typical point of Φ is equal to the total mass received at the typical point of Φ'.

Remark 10.2.9. *Unimodulartiy of the Palm probability for point processes.* Assume now $\Phi'=\Phi$. Then equation (10.2.4) by $\lambda=\lambda'$ one obtains

$$\mathbf{E}^0\left[\int_{\mathbb{R}^d}g(y,\omega)\,\Phi(\mathrm{d}y)\right]=\mathbf{E}^0\left[\int_{\mathbb{R}^d}g(-x,\theta_x)\,\Phi(\mathrm{d}x)\right] \tag{10.2.6}$$

for any measurable functions g on $\mathbb{R}^d\times\Omega$, not necessarily compatible with the flow. Equivalently, (10.2.5) becomes

$$\mathbf{E}^0\left[\int_{\mathbb{R}^d}m(0,y,\omega)\,\Phi(\mathrm{d}y)\right]=\mathbf{E}^0\left[\int_{\mathbb{R}^d}m(x,0,\omega)\,\Phi(\mathrm{d}x)\right], \tag{10.2.7}$$

for any measurable function on $\mathbb{R}^d\times\mathbb{R}^d\times\Omega$ compatible with the flow. Observe that the completely analog mass transport formula was also used to characterize unimodular graphs

(cf Lesson on Unimodular Graphs). In the context of random graphs this property was used to formalize the notion of the rooted graph with the root interpreted as its typical vertex. Similarly, under the Palm probability the point at the origin can be seen as the typical point of the point process. We shall give more arguments for this statement in the lesson on ergodicity.

10.3 Exercises

1. Construct the canonical space with the flow supporting (i) two point processes compatible with the flow, (ii) one point process and one general stochastic process both compatible with the flow.

2. Verify that the marks k_i in (10.1.5) are random variables.

3. A marked version of a simple point process Φ compatible with the flow on $(\Omega, \mathcal{A}, \{\theta_t\})$, with marks generated by a random variable K can be defined regardless of any numbering of points as a point process $\tilde{\Phi}$ on $(\mathbb{R}^d \times \mathbb{K}, \mathcal{B} \otimes \mathcal{K})$ given by

$$\tilde{\Phi}(B \times M) = \int_B \mathbf{1}(K \circ \theta_x \in M) \Phi(\mathrm{d}x) \qquad B \in \mathcal{B},\ M \in \mathcal{K}\,.$$

   Observe that for any (say non-negative) function f on $\mathbb{R}^d \times \mathbb{K}$, $\mathcal{B} \otimes \mathcal{K}$-measurable

$$\int_{\mathbb{R}^d \times \mathbb{K}} f(x,k)\, \tilde{\Phi}(\mathrm{d}(x,k)) = \int_{\mathbb{R}^d} f(x, K \circ \theta_x)\, \Phi(\mathrm{d}x)$$

   and consequently for all $t \in \mathbb{R}^d$

$$\int_{\mathbb{R}^d \times \mathbb{K}} f(x,k)\, \tilde{\Phi} \circ \theta_t(\mathrm{d}(x,k)) = \int_{\mathbb{R}^d} f(x-t, K \circ \theta_{x-t})\, \Phi(\mathrm{d}x)\,.$$

4. Let Φ be a point process compatible with the flow on $(\Omega, \mathcal{A}, \{\theta_t\})$ and θ_i its universal sequence of marks as in Example 10.1.4. Show that 0 is the atom of $\Phi \circ \theta_i$ for all i and ω.

5. Prove Fact 10.1.5.

6. Construct a stationary framework supporting independently mareked homogeneous Poisson process.

7. Prove for a stationary point process Φ that its total number of points can by only zero or infinity, $\Phi(\mathbb{R}^d) \in \{0, \infty\}$ almost surely. In case $\lambda_\Phi > 0$ we have $\Phi(\mathbb{R}^d) = \infty$ almost surely.

8. Prove that for any $B \in \mathcal{B}$, $A \in \mathcal{A}$ and $t \in \mathbb{R}^d$, $\mathcal{C}(B \times A) = \mathcal{C}(B + t \times A)$.

9. Verify that the set function $\mathbf{P}^0$ given by (12.3.1) is a probability measure on $(\Omega, \mathcal{A})$ whose value does not depend on the choice of the set B, provided $0 < |B| < \infty$.

10. Let $\tilde{\Phi}$ be a marked point process with marks generated by random variable K with values in the measurable space $(K, \mathcal{K})$ as in Exercise 3. The *Palm distribution of the mark* of $\tilde{\Phi}$ is the probability measure P_K^0 on $(K, \mathcal{K})$ defined as

$$P_K^0(M) := \frac{1}{\lambda |B|} \mathbf{E}\left[\int_{\mathbb{R}^d \times K} \mathbf{1}(x \in B) \mathbf{1}(k \in M)\, \tilde{\Phi}(\mathrm{d}(x,k)) \right]$$

$$\boxed{\text{by Exercise 3}} = \frac{1}{\lambda|B|}\mathbf{E}\left[\int_{\mathbb{R}^d} \mathbf{1}(x \in B)\mathbf{1}(K\circ\theta_x \in M)\,\Phi(\mathrm{d}x)\right]$$

$$\boxed{\text{when } \Phi = \sum_i \delta_{x_i}} = \frac{1}{\lambda|B|}\mathbf{E}\left[\sum_{x_i\in B} \mathbf{1}(k_i \in M)\right] \qquad M \in \mathcal{K}$$

for any $B \in \mathcal{B}$ with $0 < |B| < \infty$. Note $\lambda|B| = M(B) = \mathbf{E}\left[\Phi(B)\right]$ and thus P_K^0 represents some averaged distribution of the marks k_i of points x_i in a set B. More precisely, this is an averaging in expectation, since we have $\mathbf{E}\left[\cdot\right]$ in the numerator and the denominator of the expression. We shall see later in the ergodic case, that this expectation can be removed, i.e., the empirical average can be considered, asymptotically when B is approaching $\mathbb{R}^d$ in some nice way. The averaging interpretation explains why P_K^0 is also called the *distribution of the typical mark* — meaning the mark of an "arbitrary" point of Φ selected without any bias. Note, formally one cannot select uniformly a point out of the infinitely countable set of point of Φ.

11. Observe that the Palm probability $\mathbf{P}^0$ of Φ is the Palm distribution of the universal mark of Φ.

12. Prove Corollary 10.2.4.

13. Prove Corollary 10.2.5 showing that for any bounded $B \in \mathcal{B}$ and $\Gamma \in \mathcal{M}$

$$\int_B \Big(P_x(\{\mu : S_x\mu \in \Gamma\}) - \mathbf{P}^0\{\,\Phi \in \Gamma\,\}\Big)\,\mathrm{d}x = 0\,.$$

Use the general Campbell-Little-Mecke' formula and Campbell-Little-Mecke-Matthes' one for the respective expressions parts of the expression.

14. Prove Corollary 10.2.6.

15. Using unimodularity property (10.2.7) argue that any translation invariant graph with the vertexes at the points of a simple stationary point processes on $\mathbb{R}^d$ considered under its Palm probability, with $X_0 = 0$ considered as its root, is unimodular in the sense considered in the Lesson on Unimodular Graphs.

16. Using unimodularity property prove for any translation invariant, directed graph on a simple, stationary point process that the expected in-degree of the typical node is equal to the expected out degree of the typical node. Construct an example showing that in general there is no equality of the distributions of these two degrees. Further, show that the sum of the in-degrees of nodes in some given set and the sum of the out-degrees of these nodes have the same stationary expectation.

17. Derive from the Campbell-Little-Mecke-Matthes' formula the following *mass transport formula between point process* Φ *and Lebesgue measure*:

$$\lambda\mathbf{E}^0\left[\int_{\mathbb{R}^d} m(0,y,\omega)\,\mathrm{d}y\right] = \mathbf{E}\left[\int_{\mathbb{R}^d} m(x,0,\omega)\,\Phi(\mathrm{d}x)\right], \tag{10.3.1}$$

with m as in Remark 10.2.9.

18. *Little's low* is a general law in queueing theorysaying that *in a stable system the average number of customers at a given time in service is equal to the average number of customers arriving to the system per unit of time multiplied by the mean sojourn time of the typical customer*. This can be formalized as follows. Let $\tilde{\Phi} = \sum_i \delta_{(x_i,w_i)}$ by a marked point process on $\mathbb{R} \times \mathbb{R}^+$, with x_i representing arrival times of customers to the system and w_i their sojourn (eventual waiting plus service) times in the system. Let $N(x)$ be a stochastic process with values in $0, 1, 2, \ldots$ describing the number of customers in the system at time x. We assume that $\tilde{\Phi}$ and N are jointly stationary and consider them defined on some stationary framework. In particular $N(x) = N \circ \theta_x$ and $w_i = W \circ \theta_{x_i}$ for some random variables N and W. Denote by $\mathbf{P}^0$ the Palm probability with respect to the arrival process $\Phi = \sum_i \delta_{x_i}$, which we assume simple, with non-null and finite intensity λ. Then

$$\mathbf{E}[N] = \mathbf{E}[N(0)] = \lambda \mathbf{E}^0[W_0] = \lambda \mathbf{E}^0[W] .$$

Prove the above Little's law using CLMM interpreted as a mass transport formula (10.3.1). *Hint:* consider the function $m(x, y, \omega) = \mathbf{1}(x \in \Phi)\mathbf{1}(x \le y \le x + W \circ \theta_x)$. In words: at the arrival time $x = x_i$ of any customer, a unit mass is sent to all time instants y between the arrival and the exit time of this customer from the system (which is equal to $x_i + w_i$).

19. *Computer exercise.* (*Ripley's function and Poisson hypothesis testing*) The *Ripley's* $K(r)$ function is defined for a stationary point process on $\mathbb{R}^d$ as the expected number of additional points within the distance r from the typical point (located at the origin under the Palm probability $\mathbf{P}^0$) divided by λ

$$K(r) := \frac{\mathbf{E}^0[\Phi(B_0(r))] - 1}{\lambda} .$$

(a) Using CLMM theorem prove that

$$K(r) = \frac{1}{\lambda^2 |W|} \mathbf{E}\left[\int_{\mathbb{R}^d} \mathbf{1}(x \in W) \left(\Phi(B_x(t)) - 1\right) \Phi(\mathrm{d}x) \right] \tag{10.3.2}$$

for any bounded set W with non-null volume $|W|$.

(b) Using expression (10.3.2) justify the following estimator $\hat{K}$ of the function K

$$\hat{K}(r) := \frac{\sum_{X_i \neq X_j} \mathbf{1}(|X_i - X_j| \le r)}{n\hat{\lambda}}$$

of the point process $\Phi = \{X_i : i = 1, \ldots, n\}$ observed in the window W, where n is the number of points in the window and $\hat{\lambda} := n/|W|$ is the estimator of the intensity λ of this process.

(c) Using Slivnyak-Meckes's result (Corollary 10.2.6) prove for a homogeneous Poisson point process that $K(r) = \pi r^2$.

(d) $L(r)$ Ripley's function is defined as $L(r) := \sqrt{K(r)/\pi}$ and its estimator $\hat{L}(r) := \sqrt{\hat{K}(r)/\pi}$. Note that for the homogeneous Poisson process $L(r) = r$. Knowing that Slivnyak-Meckes's result characterizes Poisson process, suggests the following strategy of Poisson hypothesis testing: Observing significantly $\hat{L}(r) \not\approx r$ indicates that one cannot fit the data with Poisson process. If $\hat{L}(r) \approx r$ one can fit Poisson process to the data (at least there is no reason to reject this hypothesis).

(e) Among the data sets provided with *spatstat* [4] find one for which Poisson hypthesis can be retained and one non-Poisson data. Plot the data and use Ripley's function to test the hypothesis.

[4] *Spatstat*, `http://spatstat.org/`, R package for spatial statistics. For the package reference guide see Package ((2017)). See Baddeley et al. for the description of the data sets.

Lesson 11

Stationary Voronoi tessellation

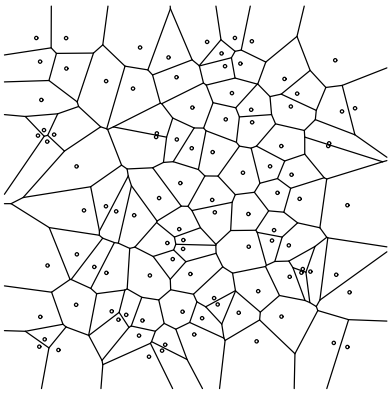


Figure 11.1: An illustration of the Voronoi tessellation generated be a finite pattern of points on the plane (some cells at the boundary are unbounded). It is named after Georgy Voronoi, an Ukrainian mathematician and sometimes also associated to a German mathematician Peter Gustav Lejeune Dirichlet. It has practical and theoretical applications in a large number of fields, mainly in science and technology, but also in visual art.

11.1 Voronoi Tessellation

Let $\Phi = \sum_i \delta_{x_i}$ be a simple point process on $\mathbb{R}^d$. The polygon

$$C_i := \{y \in \mathbb{R}^d : |y - x_i| \leq \min_{x_j \in \Phi} |y - x_j|\} \tag{11.1.1}$$

is called the *Voronoi cell* of x_i. (Note that the local finiteness of Φ justifies the usage of "min" in (11.1.1).) The collection of cells C_i is called *Voronoi tessellation* (or *mosaic*) generated by Φ. The interiors of cells are pairwise disjoint and the union of cells covers $\mathbb{R}^d$. It is an example of a tessellation of $\mathbb{R}^d$.

In this lesson we shall consider Voronoi tessellation generated by a stationary point process as a marked version of this process, with marks corresponding to the Voronoi cells (polygons) shifted (centered) at the origin. Our principal goal is to present some conservation laws derived from Campbell-Little-Mecke-Matthes' formula and the Mass Transport Principle (cf the lesson on stationary point processes). These results are of more general nature and Voronoi tessellation is here only an example of a stationary mosaic. For a supplementary material to this lesson see the working book project ((Baccelli et al., 2017, Section 7.2 and 7.3)).

11.1.1 Voronoi tessellation as a stationary marking

Throughout the whole lesson Φ is a simple, stationary point process in $\mathbb{R}^d$, compatible with the flow on a stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, having non-null, finite intensity $0 < \lambda = \mathbf{E}\left[\Phi([0,1]^d)\right] < \infty$. We denote by $\mathbf{P}^0$ its Palm probability.

Define a *random set closed set* [1] (related to Φ)

$$V := \{y \in \mathbb{R}^d : |y| \leq \min_{z\in\Phi} |y - z|\}\,. \tag{11.1.2}$$

Consider the random process (with values in the space of closed sets), compatible with the flow, generated by V

$$V(x) := V \circ \theta_x \qquad x \in \mathbb{R}^d \tag{11.1.3}$$

and the marks of Φ generated by V (cf Section 1.1.4 in the Lesson on Stationary Point Processes)

$$V_i := V \circ \theta_{x_i} \qquad \text{for } x_i \in \Phi\,. \tag{11.1.4}$$

Observe

$$\begin{aligned} V(x) &= \{y \in \mathbb{R}^d : |y| \leq \min_{z\in\Phi\circ\theta_x} |y - z|\} \\ \boxed{\text{compatibility of } \Phi} &= \{y \in \mathbb{R}^d : |y| \leq \min_{z\in\Phi} |y - (z - x)|\} \\ \boxed{y' := x + y} &= \{y' - x \in \mathbb{R}^d : |y' - x| \leq \min_{z\in\Phi} |y' - z|\} \\ &= \{y' :\in \mathbb{R}^d : |y' - x| \leq \min_{z\in\Phi} |y' - z|\} - x\,. \end{aligned}$$

Consequently, for $x_i \in \Phi$

$$V_i = V(x_i) = C_i - x_i \tag{11.1.5}$$

is the Voronoi cell of x_i translated (shifted) by the vector $-x_i$ (we say also "centered at the origin"). More generally, $V(x)$ is the Voronoi cell of a hypothetical point x (it needs not to be in Φ) generated in the presence of points of Φ and translated by $-x$.

[1] We shall see in a different lesson how to formalize the notion of random closed set.

The Voronoi tessellation generated by a stationary point process $\Phi = \sum_i \delta_{x_i}$ can be seen as the following marked version of Φ

$$\tilde{\Phi} := \sum_i \delta_{(x_i, V_i)}$$

with marks V_i in the space of closed subsets of $\mathbb{R}^d$. Remember, one recovers the original (non-centered) Voronoi cell of $x_i \in \Phi$ by taking $C_i = V_i + x_i$.

11.2 The inverse formula of Palm calculus

The random closed set V can be used to express the original stationary probability $\mathbf{P}$ in terms of the Palm probability $\mathbf{P}^0$ of the point process Φ.

Theorem 11.2.1 (Inverse formula of Palm calculus)**.** *Let Φ be a simple, stationary point process in $\mathbb{R}^d$, compatible with the flow on a stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, having non-null, finite intensity $0 < \lambda = \mathbf{E}\left[\Phi([0,1]^d)\right] < \infty$. We denote by $\mathbf{P}^0$ its Palm probability. For any measurable function on $(\Omega, \mathcal{A})$, integrable with respect to $\mathbf{E}$, we have*

$$\mathbf{E}\left[f\right] = \lambda \int_{\mathbb{R}^d} \mathbf{E}^0\left[\mathbf{1}(x \in V) f \circ \theta_x\right] \mathrm{d}x\,. \tag{11.2.1}$$

Remark 11.2.2. Observe, in the inverse formula the stationary probability is recovered by Lebesgue averaging of the Palm expectation of the stochastic process $f \circ \theta_x$ over the Voronoi cell $V = V(0)$ of the origin 0, which is called the typical point of Φ under $\mathbf{P}^0$. While in the direct counterpart of this formula, i.e., the Palm-Matthes definition of $\mathbf{P}^0$ (cf equation (2.2) in the Lesson on Stationary Point Processes), this latter probability is obtained by discrete-averaging of the stationary expectation of the marks $f \circ \theta_{x_i}$ over points of Φ in an arbitrary deterministic set B.

We need the following auxiliary result, which is an independent interest.

Lemma 11.2.3. *Under the assumptions of Theorem 11.2.1, $\mathbf{P}$-almost surely Φ does not have points equidistant to the origin.*

Proof.

$$\begin{aligned}
&\mathbf{P}\{\,\exists\, x_i, x_j \in \Phi, x_i \neq x_j, |x_i| = |x_j|\,\} \\
&\qquad \overset{\text{Markov inequality}}{\leq} \mathbf{E}\left[\int_{\mathbb{R}^d} \Phi(\{y : y \neq x, |y| = |x|\})\, \Phi(\mathrm{d}x)\right] \\
&\qquad = \mathbf{E}\left[\int_{\mathbb{R}^d} \Phi \circ \theta_{-x+x}(\{y : y \neq x, |y| = |x|\})\, \Phi(\mathrm{d}x)\right] \\
&\qquad \overset{\text{CLMM}}{=} \lambda \int_{\mathbb{R}^d} \mathbf{E}^0\left[\Phi \circ \theta_{-x}(\{y : y \neq x, |y| = |x|\})\right] \mathrm{d}x
\end{aligned}$$

$$\boxed{\text{compatibility of } \Phi} = \lambda \int_{\mathbb{R}^d} \mathbf{E}^0 \left[\Phi(\{y-x : y \neq x, |y| = |x|\})\right] \mathrm{d}x$$

$$\boxed{y' := y - x} = \lambda \int_{\mathbb{R}^d} \mathbf{E}^0 \left[\Phi(\{y' : y' + x \neq x, |y' + x| = |x|\})\right] \mathrm{d}x$$

$$= \lambda \int_{\mathbb{R}^d} \mathbf{E}^0 \left[\int_{\mathbb{R}^d} \mathbf{1}(y' \neq 0)\mathbf{1}(|y' + x| = |x|)\, \Phi(\mathrm{d}y')\right] \mathrm{d}x$$

$$\boxed{\text{Fubini's theorem}} = \lambda \mathbf{E}^0 \left[\int_{\mathbb{R}^d} \mathbf{1}(y' \neq 0) \int_{\mathbb{R}^d} \mathbf{1}(|y' + x| = |x|)\, \mathrm{d}x\, \Phi(\mathrm{d}y')\right]$$

$$\boxed{\int_{\mathbb{R}^d} \mathbf{1}(|y' + x| = |x|)\, \mathrm{d}x = 0 \text{ for } y' \neq 0} = 0\,.$$

□

Proof of Theorem 11.2.1. We shall use Campbell-Little-Mecke-Matthes' formula interpreted as the mass transport principle between point process and Lebesgue measure, presented in Exercise 17 in the Lesson on Stationary Point Processes. Let

$$m(x, y, \omega) := f \circ \theta_y \mathbf{1}(x \in \Phi)\mathbf{1}(y - x \in V(x))\,, \tag{11.2.2}$$

where $V(x)$ is defined in (11.1.3). That is, the mass $f \circ \theta_y$ is sent from every point of $x \in \Phi$ to all locations y in the Voronoi cell C of x. Indeed $y - x \in V(x)$ iff $y \in V(x) + x$ the (non-centered) Voronoi cell of x; cf (11.1.5). It is easy to verify that m is compatible with the flow; cf Exercise 1. Then

$$\mathbf{E}^0 \left[\int_{\mathbb{R}^d} m(0, y, \omega)\, \mathrm{d}y\right] = \mathbf{E}^0 \left[\int_{\mathbb{R}^d} \mathbf{1}(y \in V) f \circ \theta_y\, \mathrm{d}y\right]$$

is the total mass sent from $0 \in \Phi$ under $\mathbf{P}^0$. Moreover, by Lemma 11.2.3 under $\mathbf{P}$ there is exactly one point $x \in \Phi$ sending mass to the origin and thus

$$\mathbf{E} \left[\int_{\mathbb{R}^d} m(x, 0, \omega)\, \Phi(\mathrm{d}x)\right] = \mathbf{E}\left[f \circ \theta_0\right] = \mathbf{E}\left[f\right]$$

is the total mass received at the origin under $\mathbf{P}$. The formula (11.2.5) follows directly from equation (3.1) of Exercise 17 in the Lesson on Stationary Point Processes. □

11.2.1 Typical versus zero cell of the tessellation

The Voronoi cell $V = V(0)$ of the point $0 \in \Phi$, considered under $\mathbf{P}^0$, is called the *typical cell* of the Voronoi tessellation generated by Φ; it is the Voronoi cell of the typical point located at the origin 0 under $\mathbf{P}^0$. It may be seen as a formalization of the idea of the Voronoi cell of randomly selected (without any bias) point of Φ. [2]

[2] Note there is no uniform distribution on the countably infinite set and hence a direct formalization of randomly, uniformly selected point is not possible without asymptotic (ergodic) argument; cf the Lesson on Ergodic Theory for Point Processes.

Corollary 11.2.4. *Under the assumptions of Theorem 11.2.1, taking $f \equiv 1$, we obtain*

$$\mathbf{E}^0[V] = \frac{1}{\lambda}. \tag{11.2.3}$$

Remark 11.2.5. Recall $\lambda = \mathbf{E}\left[\Phi([0,1]^d)\right]$. An intuitive explanation of (11.2.3), which is a very general relation regarding a stationary tessellation. It can be paraphrased is as follows: When λ is the average number of points (and hence cells) per unit volume, the inverse $1/\lambda$ must be the average volume of a cell. Remark the two averages in this statement correspond to two different probabilities.

Another way of selecting some cell of the Voronoi tessellation consists in taking the one covering some given fixed location that can be assumed without loss of generality to be the origin. We can guess that this way of sampling a cell introduces a bias towards cells of larger volume.

Formally, by Lemma 11.2.3 under $\mathbf{P}$ there is exactly one point of the point process Φ whose cell covers 0, let us denote this point x_* ([3]) Let us denote this cell by C_*. The cell C_* considered under $\mathbf{P}$ is called the *zero cell* of the tessellation. To be consistent with the previous analysis, denote by

$$V_* := V(x_*) := C_* - x_*$$

the *centered zero cell*. Using the Inverse Formula, the distribution of the centered zero cell under $\mathbf{P}$ can be compared to the distribution of the typical cell V under $\mathbf{P}^0$. (Note $C_* = V$ under $\mathbf{P}^0$).

Proposition 11.2.6. *Under the assumptions of Theorem 11.2.1, let g be a (closed) set function. Then*

$$\mathbf{E}[g(V_*)] = \lambda \mathbf{E}^0[|V|\, g(V)] = \frac{\mathbf{E}^0[|V|\, g(V)]}{\mathbf{E}^0[|V|]}. \tag{11.2.4}$$

Remark 11.2.7. Note the right-hand side of 11.2.4 is the so called size biased modification of the distribution of the typical cell. In particular, taking $g(B) = 1/|B|$, and using Jensen's inequality one obtains

$$\frac{1}{\mathbf{E}[|V_*|]} \leq \mathbf{E}\left[\frac{1}{|V_*|}\right] = \frac{1}{\mathbf{E}^0[|V|]}$$

or, equivalently,

$$\mathbf{E}[|C_*|] = \mathbf{E}[|V_*|] \geq \mathbf{E}^0[|V|],$$

that is the zero-cell has larger (non-smaller) mean volume than the typical cell. This is a multidimensional version of *Feller's paradox* (called also the *waiting time paradox*, cf Exercise 3).

Proof of Proposition 11.2.6. In order to use the Inverse Formula (11.2.5) we need to ensure that V_* is well defined for all $\omega \in \Omega$ (not only $\mathbf{P}$-almost surely). In this regard take for example the following specification of V_*: $V_* = V \circ \theta_{x'_*}$, where $x'_* := x_*$ provided x_* is well defined (as

[3] Using the function h introduced in Exercise 2, $x_*(\omega) \in \Phi$ such that $h(x_*, \omega) = 1$.

the unique point of Φ the closest to the origin) and $x'_* = 0$ otherwise. By Lemma 11.2.3 $\mathbf{E}[g(V_*)] = \mathbf{E}[V \circ \theta_{x_*}]$ and using (11.2.5) we only need to show that $V_* \circ \theta_x = V$, $\mathbf{P}^0$ almost surely for x in the interior of V. Observe in this regard

$$\begin{aligned} V_* \circ \theta_x &= V \circ \theta_{x'_* \circ \theta_x} \theta_x \\ &= V \circ \theta_{x'_* \circ \theta_x + x} \\ \boxed{(*)} &= V \circ \theta_0 \\ &= V \,, \end{aligned}$$

where $(*)$ follows from the observation that $x'_* \circ \theta_x + x$ is the closest point of Φ to x provided it is unique (hence 0 for x in the interior of V), and x otherwise. □

11.2.2 Neveu exchange formula

Similar ideas allow one to relate Palm probabilities of two point processes defined in the same stationary framework.

Theorem 11.2.8 (Neveu exchange formula)**.** *Let Φ, Φ' be two simple, stationary point process in $\mathbb{R}^d$, compatible with the flow on the same stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, having non-null, finite intensities $0 < \lambda, \lambda' < \infty$, respectively. We denote by $\mathbf{P}^0$ and $\mathbf{P}^{0'}$ their Palm probabilities, respectively. Assume that $\mathbf{P}^{0'}$ almost surely there are no points of Φ equidistant to the origin. Then, for any measurable function on $(\Omega, \mathcal{A})$, integrable with respect to $\mathbf{E}^{0'}$, we have*

$$\lambda' \mathbf{E}^{0'}[f] = \lambda \mathbf{E}^0 \left[\lambda \int_{\mathbb{R}^d} \mathbf{1}(x \in V) f \circ \theta_x \, \Phi'(\mathrm{d}x) \right], \tag{11.2.5}$$

where V is the Voronoi cell of the origin generated by Φ, given by (11.1.2).

Remark 11.2.9. Condition that $\mathbf{P}^{0'}$ almost surely there are no points of Φ equidistant to the origin is equivalent to

$$\mathbf{E}^0 \left[\Phi'(\partial V) \right] = 0 \,;$$

i.e., that $\mathbf{P}^0$ almost surely the are not points of Φ' on the boundary of the set V; cf Exercise 4.

Proof of Theorem 11.2.8. Perhaps the easiest way of proving the Neveu exchange formula consist in using the Mass Transport Principle in its version expressed in equation (1.11) in the Lesson on Stationary Point processes. Indeed, let

$$m(x, y, \omega) := f \circ \theta_y \mathbf{1}(x \in \Phi) \mathbf{1}(y \in \Phi') \mathbf{1}(y - x \in V(x)) \,, \tag{11.2.6}$$

where $V(x)$ is defined in (11.1.3) (with respect to Φ). That is, the mass $f \circ \theta_y$ is sent from every point of $x \in \Phi$ to all points $y \in \Phi'$ in the Voronoi cell C of x generated by Φ. Indeed $y - x \in V(x)$ iff $y \in V(x) + x$, the (non-shifted) Voronoi cell of x; cf (11.1.5). Then

$$\mathbf{E}^0 \left[\int_{\mathbb{R}^d} m(0, y, \omega) \, \Phi'(\mathrm{d}y) \right] = \mathbf{E}^0 \left[\int_{\mathbb{R}^d} \mathbf{1}(y \in V) f \circ \theta_y \, \Phi'(\mathrm{d}y) \right]$$

is the total mass sent from $0 \in \Phi$ under $\mathbf{P}^0$. Moreover, by the assumption, under $\mathbf{P}^{0'}$ there is exactly one point $x \in \Phi$ sending mass to the origin and thus

$$\mathbf{E}^{0'}\left[\int_{\mathbb{R}^d} m(x,0,\omega)\,\Phi(\mathrm{d}x)\right] = \mathbf{E}^{0'}\left[f\circ\theta_0\right] = \mathbf{E}^{0'}\left[f\right]$$

is the total mass received at the origin under $\mathbf{P}$. The formula (11.2.5) follows directly from equation (1.11) in the Lesson on Stationary Point Processes. □

TODO

- Mean formulas for planar VT.
- Perfect sampling of the typical cell.

11.3 Exercises

1. Show that the mass transport funcition m given in (11.2.2) satisfies $m(x-t, y-t, \theta_t) = m(x, y, \omega)$ for all $x, y, t \in \mathbb{R}^d$ and $\omega \in \Omega$.

2. Define function $h(x, \omega) := \mathbf{1}(\Phi(B_0^o(|x|)) = 0)$, where $B_a^o(r)$ is the open ball centered at a of radius r. Observe that Lemma 11.2.3 implies $\int_{\mathbb{R}^d} h(x, \omega)\, \Phi(\mathrm{d}x) = 1$, $\mathbf{P}$-almost surely. Write $\mathbf{E}[f] = \mathbf{E}\left[f \times \int_{\mathbb{R}^d} h(x, \omega)\, \Phi(\mathrm{d}x)\right]$ and using Campbell-Little-Mecke-Matthes' formula prove (11.2.5).

3. Using Remark 11.2.7 formalize and explain the following statement called *Feller's* or *waiting time paradox*: "Passengers arriving at a bus-stop see statistically larger bus inter-arrival times then the ones advertised by the bus company".

4. Using Mass Transport Principle prove Remark 11.2.9.

5. *Stationary renewal point process.* Renewal point process is a generalization of the homogeneous Poisson point process on the line $\mathbb{R}$. It is more easy to define first the *non-delayed* version of the renewal process. In this case t here is a point of the process at the origin and the distances between all consecutive points (in the negative and positive half-line) are assumed to be independent with some general distribution; denote its distribution function by F. Observe that in the case of exponential F the above description of the non-delayed renewal process coincides with the description of the Palm distribution of the homogeneous Poisson point process. Also, by Slivnyak-Mecke's theorem, one knows that forgetting about the point at the origin yields the stationary Poisson process. Regarding the general renewal process, we may ask what modification of the non-delayed renewal process (besides the removal of the point at the origin) makes it stationary. The solution to this question is an appropriate *stationary delaying* of the points in the positive and negative parts, described in what follows.

   Consider a simple, stationary point process $\Phi = \sum_{n\in\mathbb{Z}} \delta_{x_n}$ on $\mathbb{R}$ of non-null, finite intensity $\lambda = \left(\int_0^\infty u F(\mathrm{d}u)\right)^{-1}$. We assume the points of Φ are numbered such that $x_n < x_{n+1}$ and $x_0 = \max\{x \in \Phi, x \le 0\}$. Denote $T_n := x_{n+1} - x_n$ for $n \in \mathbb{Z} \setminus \{0\}$. Denote by $\mathbf{P}^0$ the Palm probability of Φ (say on the canonical stationary framework). Assume that under $\mathbf{P}^0$

$$x_0 = 0,\ x_1 \text{ has distribution } F \text{ and is independent of } \{T_n\}_{n\neq 0}.$$

   (This is the *non-delayed* renewal process with generic inter-point distance T of distribution function F.) Using the Inverse Formula of Palm calculus show that under stationary probability $\mathbf{P}$ $\{T_n\}_{n\in\mathbb{Z}\setminus\{0\}}$ are also i.i.d. non-negative r.v. with distribution function F, the pair (x_0, x_1) is independent of of $\{T_n\}_{n\neq 0}$ and has the joint distribution

$$\mathbf{P}\{-x_0 > s, x_1 > t\} = \frac{\int_{s+t}^\infty (1 - F(u))\, \mathrm{d}u}{\int_0^\infty (1 - F(u))\, \mathrm{d}u}\,.$$

Hint: With $T := x_1 \sim F$ under $\mathbf{P}^0$

$$\begin{aligned}\mathbf{P}\{-x_0 > s, x_1 > t\} &= \lambda \mathbf{E}^0 \left[\int_{x_{-1}/2}^{x_1/2} \mathbf{1}(-x_0 \circ \theta_u > s) \mathbf{1}(x_1 \circ \theta_u > t) \, \mathrm{d}u \right] \\ &= \lambda \mathbf{E}^0 \left[\int_{-T_{-1}/2}^{0} \mathbf{1}(T_{-1} + u > s) \mathbf{1}(-u > t) \, \mathrm{d}u \right] \\ &\quad + \lambda \mathbf{E}^0 \left[\int_{0}^{T/2} \mathbf{1}(u > s) \mathbf{1}(T - u > t) \, \mathrm{d}u \right] \\ \boxed{T_{-1} \stackrel{\mathcal{L}}{=} T \text{ under } \mathbf{P}^0} &= \lambda \mathbf{E}^0 \left[\int_{0}^{T} \mathbf{1}(u > s) \mathbf{1}(T - u > t) \, \mathrm{d}u \right].\end{aligned}$$

6. *Computer exercise.* Voronoi (called also Dirichlet) tessellation of a planar point set (with respect to the entire plane) can be calculated using *deldir* R package [4]. It also calculates the *Delaunay triangulation* on the plane, which is obtained connecting by segments the points, whose Voronoi cells share some boundary edge, see Figure 11.2. Also calculates the perimeters of tessellations, summarizes information about the tiles of the tessellation.

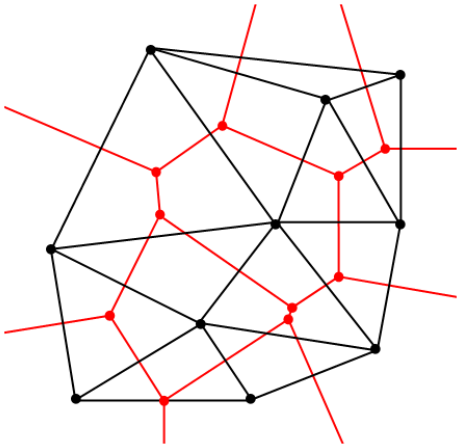


Figure 11.2: In red: Voronoi cells on black points. In black: Delaunay triangles.

(a) Using *spatstat* generate a Poisson point process in some regular window (disk or rectangle) having a reasonably large number of points.

[4] *deldir* see https://cran.r-project.org/web/packages/deldir/index.html. For the package reference guide see Turner ((2017)).

(b) Using *deldir* calculate the Voronoi tessellation for this Poisson process.

(c) Verify numerically the following theoretical result regarding the Voronoi tessellation of a homogeneous Poisson process on the plane (holds also for a general ergodic point process, under some additional conditions):

i. The typical Voronoi cell has on average 6 edges.

ii. On average, there are two times more Delaunay triangles than Voronoi cells. More formally: the intensity of triangle centers (red points on Figure 11.2, note that the triangle centers are not always inside triangles) is 2 times bigger than the intensify of Voronoi cell centers (black points on Figure 11.2).

For the verification use the information about the tiles of the tessellation provided by *deldir*.

Lesson 12

Ergodicity and point-shift invariance

Ergodic theory bridges the gap between the probability theory and real-life applications. Probability space is a mathematical abstraction meant to represent all physical circumstances, escaping from exact analysis, leading to some particular observations, which are modeled by random variables. In other words, in probabilistic modeling we consider some observations as random, since we are not able, or simply do not want, to describe and examine the exact reasons for these observations to appear. If the probabilistic model is supposed to say something pertinent to reality, theoretically calculated expectations (which are integrals over the probability space!) should be related to the empirical averages calculated with respect to real observations.

Applications		**Probability**
observations $X_1, \ldots, X_n \in \mathbb{R}$		random variable $X = X(\omega)$ on some abstract probability space $(\Omega, \mathcal{A}, \mathbf{P})$
	Ergodicity	
mean: $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$	$\underset{n\to\infty}{\longrightarrow}$	$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\,\mathbf{P}(\mathrm{d}\omega)$; expectation

Ergodic theory provides precise conditions for the above converge result, thus bridging the gap between the probability theory and real-life applications. It is particularly important in statistics.

In the first part of this lesson we show how this theory manifests in case when observed random elements are stationary point processes. Remember, in this case one considers two mathematical expectations: the stationary $\mathbf{E}$ and the Palm $\mathbf{E}^0$ one, with two types of corresponding empirical averages.

Applications	**Probability**
"homogeneous" pattern $\Phi = \{x_1, \dots, x_n\}$ of points in some observation window $B \subset \mathbb{R}^d$	stationary point process $\Phi(\omega)$ on $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, of intensity λ and Palm probability $\mathbf{P}^0$, modeling the observations;

Continuous ergodicity

Averaging observations $X(x) = X(x, \Phi) = f(S_x\Phi)$ of Φ *from all locations* $x \in B$ in the window	*Stationary expectation* of the observation $X(0) = X(0, \Phi(\omega)) = f(\Phi(\omega))$ of Φ from the origin

$$\frac{1}{|B|}\int_B X(x)\,\mathrm{d}x = \frac{1}{|B|}\int_B f(S_x\Phi)\,\mathrm{d}x \qquad \underset{B\nearrow\mathbb{R}^d}{\longrightarrow} \qquad \mathbf{E}\left[X(0)\right] = \mathbf{E}\left[f(\Phi)\right]$$

Observe by the invariance of $\mathbf{P}$ that on average

$$\frac{1}{|B|}\mathbf{E}\left[\int_B f(S_x\Phi)\,\mathrm{d}x\right] = \frac{1}{|B|}\int_B \mathbf{E}\left[f(S_x\Phi)\right]\,\mathrm{d}x = \mathbf{E}\left[f(\Phi)\right]$$

Discrete ergodicity

Averaging observations $X(x_i) = X(x_i, \Phi) = f(S_{x_i}\Phi)$ of Φ *from all (discrete) points of* Φ $x_i \in \Phi \cap B$ in the window	*Palm expectation* of the observation $X(0) = X(0, \Phi(\omega)) = f(\Phi(\omega))$ of Φ from the typical point at the origin

$$\frac{1}{\Phi(B)}\sum_{x_i\in\Phi\cap B} X(x_i) = \frac{1}{\Phi(B)}\sum_{x_i\in\Phi\cap B} f(S_{x_i}\Phi) \qquad \underset{B\nearrow\mathbb{R}^d}{\longrightarrow} \qquad \mathbf{E}^0\left[X(0)\right] = \mathbf{E}^0\left[f(\Phi)\right]$$

Observe by the CLMM theorem that on average

$$\mathbf{E}\left[\sum_{X_i\in\Phi\cap B} f(S_{X_i}\Phi)\right] = \lambda|B|\,\mathbf{E}^0\left[f(\Phi)\right] \quad \text{and} \quad \mathbf{E}\left[\Phi(B)\right] = \lambda|B|.$$

Again, ergodic theory for point processes, provides conditions under which the above convergence results hold true. For a supplementary material to this lesson see the working book project ((Baccelli et al., 2017, Chapeter 9)).

In the second part of the lesson we shall discuss some invariance of the Palm probability, which, under some conditions, allows one to move the observer from the typical node at the origin to some other point, such that the new point is also the typical one. This is partially related to the ergodicity of point processes (with a more complete relation holding however

only in 1D). More importantly this invariance is an intrinsic property of all Palm distributions (regardless of the dimension) and related to the unimodularity property.

12.1 Ergodicity of point processes

12.1.1 Continuous ergodicity

The following notions and results are borrowed from the general ergodic theory. They perfectly fit to our stationary framework for point processes $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$.

An invent $A \in \mathcal{A}$ is called $(\{\theta_t\}, \mathbf{P})$-*invariant* (or just *invariant* for short) if for all $t \in \mathbb{R}^d$, $\mathbf{P}(A \triangle \cap\theta_t A) = 0$ where $\triangle$ denotes the symmetric difference: $A \triangle B = (A \cup B) \setminus (A \cap B)$.

We next consider the family of invariant events:

$$\mathcal{I} := \{A \in \mathcal{A} : A \text{ is invariant}\}.$$

It is easy to see that $\mathcal{I}$ is a σ-algebra (prove it). We call it *invariant σ-algebra*.

We say that the stationary framework is *metrically transitive* if its invariant σ-algebra $\mathcal{I}$ is $\mathbf{P}$-*trivial*, i.e., if $\forall A \in \mathcal{I}, \mathbf{P}(A) \in \{0, 1\}$.

Remark 12.1.1. Note, when the event $A = \{\omega : \Phi(\omega) \in \Gamma\}$ corresponds to some property Γ of a stationary point process Φ in $\mathbb{R}^d$ (say defined on the canonical stationary framework) then A being invariant means that for all $t \in \mathbb{R}^d$, $\mathbf{1}(S_t\Phi \in \Gamma) = \mathbf{1}(\Phi \in A)$ $\mathbf{P}$-almost surely; i.e., the property Γ of Φ is almost surely invariant with respect to any translation of Φ. Examples of such properties Γ are: "the average number of points of Φ in the window increasing to $\mathbb{R}^d$ is equal to some constant λ", or "some translation-invariant graph on the points of Φ has an infinite component". Consequently, if the invariant σ-field $\mathcal{I}$ generated by Φ is $\mathbf{P}$-trivial then these properties Γ hold with probability 0 or 1. We shall provide later some necessary and sufficient conditions for $\mathcal{I}$ of a point process to be $\mathbf{P}$-trivial.

A sequence of sets $(B_n)_{n\geq 1}$ in $\mathbb{R}^d$ is said to be a *convex averaging sequence* if each B_n is bounded Borel and convex set; $B_n \subset B_{n+1}$, $\forall n$; and
$\sup\{r \geq 0 : B_n \text{ contains a ball of radius } r\} \to \infty$ when $n \to \infty$.

We recall now a famous result from ergodic theory using the terminology of our stationary framework. See e.g. ((Kallenberg, 2002, Theorem 10.6)) for its proof.

Theorem 12.1.2 (Birkhoff's Individual (or Pointwise) Ergodic Theorem)**.** *Let $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ be a stationary framework, $\mathcal{I}$ the invariant σ-algebra, $(B_n)_{n\geq 1}$ a convex averaging sequence in $\mathbb{R}^d$. For a measurable function on $(\Omega, \mathcal{A})$, such that $\mathbf{E}[|f|] < \infty$*

$$\lim_{n\to\infty} \frac{1}{|B_n|} \int_{B_n} f\circ\theta_x \, \mathrm{d}x = \mathbf{E}[f \,|\, \mathcal{I}], \quad \mathbf{P}\text{-a.s.} \tag{12.1.1}$$

where $\mathbf{E}[f \,|\, \mathcal{I}]$ is the conditional expectation with respect to $\mathcal{I}$.

Corollary 12.1.3. *Under the assumptions of Theorem 12.1.2, if the stationary framework is metrically transitive then (12.1.1) holds with* $\mathbf{E}[f\,|\,\mathcal{I}] = \mathbf{E}[f]$.

Verifying metrical transitivity is still not simple. In what follows we provide an equivalent and a sufficient condition.

We say stationary framework is *ergodic* if

$$\lim_{a\to\infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathbf{P}(A_1 \cap \theta_x A_2)\mathrm{d}x = \mathbf{P}(A_1)\mathbf{P}(A_2), \quad \forall A_1, A_2 \in \mathcal{A} \tag{12.1.2}$$

We say it is *mixing* if

$$\lim_{|x|\to\infty} \mathbf{P}(A_1 \cap \theta_x A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2), \quad \forall A_1, A_2 \in \mathcal{A} \tag{12.1.3}$$

Proposition 12.1.4. *For a stationary framework* $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ *the following relations hold true:*

$$\textit{mixing} \quad \Rightarrow \quad \textit{ergodicity} \quad \Leftrightarrow \quad \textit{metrical transitivity.} \tag{12.1.4}$$

Proof. We leave the first implication for the Reader. We show now that ergodicity implies metrical transitivity. Assume that the framework is ergodic. Consider some $A \in \mathcal{I}$. For any $t \in \mathbb{R}^d$, $\mathbf{P}(A \triangle \theta_t A) = 0$; and since $A \cap \theta_t A = A \setminus (A \setminus \theta_t A)$ and $A \setminus \theta_t A \subset A \triangle \theta_t A$, then $\mathbf{P}(A \cap \theta_t A) = \mathbf{P}(A) - \mathbf{P}(A \setminus \theta_t A) = \mathbf{P}(A)$. On the other hand, we deduce from ergodicity that

$$\lim_{a\to\infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathbf{P}(A \cap \theta_x A)\mathrm{d}x = \mathbf{P}(A)^2$$

Then $\mathbf{P}(A) = \mathbf{P}(A)^2$, thus $\mathbf{P}(A) \in \{0, 1\}$. Therefore the invariant σ-algebra $\mathcal{I}$ is $\mathbf{P}$-tirival, and the framework is consequently metrically transitive.

To prove the opposite implication, assume that the framework is metrically transitive. Let $A_1, A_2 \in \mathcal{A}$. By Birkhoff's theorem 12.1.2 and Corollary 12.1.3, we have

$$\lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} 1\{\theta_x(\omega) \in A_2\}\,\mathrm{d}x = \mathbf{E}[\mathbf{1}\{\omega \in A_2\}] = \mathbf{P}(A_2)$$

Then

$$\mathbf{P}(A_1)\mathbf{P}(A_2) = \mathbf{E}[\mathbf{1}\{\omega \in A_1\}] \left(\lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{1}\{\theta_x\omega \in A_2\}\,\mathrm{d}x \right)$$

$$\boxed{\text{since (lim ...) is } \mathbf{P}\text{-a.s. constant}} = \mathbf{E}\left[\mathbf{1}\{\omega \in A_1\} \lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{1}\{\theta_x\omega \in A_2\}\,\mathrm{d}x \right]$$

$$\boxed{\text{Dominated Convergence Theorem}} = \lim_{n\to\infty} \mathbf{E}\left[\mathbf{1}\{\omega \in A_1\} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{1}\{\theta_x\omega \in A_2\}\,\mathrm{d}x \right]$$

$$\boxed{\text{Fubini's Theorem}} = \lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{E}\left[\mathbf{1}\{\omega \in A_1\}\mathbf{1}\{\theta_x\omega \in A_2\}\right] \mathrm{d}x$$

$$= \lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{P}\left(A_1 \cap \theta_x A_2\right) \mathrm{d}x\,,$$

which completes the proof. □

Sometimes one says that a stationary point process Φ (meaning its distribution) is ergodic or mixing. By this we mean that the canonical space $(\mathbb{M}, \mathcal{M}, \{S_x\}, \mathbf{P}_\Phi)$ with the distribution of Φ as the probability measure is ergodic or mixing, respectively. The following result simplifies verification of these condition.

Proposition 12.1.5. *Let Φ be a stationary point process with Laplace transform $\mathcal{L} = \mathcal{L}_\Phi$. Then*

(i) Φ is ergodic if and only if

$$\lim_{a\to\infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathcal{L}_\Phi(f_1 + S_x f_2)\mathrm{d}x = \mathcal{L}_\Phi(f_1)\mathcal{L}_\Phi(f_2)$$

for any measurable $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}_+$ bounded with bounded support.

(ii) Φ is mixing if and only if

$$\lim_{|x|\to\infty} \mathcal{L}_\Phi(f_1 + S_x f_2) = \mathcal{L}_\Phi(f_1)\mathcal{L}_\Phi(f_2)$$

for any measurable $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}_+$ bounded with bounded support.

Proof. Cf. ((Daley and Vere-Jones, 2007, Proposition 12.3.VI)). □

Remark 12.1.6. Using Proposition 12.1.5 it is easy to see that a homogeneous Poisson point process Φ on $\mathbb{R}^d$ is mixing and hence ergodic. Consequently, (12.1.1) holds with $f(\omega) = f(\Phi(\omega))$ and $\mathbf{E}\left[f \,|\, \mathcal{I}\right] = \mathbf{E}\left[f\right]$. Also, we have 0-1 laws for its invariant events mentioned in Remark 12.1.1.

12.1.2 Discrete ergodicity

Here is our discrete ergodic result.

Theorem 12.1.7. *Let $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ be a stationary and ergodic framework, Φ a point process on $\mathbb{R}^d$ compatible with the flow $\{\theta_t\}_{t\in\mathbb{R}^d}$ with finite and non-null intensity λ and Palm probability $\mathbf{P}^0$. Let $(B_n)_{n\geq 1}$ be a convex averaging sequence in $\mathbb{R}^d$. For a measurable function f on $(\Omega, \mathcal{A})$, such that $\mathbf{E}^0\left[|f|\right] < \infty$*

$$\lim_{n\to\infty} \frac{1}{|B_n|} \int_{B_n} f\circ\theta_x \,\Phi\left(\mathrm{d}x\right) = \lambda \mathbf{E}^0\left[f\right], \quad \mathbf{P}\text{-a.s.} \tag{12.1.5}$$

Corollary 12.1.8. *Under assumptions of Theorem 12.1.7 we have*

$$\lim_{n\to\infty} \frac{1}{\Phi\left(B_n\right)} \int_{B_n} f\circ\theta_x \,\Phi\left(\mathrm{d}x\right) = \mathbf{E}^0\left[f\right] \quad \mathbf{P}\text{-a.s.}$$

Remark 12.1.9. For the proof of Theorem 12.1.7, one might want use a discrete version of Birkhoff's individual ergodic result. This is possible only in one dimension, i.e.; for stationary, ergodic framework with one dimensional flow $\{\theta_t\}_{t\in\mathbb{R}}$. Such framework supports point processes on the line. The reason is that Palm probability $\mathbf{P}^0$ is invariant with respect to natural discrete point shifts only in dimension $d = 1$. We shall explain this in Section 12.2, cf Corollary 12.2.3.

Proof of Theorem 12.1.7. The idea of the proof consists in approximating the discrete sum by the integral of some stochastic process and using the continuous ergodic result. In this regard for $\epsilon > 0$, let $g_\epsilon : \mathbb{R}^d \to \mathbb{R}_+$ be a measurable mapping with the following properties: (i) $g_\epsilon(x)$ is non-negative and continuous; (ii) $g_\epsilon(x) \equiv 0$ for $x \notin B_0(\epsilon)$ ($B_x(r)$ is the ball centered at x of radius r) and (iii) $\int_{\mathbb{R}^d} g_\epsilon(x)\,\mathrm{d}u = 1$. Let

$$h = h(\omega) = \int_{\mathbb{R}^d} g_\epsilon(x)\,\Phi(\mathrm{d}x).$$

By Campbell's averaging formula

$$\mathbf{E}\,[h] = \int_{\mathbb{R}^d} g_\epsilon(x)\lambda\,\mathrm{d}x = \lambda.$$

In addition,

$$h\circ\theta_t = \int_{\mathbb{R}^d} g_\epsilon(x)\,\Phi\circ\theta_t\,(\mathrm{d}x) = \int_{\mathbb{R}^d} g_\epsilon(xt)\,\Phi(\mathrm{d}x).$$

Let

$$B_n^\epsilon = \bigcup_{x\in B_n} B_x(\epsilon)$$

$$B_n^{-\epsilon} = \{x \in B_n : B_x(\epsilon) \subset B_n\}.$$

From our assumptions on g_ϵ we have $y \in B_n$ implies $\int_{B_n^\epsilon} g_\epsilon(y-t)\,\mathrm{d}t = 1$. Moreover, $y \notin B_n$ implies $\int_{B_n^{-\epsilon}} g_\epsilon(y-t)\,\mathrm{d}t = 0$. Hence

$$\int_{B_n^{-\epsilon}} g_\epsilon(y-t)\,\mathrm{d}t \le \mathbf{1}(y \in B_n) \le \int_{B_n^\epsilon} g_\epsilon(y-t)\,\mathrm{d}t. \tag{12.1.6}$$

Integrating over $\mathbb{R}^d$ w.r.t. $\Phi(\mathrm{d}y)$ we get

$$\int_{\mathbb{R}^d}\int_{B_n^{-\epsilon}} g_\epsilon(y-t)\,\mathrm{d}t\Phi(\mathrm{d}y) \le \Phi(B_n) \le \int_{\mathbb{R}^d}\int_{B_n^\epsilon} g_\epsilon(y-t)\,\mathrm{d}t\Phi(\mathrm{d}y).$$

But, for all $A \in \mathcal{B}\left(\mathbb{R}^d\right)$

$$\int_{\mathbb{R}^d}\left(\int_A g_\epsilon(y-t)\,\mathrm{d}t\right)\Phi(\,\mathrm{d}y) = \int_A\left(\int_{\mathbb{R}^d} g_\epsilon(y-t)\Phi(\,\mathrm{d}y)\right)\,\mathrm{d}t = \int_A h\circ\theta_t\,\mathrm{d}t.$$

Then,

$$\int_{B_n^{-\epsilon}} h\circ\theta_t \,\mathrm{d}t \le \Phi(B_n) \le \int_{B_n^{\epsilon}} h\circ\theta_t \,\mathrm{d}t.$$

Hence

$$\limsup_n \frac{\Phi(B_n)}{|B_n|} \le \limsup_n \frac{|B_n^{\epsilon}|}{|B_n|} \times \left(\frac{1}{|B_n^{\epsilon}|}\int_{B_n^{\epsilon}} h\circ\theta_t \,\mathrm{d}t\right).$$

The second term in the right-hand side of the above inequality tends $\mathbf{P}$-a.s. to λ from Birkhoff's theorem 12.1.2. The first term tends to 1 because B_n are convex with $\sup\{r \ge 0 : B_n$ contains a ball o ∞ when $n \to \infty$.

By the same arguments

$$\liminf_n \frac{\Phi(B_n)}{|B_n|} \ge \liminf_n \frac{|B_n^{-\epsilon}|}{|B_n|} \times \frac{1}{\left|B_n^{-\epsilon}\right|}\int_{B_n^{-\epsilon}} h\circ\theta_t \,\mathrm{d}t = \lambda, \quad \mathbf{P}\text{-a.s.}$$

Hence

$$\lim_{n\to\infty} \frac{\Phi(B_n)}{|B_n|} = \lambda \quad \mathbf{P} \text{ a.s.}$$

which proves the result with $f \equiv 1$. Consider now a general f. Let

$$h(\omega) = \int_{\mathbb{R}^d} g_\epsilon(x) f\circ\theta_x \Phi(\,\mathrm{d}x).$$

By the Campbell-Little-Mecke-Matthes' Theorem

$$\mathbf{E}\left[h\right] = \lambda \int_{\mathbb{R}^d} \mathbf{E}^0\left[g_\epsilon(x) f\right] \,\mathrm{d}x = \lambda \mathbf{E}^0\left[f\right].$$

Observe that,

$$h\circ\theta_t = \int_{\mathbb{R}^d} g_\epsilon(x) f\circ\theta_x\circ\theta_t \,\Phi\circ\theta_t\,(\mathrm{d}x) = \int_{\mathbb{R}^d} g_\epsilon(x-t) f\circ\theta_x \Phi(\mathrm{d}x).$$

Multiplying (12.1.6) by $f\circ\theta_y$ and then integrating over $\mathbb{R}^d$ w.r.t. $\Phi(\mathrm{d}y)$ we get

$$\int_{\mathbb{R}^d}\left(\int_{B_n^{-\epsilon}} g_\epsilon(y-t)\mathrm{d}t\right) f\circ\theta_y\Phi(\mathrm{d}y) \le \int_{B_n} f\circ\theta_y\Phi(\mathrm{d}y) \le \int_{\mathbb{R}^d}\left(\int_{B_n^{\epsilon}} g_\epsilon(y-t)\,\mathrm{d}t\right) f\circ\theta_y\Phi(\mathrm{d}y).$$

But, for all $A \in \mathcal{B}\left(\mathbb{R}^d\right)$

$$\int_{\mathbb{R}^d}\left(\int_A g_\epsilon(y-t)\,\mathrm{d}t\right) f\circ\theta_y\Phi(\mathrm{d}y) = \int_A\left(\int_{\mathbb{R}^d} g_\epsilon(y-t) f\circ\theta_y\Phi(\mathrm{d}y)\right)\,\mathrm{d}t = \int_A h\circ\theta_t\,\mathrm{d}t.$$

Then,

$$\int_{B_n^{-\epsilon}} h\circ\theta_t \,\mathrm{d}t \le \int_{B_n} f\circ\theta_y\Phi(\mathrm{d}y) \le \int_{B_n^{\epsilon}} h\circ\theta_t \,\mathrm{d}t.$$

Hence

$$\limsup_n \frac{\int_{B_n} f\circ\theta_y\Phi(\mathrm{d}y)}{|B_n|} \le \limsup_n \frac{|B_n^{\epsilon}|}{|B_n|} \times \frac{1}{|B_n^{\epsilon}|}\int_{B_n^{\epsilon}} h\circ\theta_t \,\mathrm{d}t.$$

The second term in the right-hand side of the above inequality tends $\mathbf{P}$-a.s. to $\mathbf{E}[h] = \lambda \mathbf{E}^0[f]$ from Birkhoff's theorem 12.1.2. The first term tends to 1 as seen in the case $f \equiv 1$. By the same arguments,

$$\liminf_n \frac{\int_{B_n} f\circ\theta_y \Phi(\mathrm{d}y)}{|B_n|} \geq \liminf_n \frac{|B_n^{-\epsilon}|}{|B_n|} \times \frac{1}{|B_n^{-\epsilon}|} \int_{B_n^{-\epsilon}} h\circ\theta_t \,\mathrm{d}t = \lambda \mathbf{E}^0[f], \quad \mathbf{P}\text{-a.s.}$$

Hence

$$\lim_{n\to\infty} \frac{\int_{B_n} f\circ\theta_y \Phi(\mathrm{d}y)}{|B_n|} = \lambda \mathbf{E}^0[f], \quad \mathbf{P}\text{-a.s.}$$

□

12.2 Point-shift invariance of Palm probability

Recall, the stationary probability $\mathbf{P}$ on the stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ is invariant with respect to all mappings θ_t, $t \in \mathbb{R}^d$: $\mathbf{P}\theta_t^{-1} = \mathbf{P}$. This is required in the formulation of Birkhoff's ergodic theorem. Discrete ergodic theorem for point processes triggers a question whether Palm probability of the point process Φ has similar property at least with respect to some particular flow-shifts; i.e., whether

$$\mathbf{P}^0\theta_\pi^{-1} = \mathbf{P}^0 \tag{12.2.1}$$

holds true for some $\pi \in \mathbb{R}^d$. Recall that $0 \in \Phi$ $\mathbf{P}^0$-a.s. and hence a necessary condition for (12.2.1) is that $\pi \in \Phi$ $\mathbf{P}^0$-a.s. Indeed: $0 \in \Phi\circ\theta_\pi$ iff $\pi \in \Phi$. Hence $\pi = \pi(\omega)$ needs to be a *random* point, and θ_π will be a *random-point shift*.

For a simple point process Φ, denote $\Omega^0 = \Omega^0_\Phi = \{\omega \in \Omega : 0 \in \Phi(\omega)\}$ and assume π is a measurable mapping from $(\Omega^0, \mathcal{A})$ into $\mathbb{R}^d$, having property $\pi(\omega) \in \Phi(\omega)$ for all $\omega \in \Omega^0$. We call such π *point map* (related to Φ). It is customary to interpret π as the point of the point process to which an observer is moving from the origin 0. Any such point map π generates a unique function $\pi_x = \pi_x(\omega)$, compatible with the flow, describing the point of the point process to which an observer should move from $x \in \Phi$:

$$\Phi(\omega) \ni x \longrightarrow \pi_x(\omega) := \pi\circ\theta_x(\omega) + x \in \Phi(\omega) \qquad \omega \in \Omega^0. \tag{12.2.2}$$

Indeed, it is easy to see that $\pi_x(\omega) \in \Phi(\omega)$. Observe also $\pi = \pi_0$.

We shall say that the point map π is *bijectif* if for all $\omega \in \Omega^0$ the mapping (12.2.2) is 1-1 mapping of Φ on itself.

Fact 12.2.1. *If π is bijectif then the inverse π_y^{-1} of π_x satisfies*

$$\pi_y^{-1}(\omega) = \pi^{-1}\circ\theta_y(\omega) + y \qquad \omega \in \Omega^0, \tag{12.2.3}$$

where $\pi^{-1} = \pi_0^{-1}$.

We formulate now and prove the following result.

Theorem 12.2.2. *Let $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ be a stationary framework, Φ a point process on $\mathbb{R}^d$ compatible with the flow $\{\theta_t\}_{t\in\mathbb{R}^d}$ with finite and non-null intensity λ and Palm probability $\mathbf{P}^0$. Let π be bijectif point shift related to Φ. Then*

$$\mathbf{P}^0\theta_\pi^{-1} = \mathbf{P}^0\,. \tag{12.2.4}$$

Proof. The result follows easily from the unimoudlarity of the Palm probability $\mathbf{P}^0$ (see Lesson on Stationary point processes). Indeed, for any measurable function f defined on Ω^0 and arbitrarily extended to Ω, let

$$m(x, y, \omega) := f\circ\theta_x(\omega)\mathbf{1}(x \in \Phi(\omega))\mathbf{1}(x = \pi_y(\omega))\,. \tag{12.2.5}$$

One can show that $m(x, y, \omega)$ is compatible with the flow; cf Exercise 12. By the unimodularity property (mass transport formula for the point process under its Palm probability, cf eq. (2.7) in the lesson on Stationary Point Processes)

$$\mathbf{E}^0\left[f\circ\theta_\pi\right] = \mathbf{E}^0\left[f\right]\,. \tag{12.2.6}$$

□

Corollary 12.2.3. *For a stationary, simple point process Φ on the real line $\mathbb{R}$, with the standard numbering of points ($x_n < x_{n+1}$, $x_0 = \max\{x_n \le 0\}$), for any fixed $k \in \mathbb{Z}$, the point map $\pi := x_k$ is bijectif. Consequently $\mathbf{P}^0 = \mathbf{P}^0\circ\theta_{x_k}^{-1}$.*

Example 12.2.4. For a simple stationary point process Φ, assume that $\mathbf{P}^0$ almost surely there is a unique point $x^* = x^*(\Phi)$ of Φ, different than 0, being the closest neighbour of 0. Note $\pi = x^*$ is *not* bijectif map. However the following involution

$$\pi := \begin{cases} x^* & \text{if } x^*\circ\theta_{x^*} + x^* = 0 \quad \text{i.e., if 0 and } x^* \text{ are mutually nearest neighbours} \\ 0 & \text{otherwise} \end{cases}$$

is a bijectif map.

12.3 Direct and inverse construction of Palm theory

We conclude this (last) lesson on stationary point processes by discussing a problem of the inverse construction in Palm theory.

Assume $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d})$ is measurable space with a d-dimensional flow, but without yet any given probability measure. Let Φ be a point process with points in $\mathbb{R}^d$, defined on this probability space, compatible with the flow.

12.3.1 Direct construction: from stationary to Palm probability

Assume a probability $\mathbf{P}$ is given on $(\Omega, \mathcal{A}, \{\theta_t\})$, invariant with respect to the flow. We have called it *stationary probability* in *Lesson on Stationary Framework for Point Processes*. We have shown that if the intensity λ of Φ, defined as $\lambda := \mathbf{E}\left[\Phi((0,1]^d)\right]$ (using the expectation with respect to $\mathbf{P}$), satisfies $0 < \lambda < \infty$, then there is a unique probability $\mathbf{P}^0$ on $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d})$ called *Palm probability of* Φ defined as

$$\mathbf{P}^0(A) = \frac{1}{\lambda|B|}\mathbf{E}\left[\int_B \mathbf{1}(\theta_x \in A)\,\Phi(\mathrm{d}x)\right] \qquad A \in \mathcal{A}\,; \tag{12.3.1}$$

with any set $B \in \mathcal{B}$ of finite, non-null Lebesgue measure $|B|$ on $\mathbb{R}^d$ (Palm-Matthes definition). Palm probability is obviously *not* compatible with the flow (the most evident argument is the fixed atom at the origin). It satisfy the following properties:

1. $\mathbf{P}^0\{\Omega^0\} = 1$, where $\Omega^0 = \Omega^0(\Phi) := \{\omega \in \Omega : 0 \in \Phi\}$.
2. $0 < \mathbf{E}^0\left[|V|\right] < \infty$, where $V = V(\Phi) := \{y \in \mathbb{R}^d : |y| \le \min_{z\in\Phi}|y - z|\}$. Indeed, recall $\mathbf{E}^0\left[|V|\right] = 1/\lambda$, (cf the Lesson on Stationary Voronoi Tessellation).
3. $\mathbf{P}^0$ is *unimodular* with respect to Φ in the sense that it satisfies the following Mass Transport Principle regarding Φ

$$\mathbf{E}^0\left[\int_{\mathbb{R}^d} g(y,\omega)\,\Phi(\mathrm{d}y)\right] = \mathbf{E}^0\left[\int_{\mathbb{R}^d} g(-x,\theta_x)\,\Phi(\mathrm{d}x)\right] \tag{12.3.2}$$

   for any (say non-negative) measurable functions g on $\mathbb{R}^d \times \Omega$ (not necessarily compatible with the flow).

4. $\mathbf{P}^0$ is *point-shift invariant* with respect to Φ, i.e. it is invariant with respect to all bijectif point shifts on Φ (Theorem 12.2.2).

12.3.2 Inverse construction: from Palm to stationary probability

Assume a probability $\mathbf{P}^0$ is given on $(\Omega, \mathcal{A})$, not necessarily (in fact necessarily not) compatible with the flow.

Fact 12.3.1. *Assume* $\mathbf{P}^0$ *satisfies conditions 1,2, and one of the two conditions: 3 (unimodularity) or 4 (point-shift invariance) above. Then, there exists a stationary (compatible with the flow) probability* $\mathbf{P}$ *on* $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d})$*, such that* $\mathbf{P}^0$ *is the Palm probability of* Φ *in the stationary framework* $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$*; i.e., (12.3.1) holds true. Moreover* $\mathbf{P}$ *is unique on* $\{\omega \in \Omega : \Phi(\mathbb{R}^d) \neq 0\}$ *and can be expressed by the inverse formula of Palm calculus*

$$\mathbf{P}(A) = \frac{1}{\mathbf{E}^0\left[|V|\right]}\mathbf{E}^0\left[\int_V \mathbf{1}(\theta_x \in A)\,\mathrm{d}x\right], \tag{12.3.3}$$

(cf Theorem 2.1 in the lesson on Stationary Voronoi tessellation).

Mecke Mecke ((1967)) proved the result with the unimodularity assumption. Heveling and Last Heveling et al. ((2005)) have shown that the point-shift invariance implies unimodularity.

Finally, recall that the completely analog mass transport formula was also used to characterize unimodular graphs (cf Lesson on Unimodular Graphs). In both scenario the aim was to formalize the notion of the typical point (node) of the point process (random graph).

12.4 Exercises

1. Prove that the family of invariant events $\mathcal{I}$ is a σ-algebra.

2. Prove that mixing implies ergodicity; cf (12.1.4).

3. Prove that a homogeneous Poisson point process on $\mathbb{R}^d$ is mixing.

4. Considering appropriate stationary framework prove that an independently marked homogeneous Poisson point process is mixing.

5. Consider the following stationary square-lattice point process $\Phi_{lattice}$ in $\mathbb{R}^d$: $\Phi_{lattice} := S_U\mathbb{Z}^d = \{z - U : z \in \mathbb{Z}^d\}$, where $\mathbb{Z}$ are integers and U is a uniform random variable on $[0,1]^d$. Show that $\Phi_{lattice}$ is ergodic but not mixing.

6. Prove that a mixture $\Phi_{mixture} = I \times \Phi_1 + (1 - I) \times \Phi_2$, of two homogeneous Poisson point processes Φ_1 and Φ_2 of different intensities (I, Φ_1, Φ_2 are independent, $\mathbf{P}\{\, I = 1 \,\} = 1 - \mathbf{P}\{\, I = 0 \,\} = p$, $0 < p < 1$) is *not* ergodic. Find invariant event which is not $\mathbf{P}$-trivial.

7. Prove Corollary 12.1.8.

8. For ergodic point process Φ of non-null intensity, show that $\mathbf{P}\left(\Phi\left(\mathbb{R}^d\right) = \infty\right) = 1$.

9. Let $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ be a stationary framework, Φ a point process compatible with the flow $\{\theta_t\}_{t\in\mathbb{R}^d}$ with finite and non-null intensity λ, such that $\mathbf{P}\left(\Phi\left(\mathbb{R}^d\right) = 0\right) = 0$. Let $A \in \mathcal{A}$ be *strictly* $\{\theta_t\}$*-invariant*; i.e. for all $t \in \mathbb{R}^d$, $A = \theta_{-t}A$. Show that $\mathbf{P}(A) = 1$ if and only if $\mathbf{P}^0(A) = 1$. *Hint:* Using the Inverse Formula of Palm calculus show that $\mathbf{P}(A) = 1 - \mathbf{E}^0\left[1_{A^c}\,|V|\right]$.

10. *Cross-ergodic theorems.* Under assumptions of Theorem 12.1.7, with $B_n = B_0^o(n)$ being open balls centered at 0 of radius n, show that (12.1.7) and (12.1.1) with $\mathbf{E}[f\,|\mathcal{I}] = \mathbf{E}[f]$ hold $\mathbf{P}^0$ almost surely. *Hint:* Show that the event

$$A = \left\{\omega \in \Omega : \frac{1}{|B_n|}\int_{B_n} f\circ\theta_x(\omega)\,\mathrm{d}x \underset{n\to\infty}{\longrightarrow} \mathbf{E}[f]\right\}.$$

    is strictly $\{\theta_t\}$-invariant. Use Exercise 9.

11. Let π be a point map. Show that $\pi_{a+b} = \pi_a \circ \theta_b + b$ for all $a, b \in \mathbb{R}^d$. Using this prove Fact 12.2.1.

12. Show that the mass function m given in (12.2.5) satisfies $m(x - t, y - t, \theta_t) = m(x, y, \omega)$ for all $\omega \in \Omega$, $x, y, t \in \mathbb{R}^d$. Prove (12.2.6).

13. Verify the statements of Example 12.2.4.

14. Argue that for a homogeneous Poisson point process on the line $\mathbf{P}^0 \circ \theta_{x^*}^{-1} \neq \mathbf{P}^0$, where x^* is the nearest neighbour of 0. Recall Slivnyak-Mecke's theorem to characterize $\mathbf{P}^0$. Observe that under $\mathbf{P}^0$, $\Phi \circ \theta_{x^*}$ has the following property: almost surely, in the positive or negative half-line, the first two consecutive points of the point process are separated by increasing distances: $x_1 \circ \theta_{x^*} < x_2 \circ \theta_{x^*} - x_1 \circ \theta_{x^*}$ or $-x_{-1} \circ \theta_{x^*} < x_{-1} \circ \theta_{x^*} - x_{-2} \circ \theta_{x^*}$ This contradicts with the properties of Φ under $\mathbf{P}^0$.

15. Using the discrete ergodic result (Theorem 12.1.7) argue that the estimator

$$\hat{K}_n(r) := \frac{|B_n| \sum_{X_i \neq X_j \in B_n} \mathbf{1}(|X_i - X_j| \leq r)}{(\Phi(B_n))^2}$$

of the Ripley's K function (see Exercise 19 to the Lesson on stationary point processes) calculated on the realizations of an ergodic point process Φ observed within an increasing convex averaging sequence B_n of observation windows is *consistent* i.e., it convergence to

$$K(r) := \frac{\mathbf{E}^0 \left[\Phi(B_0(r))\right] - 1}{\lambda}$$

when $n \to \infty$. (Almost sure convergence is called *strong consistency.*)

16. *Computer exercise.* Simulate a homogeneous Poisson point process in increasing square window B_n and verify numerically that $\hat{L}_n(r) := \sqrt{\hat{K}_n(r)/\pi} \to r$ when $n \to \infty$ (strong consistency of the estimator of $\hat{L}(r)$).

Lesson 13

Random closed sets

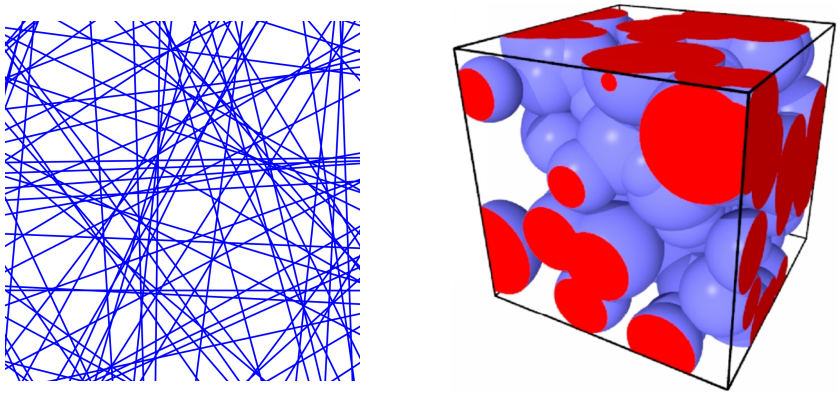


Figure 13.1: Two popular examples of random closed sets: Poisson line process on the plane (left) and a spherical Boolean model in $\mathbb{R}^3$ (right, image borrowed from Hermann and Elsner ((2014))).

We have already seen examples of random sets, mainly generated by point processes, such as bond percolation, random geometric (Gilbert) graph, the cells of a Voronoi tessellation. There is however a need, motivated by many applied sciences (e.g. material sciences and biology) for a framework allowing one to consider even more general random sets. The theory of random closed sets offers such possibility. As we shall see, it is a natural extension of the theory of point processes in the sense that the support of each point process is a random closed set. We shall briefly sketch foundations of the theory and present a few popular examples.

Some supplementary material to this lesson can be found in the book project ((Baccelli et al., 2017, Chapters 10–11)). For more reading on the foundations of the theory of random closed sets see Matheron ((1975)); Molchanov ((2005)); Schneider and Weil ((2008)) and Chiu et al. ((2013)) for various examples.

13.1 Random closed sets framework

As for general theory of point processes (see the lesson on Poisson point process), we consider a topological space $\mathbb{E}$ that is locally compact, second countable and Hausdorff, abbreviated to LCSCH space.

13.1.1 Space of closed sets

Denote by $\mathcal{F}$ the set of closed subsets of $\mathbb{E}$, that will be the space of realizations of our random closed sets, with some topology and σ-field considered on it. Note that the elements (points) of $\mathcal{F}$ are closed subsets of $\mathbb{E}$, and subsets of $\mathcal{F}$ are sets of subsets of $\mathbb{E}$ ([1]). We denote also by $\mathcal{K}$ (respectively $\mathcal{G}$) the set of compact (respectively open) subsets of $\mathbb{E}$.

We need the following notation to define the topology on $\mathcal{F}$. For any $A \subset \mathbb{E}$, we denote by $\mathcal{F}_A$ the subset of all closed set which intersect (hit) A

$$\mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \emptyset\}.$$

Similarly, we denote by $\mathcal{F}^A$ the subset of all closed sets which do not intersect (miss) A

$$\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \emptyset\}.$$

For all $A, A_1, \ldots, A_n \subset \mathbb{E}$, $n \geq 1$, let $\mathcal{F}^A_{A_1,\ldots,A_n}$ be the subset of all closed sets which hit all sets $A_1, \ldots, A_n$ but miss A

$$\mathcal{F}^A_{A_1,\ldots,A_n} = \mathcal{F}^A \cap \mathcal{F}_{A_1} \cap \ldots \cap \mathcal{F}_{A_n}.$$

Some simple relations regarding $\mathcal{F}^A_{A_1,\ldots,A_n}$ are given in Exercise 1.

Recall, defining a topology on $\mathcal{F}$ consists in specifying a family $\mathcal{T}$ of open subsets of $\mathcal{F}$. We consider the *Fell topology* $\mathcal{T}$ on $\mathcal{F}$; it is generated by the base

$$\mathcal{O} = \left\{\mathcal{F}^K_{G_1,\ldots,G_n} : K \in \mathcal{K}, G_1, \ldots, G_n \in \mathcal{G}, n \geq 1\right\}. \tag{13.1.1}$$

In other words, open subsets of $\mathcal{F}$, i.e., its topology $\mathcal{T}$, are all subsets of $\mathcal{F}$, which can be expressed as a union of elements of the given basis $\mathcal{O}$, which needs to satisfy in this regard some conditions, cf Exercise 2. Note $\mathcal{T}$ and $\mathcal{O}$ are two sets of subsets of subsets of $\mathbb{E}$ ([2]).

We will also consider the space of all *non-empty* closed sets $\mathcal{F}' = \mathcal{F} \setminus \{\emptyset\}$, with the corresponding subspace-topology induced by the Fell topology on $\mathcal{F}$ (the open sets on $\mathcal{F}'$ are open sets of $\mathcal{T}$ intersected with $\mathcal{F}'$). This restriction is important both from the theoretical and practical point of view, with the empty set not being observable.

Remark 13.1.1. The following topological properties of the two spaces $\mathcal{F}$ and $\mathcal{F}'$ are of interest for us; see for their proofs ((Baccelli et al., 2017, Chapter 10)) and the references

[1] They could be called *meta-subsets* of $\mathbb{E}$ although this is not a standard terminology.
[2] I.e., $\mathcal{O}, \mathcal{T}$ are meta-meta-subsets of $\mathbb{E}$ in the nonstandard terminology of Footnote 1.

cited therein. The most important message from these properties is that $\mathcal{F}$ *and* $\mathcal{F}'$ *with the Fell topology inherit the fundamental topological properties assumed for* $\mathbb{E}$. This allows us, in particular, to define and study the point process on these two spaces by a straightforward extension of the theory previously consider on $\mathbb{E}$.

1. The space $\mathcal{F}$ of all closed subsets of $\mathbb{E}$, with the Fell topology $\mathcal{T}$ is a compact, second countable and Hausdorff space, CSCH space for short. These are the same topological properties as the original LCSCH space $\mathbb{E}$, with, moreover, the whole space $\mathcal{F}$ being a compact set (while $\mathbb{E}$ is assumed to be only locally compact).

2. The space $\mathcal{F}' \subset \mathcal{F}$ of all nonempty closed subsets of $\mathbb{E}$, with the sub-space topology induced by the Fell topology $\mathcal{T}$, is a LCSCH space. These are the same topological properties as the original LCSCH space $\mathbb{E}$, with the whole space $\mathcal{F}'$ being compact, iff $\mathbb{E}$ is compact.

3. Each compact subset of $\mathcal{F}'$ can be included in a set $\mathcal{F}_K$, for some $K \in \mathcal{K}$.

4. The Borel σ-algebra $\mathcal{B}_{\mathcal{F}}$ on $\mathcal{F}$, i.e., the σ-algebra generated by all open sets $\mathcal{T}$, is also generated by any of the following families of subsets of $\mathcal{F}$

$$\left\{\mathcal{F}^K : K \in \mathcal{K}\right\},\ \left\{\mathcal{F}_K : K \in \mathcal{K}\right\}, \left\{\mathcal{F}_G : G \in \mathcal{G}\right\}, \text{ or } \left\{\mathcal{F}^G : G \in \mathcal{G}\right\}.$$

5. $\mathcal{F}'$ is a Borel subset of $\mathcal{F}$ (i.e., $\mathcal{F}' \in \mathcal{B}_{\mathcal{F}}$, as a consequence of $\mathcal{F}' = \mathcal{F}_{\mathbb{E}}$) and the Borel σ-algebra $\mathcal{B}_{\mathcal{F}'}$ on $\mathcal{F}'$ (induced by the subspace topology of $\mathcal{F}' \subset \mathcal{F}$) is equal to the restriction of $\mathcal{B}_{\mathcal{F}}$ to $\mathcal{F}'$.

13.1.2 Random closed set

A *random closed set* is a random mapping Z from some probability space $(\Omega, \mathcal{A}, \mathbf{P})$ to the measurable space of closed sets $(\mathcal{F}, \mathcal{B}_{\mathcal{F}})$. As usual, the probability distribution of Z is the probability measure $\mathbf{P}_Z$ on $(\mathcal{F}, \mathcal{B}_{\mathcal{F}})$ being the image of $\mathbf{P}$ by Z, i.e.; $\mathbf{P}_Z(B_{\mathcal{F}}) = \mathbf{P}\{\, Z \in B_{\mathcal{F}} \,\}$, for $B_{\mathcal{F}} \in \mathcal{B}_{\mathcal{F}}$. In the similar way we consider random non-empty closed set on $\mathcal{F}'$.

Natural examples of random closed sets such as random balls, triangles, orthants, level-sets are generated by random variables, vectors or stochastic processes in the space $\mathbb{E}$, see Exercise 3, 4, 5. We have already also seen several more complicated random sets in previous lessons, which can be now formally recognized as random closed sets.

Example 13.1.2 (Support of a point process)**.** Let Φ be a point process on $\mathbb{E}$. The support of Φ, $\mathrm{supp}(\Phi) = \{x \in \mathbb{E} : \Phi(\{x\}) > 0\}$ is a random closed set on $\mathcal{F}$; Indeed, $\mathrm{supp}(\Phi)$ is a closed set, since Φ is a Radon measure (atoms of Φ do not have accumulation points and hence their union, albeit countable is a closed set). Moreover, for $K \in \mathcal{K}$, $\{\mathrm{supp}(\Phi) \in \mathcal{F}_K\} = \{\Phi(K) > 0\}$ is a measurable event since Φ is a point process and by Remark 13.1.1 point 4 it is enough to conclude that $\mathrm{supp}(\Phi)$ is $\mathcal{B}_{\mathcal{F}}$ measurable.

Note that $\mathrm{supp}(\Phi)$ caries the information about the locations of atoms of Φ, but not their multiplicities.

Example 13.1.3 (Voronoi cells)**.** Let $\Phi = \sum_i \delta_{X_i}$ be a (say simple, stationary) point process on $\mathbb{R}^d$. The following random cells considered in the lesson on Stationary Voronoi tessellation are random closed set on the space $\mathcal{F}$ of closed subsets of $\mathbb{E} = \mathbb{R}^d$. It is clear that they are closed sets, but it is more tedious to prove their measurability.

- Centered cell of a given location $x \in \mathbb{R}^d$, $V(x) := \{y \in \mathbb{R}^d : |y| \leq \min_{X_i \in \Phi} |y - X_i|\}$, in particular $V = V(0)$, which is the typical cell under the Palm probability of Φ.
- Centered zero cell $V_* := V(X_*)$, where X_* is the closest pint of Φ to the origin. It is well defined only on a subset of the probability space where X_* is unique, recall this event has stationary probability equal to 1. Also, the non-centered zero-cell $C_* = V_* + X_*$.
- Voronoi cells $C_i := \{y \in \mathbb{R}^d : |y - X_i| \leq \min_{X_j \in \Phi} |y - X_j|\}$ and their centered variants $V_i = C_i - X_i$ of all points (this requires some measurable numbering of points).

13.1.3 The capacity functional of a random closed set

We introduce now a fundamental characteristic of the distribution of a random closed set. As we shall see, it has properties analogous to these of the probability distribution function of a random vector, in particular, it uniquely characterizes the distribution of the random closed set.

Definition 13.1.4. Given a random closed set Z, its *capacity functional* is defined as

$$T_Z(K) = \mathbf{P}(Z \cap K \neq \emptyset) = \mathbf{P}(Z \in \mathcal{F}_K)$$

for all $K \in \mathcal{K}$.

Remark 13.1.5. The capacity functional of a random closed set characterizes its probability distribution because the sets of the form $\mathcal{F}_K$, $K \in \mathcal{K}$, generate the σ-field $\mathcal{B}_{\mathcal{F}}$. The functional $T_Z(B)$ can extended to sets B for which $F_B \in \mathcal{B}_{\mathcal{F}}$, in particular open and closes sets $B \subset \mathbb{E}$.

Example 13.1.6. The support of a point process Φ is a random closed set (cf. Example 13.1.2) with capacity functional

$$\begin{aligned} T_{\mathrm{supp}(\Phi)}(K) &= \mathbf{P}(\mathrm{supp}(\Phi) \cap K \neq \emptyset) \\ &= \mathbf{P}(\Phi(K) \neq 0) \\ &= 1 - \mathbf{P}(\Phi(K) = 0) \\ &= 1 - \nu_\Phi(K), \end{aligned}$$

where $\nu_\Phi(\cdot)$ is the void probability of Φ. (Recall the Rényi's theorem from the lesson on Poisson point process, which says that the void probabilities characterize the distribution of a simple point process.)

Let us define the following *difference operators* S_n, $n = 0, 1, \ldots$ of a functional T on $\mathcal{K}$: for $K, K_1, K_2, \ldots \in \mathcal{K}$, let

$$S_0(K) = 1 - T(K)$$

and recursively, for $n \geq 1$

$$S_n(K; K_1, \ldots, K_n) = S_{n-1}(K; K_1, \ldots, K_{n-1}) - S_{n-1}(K \cup K_n; K_1, \ldots, K_{n-1}).$$

The following result states some important properties of the capacity functional of a random closed set, analogous to the classical properties of the distribution function of a real random vector, concerning its bounds, monotonicity and right-continuity.

Proposition 13.1.7. *The capacity functional $T = T_Z$ of a random closed set Z has the following properties:*

1. (bounds) $0 \leq T(K) \leq 1$, $T(\emptyset) = 0$.
2. (monotonicity) [3] $S_n(K; K_1, \ldots, K_n) \geq 0$ *for all* $K, K_1, K_2, \ldots K_n \in \mathcal{K}$, $n \geq 1$.
3. (upper semi-continuity) [4] *Let* $K, K_1, K_2, \ldots \in \mathcal{K}$. *If* $K_n \downarrow K$ *then* $T(K_n) \to T(K)$.

Proof. The bounds are obvious. The upper semi-continuity follows from the continuity of property of the probability measure on the decreasing sets $\mathcal{F}_{K_n} \downarrow \mathcal{F}_K$ when $K_n \downarrow K$. For the monotonicity observe that

$$\mathbf{P}\{ Z \in \mathcal{F}^K_{K_1,\ldots;K_n} \} = \mathbf{P}\Big\{ Z \in \mathcal{F}^K_{K_1,\ldots,K_{n-1}} \Big\} - \mathbf{P}\Big\{ Z \in \mathcal{F}^{K\cup K_n}_{K_1,\ldots,K_{n-1}} \Big\}$$

and consequently, by the induction with respect to n

$$S_n(K; K_1, \ldots, K_n) = \mathbf{P}\{ Z \in \mathcal{F}^K_{K_1,\ldots,K_n} \},$$

with the lase expression obviously non-negative. □

In full analogy to the classical existence result regarding the distribution function of a random vector, we have the following theorem characterization of the probability distribution on the space $(\mathcal{F}, \mathcal{B}_\mathcal{F})$.

Theorem 13.1.8 (Choquet's theorem)**.** *Let T be real valued function defined on $\mathcal{K}$. Then there exists a (necessarily unique) probability distribution $\mathbf{P}_T$ on $(\mathcal{F}, \mathcal{B}_\mathcal{F})$ such $\mathbf{P}_T(\mathcal{F}_K) = T$ if and only if T satisfies properties 1,2,3 of Proposition 13.1.7.*

Remark 13.1.9. The direct part of Theorem 13.1.8 follows from Proposition 13.1.7. The uniqueness from the first statement in Remark 13.1.5. The most difficult part consists in proving that if T satisfies properties 1,2,3 then there exists a random closed set Z on $\mathcal{F}$ with $T_Z = T$. The proof given by Matheron in ((Matheron, 1975, Theorem 2-2-1)) is based on the routine application of the measure-theoretic arguments related to extension of measures from algebras to σ-algebras. Molchanov gives also in ((Molchanov, 2005, Section 1.1.3)) a different proof, based on some arguments from harmonic analysis.

[3] In a more standard terminology this condition is called *complete alternation*, cf Molchanov ((2005)).

[4] This condition is in fact equivalent to the upper semi-continuity of T as a functional on $\mathcal{K}$ with the subspace topology induced by the Fell topology on $\mathcal{F}$, cf ((Matheron, 1975, Proposition 1.4.2)) or ((Molchanov, 2005, Proposition D.7)).

13.2 Poisson set process

The fact that the space of closed subsets inherits crucial topological properties of $\mathbb{E}$ (cf Remark 13.1.1) allow us to model random patters of (closest) subsets of $\mathbb{E}$ as point processes on $\mathcal{F}$. We call them *set processes*. However, since the whole space $\mathcal{F}$ is a compact set, these set processes would be configurations of only finite total number of sets. In order to allow for infinite set processes, we consider them on the space of non-empty closed sets $\mathcal{F}'$, which is only LCSCH space.

Denote by $\mathbb{M}_{\mathcal{F}'}$ and $\mathcal{M}_{\mathcal{F}'}$ the space of counting measures on $\mathcal{F}'$ and its σ-field, respectively, defined exactly is the same way as $\mathbb{M}$ and $\mathcal{M}$ was defined for counting measures on $\mathbb{E}$ ([5]). A set process Φ_f is a point process on $\mathcal{F}'$; i.e., a measurable mapping from some probability space to $(\mathbb{M}_{\mathcal{F}'}, \mathcal{M}_{\mathcal{F}'})$. From the general theory of point processes we know that such a process admits a representation $\Phi_{\mathrm{f}} = \sum_{n\in\mathbb{Z}} \delta_{F_n}$, where the sets $F_n \in \mathcal{F}'$ are the set-atoms of Φ_{f}.

In what follows we shall focus on Poisson processes in this setting. As usual, they are characterized by intensity measures Λ_f, which need to be locally finite (Radon) measures or, equivalently, satisfy $\Lambda_f(\mathcal{F}_K) < \infty$ for all $K \in \mathcal{K}$.

Definition 13.2.1 (Poisson set process)**.** A *Poisson set process* of Radon intensity measures Λ_{f} on $\mathcal{F}'$ is a Poisson point process on $\mathcal{F}'$ with intensity measure Λ_f.

Remark 13.2.2. Observe directly from the definition of Poisson process, that the number of sets-atoms of a Poisson set process Φ_f of intensity Λ_f hitting a given compact set $K \in \mathcal{K}$ is a Poisson random variable with parameter $\Lambda_f(\mathcal{F}_K)$.

We present now two popular examples of Poisson set processes.

Example 13.2.3 (Poisson line process on the plane)**.** Let $\Phi = \sum_i \delta_{(r_i,\theta_i)}$ be a Poisson point process on $\mathbb{R} \times (0, \pi]$ of intensity Λ. Consider the following mapping ℓ of $\mathbb{R} \times (0, \pi]$ into the space of closed subsets $\mathcal{F}$ of $\mathbb{R}^2$: for $r \in \mathbb{R}$, $\theta \in (0, \pi]$, let $\ell(r, \theta)$ be a unique non-oriented line on $\mathbb{R}^2$ such that the orthogonal projection of the origin on it has the polar coordinates (r, θ). (Equivalently, it is the line with the Cartesian coordinates (x, y) satisfying $x \cos\theta + y \sin\theta = r$.) It can be shown that ℓ is a measurable mapping and that for any relatively compact (topologically bounded) $B_{\mathcal{F}} \in \mathcal{F}$ the inverse set $\ell^{-1}(B_{\mathcal{F}})$ is a bounded subset of $\mathbb{R} \times (0, \pi]$. Consequently, by the Poisson displacement theorem, the point process $\Phi_\ell := \sum_i \delta_{\ell(r_i,\theta_i)}$ is a Poisson set process on $\mathbb{R}^2$. Note, all set-atoms of Φ_ℓ are lines and we call it *Poisson line process*. The intensity measure M_ℓ of Φ_ℓ is related to Λ by the following relation:

$$M_\ell(\mathcal{F}_K) = \int_{\mathbb{R}\times(0,\pi]} \mathbf{1}(\ell(r,\theta) \cap K \neq \emptyset)\, \Lambda(\mathrm{d}(r,\theta)\,.$$

It can be show that if $\Lambda(\mathrm{d}(r,\theta)) = \mathrm{d}r\, G(\mathrm{d}\theta)$ for some finite measure on $(0, \pi]$ then the distribution of Φ_ℓ is invariant with respect to any translation in $\mathbb{R}^2$. It is called then *stationary*

[5] $\mathbb{M}_{\mathcal{F}'}$ are meta counting measures in the nonstandard terminology of Footnote 1.

Poisson line process of intensity $\lambda = G((0,\pi])$. If moreover $G(\mathrm{d}\theta) = \lambda/\pi\,\mathrm{d}\theta$ then the distribution of Φ_ℓ is also rotation invariant and the process is called a *motion invariant Poisson line process* of intensity λ. It can be shown that the intensity of a motion invariant Poisson line process of intensity λ satisfies for *convex compact* set K

$$M_\ell(\mathcal{F}_K) = \frac{\lambda}{\pi} L(K)$$

where $L(K)$ is the perimeter of K.

Example 13.2.4 (Poisson germ-grain model)**.** Let $\tilde{\Phi}$ be a Poisson process on $\mathbb{R}^d$ with non-null intensity measure Λ, independently, identically marked by random, closed, nonempty subsets $Z_i \in \mathcal{F}'$ of $\mathbb{R}^d$

$$\tilde{\Phi} = \sum_i \delta_{(X_i,Z_i)}\,. \tag{13.2.1}$$

Points X_i are called *germs* and marks — random sets Z_i are called *grains*. We assume that

$$\mathbf{E}\left[\Lambda(\check{Z}_0 \oplus K)\right] < \infty \quad \text{for all } K \in \mathcal{K} \tag{13.2.2}$$

where Z_0 is the generic grain, with $A \oplus B = \{x+y : x \in A, y \in B\}$ and $\check{A} = \{-x : x \in A\}$. As an independently marked Poisson point process $\tilde{\Phi}$ is a Poisson process on $\mathbb{R}^d \times \mathcal{F}'$ with intensity measure $\Lambda(\mathrm{d}x)\mathbf{P}_{Z_0}(\mathrm{d}F)$, $x \in \mathbb{R}^d$, $F \in \mathcal{F}'$, where $\mathbf{P}_{Z_0}$ is the distribution of Z_0. It can be shown, again by the displacement theorem, that

$$\Phi_f = \sum_{(X_i,Z_i)\in\tilde{\Phi}} \delta_{X_i+Z_i} \tag{13.2.3}$$

is a Poisson set process on the space of closed subsets $\mathcal{F}$ of $\mathbb{R}^d$. We call it *Poisson germ-grain model*. It's intensity measure admits the following expression

$$\Lambda_f(\mathcal{F}_K) = \mathbf{E}\left[\Lambda(\check{Z}_0 \oplus K)\right]$$

for $K \in \mathcal{K}$, which is finite by the assumption (14.1.11). Indeed

$$\begin{aligned}
\Lambda_f(\mathcal{F}_K) &= \mathbf{E}\left[\int_{\mathbb{R}^d\times\mathcal{F}'} \mathbf{1}((x+F)\cap K \neq \emptyset)\,\tilde{\Phi}(\mathrm{d}(x,F))\right] \\
\boxed{\text{Campbell averaging formula}} &= \int_{\mathcal{F}'}\int_{\mathbb{R}^d} \mathbf{1}((x+F)\cap K \neq \emptyset)\,\Lambda(\mathrm{d}x)\mathbf{P}_{Z_0}(\mathrm{d}F) \\
\boxed{(x+F)\cap K \neq \emptyset \text{ equivalent to } x \in \check{F}\oplus K} &= \int_{\mathcal{F}'} \Lambda(\check{F}\oplus K)\,\mathbf{P}_{Z_0}(\mathrm{d}F) \\
&= \mathbf{E}\left[\Lambda(\check{Z}_0\oplus K)\right]\,.
\end{aligned}$$

13.2.1 Boolean model

Finally, we present a famous random closed set model, which is constructed here as a union of set-atoms of a Poisson set process.

Definition 13.2.5 (Boolean coverage model)**.** Let Λ_f be a Radon measure on $\mathcal{F}'$ with a support contained in $\mathcal{K}$, $\Lambda_f(\mathcal{F}' \setminus \mathcal{K}) = 0$. Let Φ_f be a Poisson set process of intensity measures Λ_f. We define the *Boolean coverage model* of intensity Λ_f as the union of the set-atoms of Φ_f

$$\Xi := \bigcup_{F \in \Phi_f} F. \tag{13.2.4}$$

Remark 13.2.6. Note that we can take the Poisson germ-grain model of Example 14.1.4 as the Poisson set process Φ_f in the Definition 13.2.5 of the Boolean coverage model. It is enough to assume that the grains are compact sets $Z_i \in \mathcal{K}$ to guarantee that Λ_f is supported by $\mathcal{K}$. In this case we call Ξ a *Boolean germ-grain model.*

Proposition 13.2.7. *The Boolean coverage model as in Definition 13.2.5 is a random closed set with the capacity functional*

$$T_\Xi(K) = 1 - e^{-\Lambda_f(\mathcal{F}_K)}$$

for $K \in \mathcal{K}$.

Proof. Note that for all $K \in \mathcal{K}$

$$\#\{F \in \Phi_\mathrm{f} : F \cap K \neq \emptyset\} = \#\{F \in \Phi_\mathrm{f} \cap \mathcal{F}_K\} = \Phi_\mathrm{f}(\mathcal{F}_K) < \infty, \tag{13.2.5}$$

with the last inequality following from the fact that $\mathcal{F}_K$ is compact and Φ_f is a point process on $\mathcal{F}'$. This implies that Ξ, which is the union (13.2.4) of (possibly countably many) closed sets is a closed set. Indeed, points of a convergent sequence on Ξ are eventually contained in a compact subset K of $\mathbb{E}$. By (14.1.3) they belong hence to some union of finitely many set-atoms F of Φ_f. As a finite union of closed sets, it is a closed set, thus containing the sequence limit.

Regarding $\mathcal{B}_{\mathcal{F}'}$ measurability of Ξ, note for all $K \in \mathcal{K}$

$$\left\{\Xi \in \mathcal{F}^K\right\} = \{\Xi \cap K = \emptyset\} = \{\Phi_f(\mathcal{F}_K) = 0\},$$

which is measurable set since Φ_f is a point process on $\mathcal{F}'$.

Let us calculate the capacity functional T_Ξ of Ξ

$$\begin{aligned} T_\Xi(K) &= \mathbf{P}\{\Xi \in \mathcal{F}_K\} \\ &= 1 - \mathbf{P}\left\{\Xi \in \mathcal{F}^K\right\} \\ &= 1 - \mathbf{P}\{\Phi_f(\mathcal{F}_K) = 0\} \\ &\boxed{\Phi_f \text{ is Poisson process of intensity } \Lambda_f} = 1 - e^{-\Lambda_f(\mathcal{F}_K)}. \end{aligned}$$

□

Together with the Voronoi tessellation, Boolean model is one of the most popular models in stochastic geometry. Two remaining lessons will be devoted to the study of its properties.

TODO

- Introduce Minkowski functionals.

13.3 Exercises

1. Prove the following relations from the definition of the respective sets:
   (a) $\mathcal{F}^A = \mathcal{F}_A^{\mathrm{c}}, \quad \forall A \subset \mathbb{E}$;
   (b) $\mathcal{F}^{\emptyset} = \mathcal{F}$;
   (c) $\mathcal{F}_{\emptyset} = \emptyset$; note $\emptyset$ on the left-hand-side is the empty subset of $\mathbb{E}$, while $\emptyset$ on the right-hand-side is the empty subset of $\mathcal{F}$.
   (d) $\mathcal{F}^{\mathbb{E}} = \{\emptyset\}$, this is a singleton (one element subset) of $\mathcal{F}$, consisting of only one closed set, being the empty set.
   (e) $\mathcal{F}_{\mathbb{E}} = \mathcal{F} \setminus \{\emptyset\}$;
   (f) $\mathcal{F}_A^{\emptyset} = \mathcal{F}_A, \quad \forall A \subset \mathbb{E}$;
   (g) $\emptyset \in \mathcal{F}^A, \quad \forall A \subset \mathbb{E}$;
   (h) $\emptyset \notin \mathcal{F}^A_{A_1,\ldots,A_n}$ for all $A, A_1, \ldots, A_n \subset \mathbb{E}$, $n \geq 1$;
   (i) $\mathcal{F}^A_{A_1,\ldots,A_n} = \emptyset, \quad$ if $A_1, \ldots, A_n \subset A$;
   (j) $\mathcal{F}^A \setminus \{\emptyset\} \subset \mathcal{F}_{A^{\mathrm{c}}}$.

2. Verify that the collection $\mathcal{O}$ of subsets of $\mathcal{F}$ given in (13.1.1) satisfies the following conditions:
   (a) $\mathcal{F} \in \mathcal{O}$,
   (b) $\mathcal{O}$ is closed under finite intersections.

   These are conditions allowing one to consider $\mathcal{O}$ as a topology basis, cf ((Kelley, 1955, p.47)).

3. Prove that the following sets are random closed sets:
   (a) Random singleton $Z := \{X\}$ where X is a random element on $(\mathbb{E}, \mathcal{B})$.
   (b) Random orthant $Z := (= \infty, X_1] \times \ldots \times (-\infty, X_d]$ where $(X_1, \ldots, X_d)$ is a random vector in $\mathbb{R}^d$.
   (c) Random triangle Z in $\mathbb{R}^d$ generated by a random vector (X_1, X_2, X_3) of its vertexes.

4. Assume $\mathbb{E}$ is a metric space with distance d. Let R be a non-negative random variable and ξ be a random element in $\mathbb{E}$. Prove that the random (closed) ball $Z = B_{\xi}(R)$ centered at ξ with radius R is a random closed set and show that its capacity functional is equal to $T_Z(K) = \mathbf{P}(R \geq d(\xi, K))$, where $d(x, K)$ is the distance between x and the set K.

5. Let $\{\xi(t)\}_{t \in \mathbb{E}}$ be a random field on a LCSCH space $\mathbb{E}$, with values in $\mathbb{R}$ and continuous sample paths. Prove that the level set $Z := \{x \in \mathbb{E} : \xi(x) \geq u\}$, with a given $u \in \mathbb{R}$, is a random closed set the space of closed subsets $\mathcal{F}$ of $\mathbb{E}$, with capacity functional $T_Z(K) = \mathbf{P}(\sup_{x \in K} \xi(x) \geq u)$.

6. Prove Rényi's theorem saying that the void probabilities characterize the distribution of a simple point process. Hint: In view of Remark 13.1.5 and Example 13.1.6 it is enough to prove that the sets of the form $\{\mu \in \mathbb{M} : \mathrm{supp}(\mu) \in B_{\mathcal{F}}\}$ for all $B_{\mathcal{F}} \in \mathcal{B}_{\mathcal{F}}$ generate the σ-field $\mathcal{M}$ on the space of counting measures $\mathbb{M}$. Use one of the representations of $\mathcal{B}_{\mathcal{F}}$ given in point 4 of Remark 13.1.1.

7. Prove that the capacity functional $T_{V_*}(K)$ of the centered zero-cell of a stationary point process on $\mathbb{R}^d$ of finite, positive intensity λ can be expressed using the typical cell V under the Palm probability $\mathbf{P}^0$ of Φ as $T_{V_*}(K) = \lambda \mathbf{E}^0 [|V| \mathbf{1}(V \cap K \neq 0)]$. Hint: See the Inverse formula of the Palm calculus in the lesson on Stationary Voronoi tessellations.

8. *Computer exercise.* Simulate a motion invariant Poisson line process in a square window on the plane; cf Exemple 13.2.3.

Lesson 14

Boolean model and coverage processes

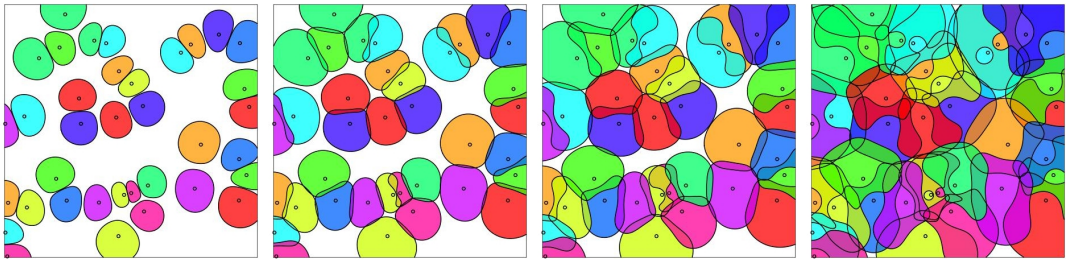


Figure 14.1: Three instances of a shot-noise germ-grain coverage process proposed to model wireless coverage cells, see Exercise 14.1.9.

Coverage processes serve as very general mathematical models for irregular geometrical patterns. They have many applications, traditionally in material and biological sciences but also, more recently, in communication networks, in particular wireless communications. In principle, a general random closed set considered in the previous lesson can be considered as a coverage model. However, really interesting coverage models, penetrating the whole space, are constructed via point processes of closed sets. Particularly popular coverage models on $\mathbb{R}^d$ arise via a so-called germ-grain construction. A very prominent example of coverage process, considered in this lesson, is a Boolean coverage model. It also admits a germ-grain construction on $\mathbb{R}^d$.

For more reading on coverage processes see a monograph book Hall ((1988)) and also ((Chiu et al., 2013, Chapter 6)).

14.1 Coverage process

14.1.1 General coverage processes

A coverage process in a general LCSCH space $\mathbb{E}$ is defined as the union of the set-atoms of a point process on the space of non-empty closed subsets $\mathcal{F}'$ of $\mathbb{E}$; see the lesson on the random closed sets. We have called such point processes set processes. Sometimes one assumes requires these set processes to have compact set-atoms.

Let $\Phi_f = \sum_i \delta_{F_i}$ be a set process on $\mathbb{E}$ i.e., a point process on the space $\mathcal{F}'$. Φ_f generates a *coverage process* being the union of its set-atoms

$$\Xi := \bigcup_i F_i. \tag{14.1.1}$$

Fact 14.1.1. *The set Ξ is a random closed set with the capacity functional $T_\Xi(K) := \mathbf{P}\{\,\Xi \cap K \neq \emptyset\,\}$ related to the void probabilities $v_f(\mathcal{F}_k) := \mathbf{P}\{\,\Phi_f(\mathcal{F}_k) = 0\,\}$ of the set process Φ_f by the expression*

$$T_\Xi(K) = 1 - v_f(\mathcal{F}_K). \tag{14.1.2}$$

for $K \in \mathcal{K}$.

Proof. The equation (14.1.2) is trivial, with both sides expressing the probability that at least one atom of Φ_f, and hence Ξ, hits K. The proof of the closure and measurability of Ξ follows the same lines as the proof that a general Boolean model is a random closed set, cf the lesson on Random closed sets. We repeat them in what follows for completeness. Note that for all $K \in \mathcal{K}$

$$\#\{F \in \Phi_\mathrm{f} : F \cap K \neq \emptyset\} = \#\{F \in \Phi_\mathrm{f} \cap \mathcal{F}_K\} = \Phi_\mathrm{f}(\mathcal{F}_K) < \infty, \tag{14.1.3}$$

with the last inequality following from the fact that $\mathcal{F}_K$ is compact and Φ_f is a point process on $\mathcal{F}'$. This implies that Ξ, which is the union (14.1.1) of (possibly countably many) closed sets is a closed set. Indeed, points of a convergent sequence on Ξ are eventually contained in a compact subset K of $\mathbb{E}$. By (14.1.3) they belong hence to some union of finitely many set-atoms F of Φ_f. As a finite union of closed sets, it is a closed set, thus containing the sequence limit.

Regarding $\mathcal{B}_{\mathcal{F}'}$ measurability of Ξ, note $\{\Xi \in \mathcal{F}^K\} = \{\Xi \cap K = \emptyset\} = \{\Phi_f(\mathcal{F}_K) = 0\}$, for all $K \in \mathcal{K}$, with the last event being measurable since Φ_f is a point process on $\mathcal{F}'$. □

Remark 14.1.2. The difference between two models Φ_f and Ξ lies in the fact that in Ξ we do not observe the shapes of individual set-atoms F_i of Φ_f. In particular, we do not recognize their boundaries when they overlap. This makes Ξ more appropriate model for some statistical analysis.

Theoretically, every non-empty random closed set Ξ can be trivially represented as in (14.1.1), with the set process Φ_f having just one atom equal to Ξ. The problem becomes only slightly more complicated if we want Φ_f to have compact set-atoms F_i. In this case it is enough to "decompose" Ξ using a countable, compact cover of $\mathbb{E}$ (it exists since $\mathbb{E}$ is a LCSCH space).

The problem is non-trivial if Ξ is stationary and/or isotropic coverage process on $\mathbb{R}^d$ (its distribution is invariant with respect to all translations and/or rotations) and we want Φ_f to have the same properties.

The most important example of a general coverage model is a general Boolean model, which is generated by a Poisson set process.

Example 14.1.3 (General Boolean model)**.** Let Φ_f be a Poisson point process on $\mathcal{F}'$ with intensity measure Λ_f. The corresponding coverage model Ξ is called the *Boolean model*, or sometimes *Poisson-Boolean model*. By (14.1.2) the capacity functional of the Boolean model can be expressed as

$$T_\Xi(K) = 1 - e^{-\Lambda_f(\mathcal{F}_K)} \tag{14.1.4}$$

for $K \in \mathcal{K}$.

14.1.2 Characteristics of a general coverage process

The following characteristics of the coverage process Ξ are of particular interest. In fact they can be considered for a general random closed set Ξ. Some of them can be explicitly expressed using the capacity functional.

(Full) coverage functional $\mathbf{P}\{\, K \subset \Xi \,\}$ for, say, compact $K \in \mathcal{K}$. Obviously $\mathbf{P}\{\, K \subset \Xi \,\} \le T_\Xi(K)$ but in general the coverage functional does not admit any explicit expression in terms of the capacity functional. Exceptions are some hard-core coverage models, cf Exercises 4 and 6d. More tight bounds and asymptotic expression are known for some Boolean models; see Section 14.2.

Coverage functions A *multi-dimensional coverage function*

$$p(x_1, \dots, x_n) := \mathbf{P}\{\, \{x_1, \dots, x_n\} \subset \Xi \,\}$$

can be seen as a discrete approximation of the full coverage functional taking $K = \{x_1, \dots, x_n\}$, $n \ge 1$. It admits explicit expressions in terms of the capacity functional T_Ξ, cf Exercise 1.

In particular, for $n = 1$, the *coverage function* $p(x) := \mathbf{P}\{\, x \in \Xi \,\} = T_\Xi(\{x\})$ coincides with the capacity and coverage functional of $K = \{x\}$, $x \in \mathbb{E}$.

For $n = 2$, $p(x_1, x_2)$ is called *covariance function* of Ξ. It is indeed a non-centered covariance function of the random field $\{I(x) = \mathbf{1}(x \in \Xi) : x \in \mathbb{E}\}$ meaning that $\mathbf{E}\,[I(x_1)I(x_2)] = p(x_1, x_2)$.

Mean measure If some deterministic measure μ is considered on $\mathbb{E}$ (for example the Lebesgue measure on $\mathbb{E} = \mathbb{R}^d$) then one can define a *mean measure of* Ξ *related to* μ $M_\Xi^\mu(B) := \mathbf{E}\,[\mu(\Xi \cap B)]$ for, say, compact, $K \in \mathcal{B}$. Note that the coverage function $p(x)$ is the density of this measure with respect to μ

$$M_\Xi^\mu(K) = \mathbf{E}\,[\mu(\Xi \cap K)]$$

$$= \mathbf{E}\left[\int_K \mathbf{1}(x \in \Xi)\, \mu(\mathrm{d}x)\right]$$

$$= \int_K \mathbf{P}\{\, x \in \Xi\,\}\ \mu(\mathrm{d}x)$$

$$= \int_K p(x)\, \mu(\mathrm{d}x). \tag{14.1.5}$$

The mean measure of coverage process on $\mathbb{R}^d$ with respect to the Lebesgue measure $\mu(\mathrm{d}x) = \mathrm{d}x$ is called the *mean volume measure* of Ξ and denoted by M_Ξ. Note that the mean (volume) measure of Ξ is *not* equal to the mean measure of the underlying set-process.

14.1.3 Stationary and motion invariant coverage processes on $\mathbb{R}^d$

If Ξ is a stationary coverage process on $\mathbb{R}^d$ then its capacity functional is invariant with respect to any translation $T_\Xi(K) = T_\Xi(K+a)$ for all $a \in \mathbb{R}^d$. In this case $p(x_1, x_2, \ldots, x_n) = p(0, x_2 - x_1, \ldots, x_n - x_1)$ and, in particular, we define the *reduced covariance function* $C(x) := p(0, x)$.

Moreover, the coverage function is a constant $p(x) = p$ called the *volume fraction* of Ξ. Indeed, for any compact $K \in \mathcal{K}$, the expected volume of K covered by Ξ is equal by the definition to mean volume measure of Ξ on K, $M_\Xi(K)$, and by (14.1.5) it admits $p(x) = p$ as its density. Consequently

$$\mathbf{E}\left[\frac{|\Xi \cap K|}{|K|}\right] = \frac{M_\Xi(K)}{|K|} = \frac{p|K|}{|K|} = p\,. \tag{14.1.6}$$

Isotropy of a random closed set Ξ is defined as the invariance of its distribution with respect to all rotations around some fixed point, that can be considered as the origin. It is usually considered together with the stationarity; the joint property is called *motion invariance*. The distribution of a motion invariant coverage set Ξ is invariant with respect to all rotations around any center point. This implies the invariance of $T_\Xi(K)$ with respect to all rotations [1]. In particular, the covariance function $p(x_1, x_2)$ depends only on the distance $|x_1 - x_2|$ and one defines the *distance covariance function* $\bar{C}(r) := C(|x|) = p(0, x)$.

Contact distribution functions

These are some probability distribution functions meant to provide some information on the local geometry of the coverage process Ξ around a point non-covered by Ξ. They are usually defined for a stationary Ξ with respect to the origin. (The generalization to a general non-stationary coverage processes on $\mathbb{R}^d$ with respect to an arbitrary point is straightforward.)

Let $K \in \mathcal{K}' := \mathcal{K} \setminus \{\emptyset\}$ be a given non-empty compact subset of $\mathbb{R}^d$ containing the origin $0 \in K$. The *contact distribution function* $H_K(r)$ of Ξ with respect to the *test set* K is defined

[1] Observe that the isotropy of Ξ alone does not imply this property!

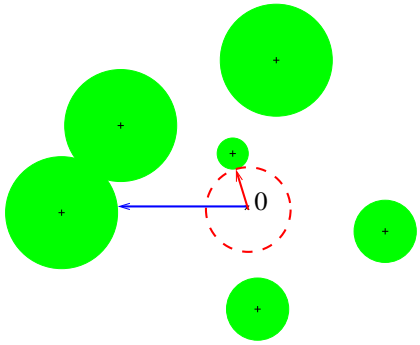


Figure 14.2: The conditional distribution functions of the radius of the smallest sphere centered at 0 and hitting Ξ and the shortest segment joining 0 with Ξ in the direction of the blue vector, given 0 is not covered by Ξ, are called, respectively, the spherical and linear contact distribution functions (in the given direction).

as the conditional probability that Ξ does not hit the dilation of the set K by the factor r, $rK := \{ry : y \in K\}$, given it does not hit 0,

$$\begin{aligned} H_K(r) &:= \mathbf{P}\{\, \Xi \cap rK = \emptyset \,|\, 0 \notin \Xi \,\} \\ \boxed{0 \in rK} &= \frac{\mathbf{P}\{\, \Xi \cap rK = \emptyset \,\}}{1-p}, \\ &= \frac{1 - T_\Xi(rK)}{1-p}, \quad r \geq 0\,, \end{aligned} \tag{14.1.7}$$

where $p = T_\Xi(\{0\})$ is the volume fraction of Ξ. Thus the contact distribution function admits the explicit expression (14.1.7) in therms of the capacity functional.

Different instances of contact distribution functions can be considered, depending on the choice of the test set K. The most popular cases are as follows:

- *The spherical contact distribution function.* This is the case with the spherical test set $K = B_0(1)$ being the ball centered at the origin with unit radius. In this case the $H_{B_0(1)}(r)$ is the conditional distribution function of the distance from 0 to Ξ given $0 \notin \Xi$; see Figure 14.2.

- *The linear contact distribution function.* This case arises when $K = [0, x]$, a unit length segment from the origin in the direction of some $x \in \mathbb{R}^d$, $|x| = 1$. In this case $H_{[0,x]}(r)$ is the conditional distribution function of the distance from 0 to Ξ in the direction of the vector x, given $0 \notin \Xi$.

  If Ξ is isotropic (rotation invariant distribution) then the linear contact distribution function does not depend on the direction x and $H_{[0,x]}(r) = H(r)$ can be seen as the conditional distribution function of the distance from the origin to Ξ in a *randomly chosen direction*; see Figure 14.2.

14.1.4 Germ-grain coverage process in $\mathbb{R}^d$

Recall, the general coverage process is constructed as a union of set-atoms of a set process. A germ-grain model arises when the set-process is first constructed from a usual point process on $\mathbb{R}^d$ marked by some random non-empty closes or compact sets. Points of this process are called *germs*, and the marks *grains*. Let

$$\tilde{\Phi} = \sum_i \delta_{(X_i, Z_i)} \,. \tag{14.1.8}$$

by such a process, with germs $X_i \in \mathbb{R}^d$ and grains $Z_i \in \mathcal{K}' = \mathcal{K} \setminus \{\emptyset\}$ non-empty compact sets of $\mathbb{R}^d$. In the most general setting the grains might be dependent and $\tilde{\Phi}$ is considered as a point process on $\mathbb{R}^d \times \mathcal{K}'$. This point process $\tilde{\Phi}$ generates a set process i.e., a point process Φ_f on $\mathcal{K}' \subset \mathcal{F}'$

$$\Phi_f = \sum_{(X_i, Z_i) \in \tilde{\Phi}} \delta_{X_i + Z_i} \,; \tag{14.1.9}$$

Φ_f is indeed a set-process, by the fact that the mapping $(x, K) \mapsto x + K$ from $\mathbb{R}^d \times \mathcal{K}'$ to $\mathcal{K}'$ is measurable and $\{(x, K) \in \mathbb{R}^d \times \mathcal{K}' : (x + F) \cap K' = \emptyset\}$ are compact sets in $\mathbb{R}^d \times \mathcal{K}'$ for any $K' \in \mathcal{K}'$. The set process Φ_f defines the corresponding *germ-grain coverage process*

$$\Xi := \bigcup_{(X_i, Z_i) \in \tilde{\Phi}} (X_i + Z_i) \,. \tag{14.1.10}$$

Example 14.1.4 (Boolean germ-grain coverage model)**.** Let $\tilde{\Phi}$ in (14.1.8) be a Poisson process on $\mathbb{R}^d$ with non-null intensity measure Λ, independently, identically marked by random, nonempty compact subsets $Z_i \in \mathcal{F}'$ of $\mathbb{R}^d$. We assume that

$$\mathbf{E}\left[\Lambda(\check{Z}_0 \oplus K)\right] < \infty \quad \text{for all } K \in \mathcal{K} \tag{14.1.11}$$

where Z_0 is the generic grain, with $A \oplus B = \{x + y : x \in A, y \in B\}$ and $\check{A} = \{-x : x \in A\}$. We have already argued that the corresponding set process $\Phi_f := \sum_{(X_i, Z_i) \in \tilde{\Phi}} \delta_{X_i + Z_i}$ is a Poisson set process with intensity measure $\Lambda_f(\mathcal{F}_K) = \mathbf{E}\left[\Lambda(\check{Z}_0 \oplus K)\right]$, for $K \in \mathcal{K}$ (cf Example 2.13 in the lesson on Random closed sets). The coverage process Ξ (14.1.10) generated by Φ_f is hence a Boolean coverage process. We call it *Boolean germ-grain coverage model*. By (14.1.4), the capacity functional of the Boolean germ-grain coverage model Ξ can be expressed as

$$T_\Xi(K) = 1 - e^{-\mathbf{E}\left[\Lambda(\check{Z}_0 \oplus K)\right]} \tag{14.1.12}$$

for $K \in \mathcal{K}$.

Example 14.1.5 (Homogeneous Boolean germ-grain model)**.** This is a special case of a Boolean germ-grain coverage model generated by a homogeneous Poisson process with intensity λ on $\mathbb{R}^d$ (i.e., $\Lambda(\mathrm{d}x) = \lambda \, \mathrm{d}x$) and general i.i.d. non-empty compact germs $Z_i \in \mathcal{K}'$ having finite expected volume $\mathbf{E}\left[|Z_0|\right] < \infty$ which is equivalent to (14.1.11). Ξ in this case is a stationary coverage process.

The volume fraction of the homogeneous BM with intensity λ and generic grain Z_0 admits the following expressions for the volume fraction and the reduced covariance function

$$p = 1 - e^{-\lambda \mathbf{E}[|Z_0|]}, \tag{14.1.13}$$

$$C(x) = 2p - 1 + (1-p)^2 e^{\lambda \mathbf{E}[|Z_0 \cap (Z_0 - x)|]}; \tag{14.1.14}$$

cf Exercise 3.

Example 14.1.6 (Homogeneous Boolean germ-grain model with random spherical grains)**.** This is a special case of a homogeneous Boolean germ-grain coverage model with grains being balls $Z_i = B_0(R_i)$ centered at the origin, with i.i.d. radii R_i distributed as some generic random variable R. Note that the $\mathbf{E}[|Z_0|] < \infty$ is equivalent to $\mathbb{E}[R^d] < \infty$. This is an example of a motion invariant coverage process. The simples example corresponds to constant deterministic radius of balls $R_i \equiv R$.

Boolean germ-grain model is one of the most popular models in stochastic geometry. It is an example of a germ-grain model with independent grains. In what follows we recall and present other models with dependent grains.

Example 14.1.7 (Hard-core coverage models)**.** General hard-core coverage models are generated by set-processes with non-overlapping set-atoms. Special cases of germ-grain hard-core models are generated by three Matérn hard-core point processes in $\mathbb{R}^d$ considered in the lesson on Hard-core point processes. Recall, these are the process of isolated Poisson points Φ_1, Matérn type I process Φ_2 and the RSA point process Φ_3, all respecting some exclusion distance R between points. Consequently, assuming spherical germs $Z_i := B_0(R/2)$ of radius $R/2$ one obtains hard-core germ-grain models. Recall, we have been comparing in the aforementioned lesson the volume fractions of these coverage processes. In general, the volume fraction and even the full coverage functional of hard-core models admit some special expressions, cf Exercises 4, 6c and 6d.

Example 14.1.8 (Johnson-Mehl coverage process)**.** Note that the Voronoi tessellation generated by a (say simple) point process $\Phi = \sum_i X_i$ in $\mathbb{R}^d$ can be also seen as a germ-grain model generated by the points of the process, with grains being the centered Voronoi cells

$$Z_i := C_i - X_i = \{y \in \mathbb{R}^d : |y - X_i| \leq \min_{X_j \in \Phi} |y - X_j|\} - X_i.$$

The coverage properties of $\Xi = \mathbb{R}^d$ in this case are trivial. However, the following modification of the Voronoi cells leads to a non-trivial coverage process. In fact, this is a family of coverage processes parametrized by $r > 0$, having Voronoi grains Z_i restricted to the ball $B_0(r)$

$$Z_i^{(r)} := Z_i \cap B_0(r).$$

The parameter r might be considered as the growth time. Initially each grain grows spherically in all directions but as soon two grains touch each other at some location, they stop growing in

the "blocked" direction. Johnson-Mehl is "almost"a hard core model with the d-dimensional volume of the overlapping of set-atoms equal to 0. As $r \to \infty$, $Z_i^{(r)} \to Z_i$. The third picture on Figure 14.1 looks as a "soft" version of the Johnson-Mehl model, allowing for a small overlapping.

Example 14.1.9 (Shot-noise coverage model)**.** This is a germ-grain coverage process on $\mathbb{R}^d$ in which the sets $F_i = X_i + Z_i$ are defined as the regions of the space where the "impact" of the point X_i exceeds the cumulative (additive) effect of the impacts of all other points. In the simplest scenario the impact is modeled be a deterministic, decreasing function of the distance. The cumulative effect is called the shot-noise process associated with the point process and the impact function (called also the response function). This germ-grain coverage process was initially proposed to model wireless coverage cells, with the shot-noise modeling the interference in radio communication. Changing model parameters, such processes can exhibit a wide range of coverage patterns approaching Boolean models, on one side and Voronoi tessellation and Johnson-Mehl model on the other, including hard core scenarios. Three instances of a shot-noise germ-grain coverage process are shown on Figure 14.1. For more reading on this model see ((Chiu et al., 2013, Section 6.5.4)), ((Baccelli and Błaszczyszyn, 2009, Chapter 7)). Some more advanced quantitative coverage results are presented in ((Błaszczyszyn et al., 2018, Chapters 5-7)).

14.2 More advanced coverage results for Boolean models

Asymptotic results

Some informative results for the full coverage of the Boolean models are known in asymptotic form when germs are dense and grains are small.

As an example of such result, consider the following parametric family of homogeneous Boolean model on the plane $\mathbb{R}^2$ with spherical grains of generic random radius rR, $r > 0$ and R having finite second moment (cf Example 14.1.6)

$$\Xi(r) = \bigcup_i \left(X_i + B_0(rR_i)\right). \tag{14.2.1}$$

where $\{X_i\}$ are the atoms of a homogeneous Poisson process on the plane of intensity λ.

Proposition 14.2.1. *Let K be a compact set in $\mathbb{R}^2$ whose boundary ∂K has zero 2-D Lebesgue measure, i.e. $|\partial K| = 0$. Consider the family of Boolean models (14.2.1) with intensity of germs λ and assume that $\mathbf{E}\left[R^{2+\epsilon}\right] < \infty$ for some $\epsilon > 0$. For a given u $(-\infty < u < \infty)$ take*

$$\lambda = \lambda(r) = \frac{1}{\pi r^2 \mathbf{E}\left[R^2\right]}\left(\log \frac{|K|}{\pi r^2 \mathbf{E}\left[R^2\right]} + 2\log\log \frac{|K|}{\pi r^2 \mathbf{E}\left[R^2\right]} + \log \frac{\mathbf{E}\left[R\right]^2}{\mathbf{E}\left[R^2\right]} + u\right) \tag{14.2.2}$$

or

$$r = r(\lambda) = \sqrt{\frac{\log\lambda + \log\log\lambda + \log\left(|K|\mathbf{E}\left[R\right]^2/\mathbf{E}\left[R^2\right]\right) + u}{\lambda\pi\mathbf{E}\left[R^2\right]}}. \tag{14.2.3}$$

Then

$$\mathbf{P}\{\, K \subset \Xi \,\} = \exp[-e^{-u}] + o(1)$$

as $r \to 0$ *or* $\lambda \to \infty$*, respectively.*

Proof. The result follows from an original results proved in ((Janson, 1986, cf. Lemma 7.3)), where a more general condition on r and λ going to 0 and ∞, respectively, is given. Dimension $d \geq 2$, more general grains, and multiple coverage is considered in the aforementioned paper too. □

Note that the above result gives approximations of the coverage probability for dense Boolean models with small grains. The two functions $\lambda(r)$ and $r(\lambda)$ are of interest if one wants to design a Boolean model with some given (approximate) probability of coverage of a given set K. The first one shows how many germs of a given size are needed, while the second one indicates how large grains should be if the density of germs is given.

Bound on the full coverage

The following bounds have been shown in ((Hall, 1988, Theorem 3.11)) in the case of the Boolean model with grains of *fixed (deterministic) radius*, see also ((Chiu et al., 2013, Eq. (3.100))).

Proposition 14.2.2. *Let* $\Xi(r)$ *be the homogeneous Boolean model given by (14.2.1) with constant* $R_i \equiv 1$ *and intensity of grains* λ*. Then for* $K = [0,1]^2$*,* $\lambda \geq 1$*,* $r \leq 0.5$

$$1 - 3\min\{1, (1+\pi r^2\lambda^2)e^{-\pi r^2\lambda}\} < \mathbf{P}\{\, K \subset \Xi \,\} < 1 - \frac{1}{20}\min\{1, (1+\pi r^2\lambda^2)e^{-\pi r^2\lambda}\}\,.$$

14.3 Exercises

1. Express $p(x_1, x_2) = \mathbf{P}\{\, x_1, x_2 \in \Xi \,\}$ in terms of the capacity functional of Ξ.

2. Let Ξ be a general coverage model with capacity functional T_Ξ, generated by a set process Φ_f with Laplace transform $\mathcal{L}_{\Phi_f}$. Prove the following relation for $K \in \mathcal{K}$

$$T_\Xi(\mathcal{F}_K) = 1 - \mathcal{L}_{\Phi_f}(\varphi_K)\,,$$

   where φ_K is a set function $\varphi_K(F) = \log 1(F \cap K = \emptyset)$ defined for $F \in \mathcal{F}'$. Show that this is a generalization of the relation (14.1.4) valid for the Boolean model.

3. Prove the expressions (14.1.13) and (14.1.14) for the volume fraction and the covariance function of a homogeneous germ-grain Boolean model; cf Example 14.1.5.

4. Let $\Phi_f = \sum_i \delta_{F_i}$ be a hard core set-process on $\mathbb{R}^d$, (i.e. the set-atoms F_i do not overlap) with F_i being balls (possibly with random radii). Examples are the three Matérn models recalled in the Example 14.1.7. Let Ξ be the corresponding coverage process (14.1.10). Argue that for any *connected* K the full coverage functional is equal to the mean measure of Φ_f on $\mathcal{F}_K$

$$\mathbf{P}\{\, K \subset \Xi \,\} = \mathbf{E}\,[\Phi_f(\mathcal{F}_K)]\ ,$$

   Hint: $K \subset \Xi$ iff K is entirely contained in of the spherical set-atoms F_i.

5. Let Ξ be a homogeneous germ-grain Boolean coverage model driven by a homogeneous Poisson process with intensity λ, $0 < \lambda < \infty$ on $\mathbb{R}^d$ and generic grain Z_0. Show that the condition $\mathbf{E}\,[|Z_0|] < \infty$ is equivalent to the volume fraction being strictly smaller than 1, $p < 1$.

6. *Stationary germ-grain coverage processes.* Let $\tilde{\Phi} = \sum_{X_i, Z_i}$ be a stationary germ-grain model on $\mathbb{R}^d$ as in (14.1.8), considered on some stationary framework (see the corresponding lesson) $(\Omega, \mathcal{A}, \{\theta_x\}_{x\in\mathbb{R}^d}, \mathbf{P})$, with intensity of germs λ, $0 < \lambda < \infty$, and its Palm probability $\mathbf{P}^0$.

   Note that this means that the grains Z_i are invariant with respect to any translation of germs. Using the flow of the stationary framework this means $Z_i = Z \circ \theta_{X_i}$ for some random closed set Z defined on the considered probability space. Stationary Voronoi tessellation and the Johnson-Mehl coverage model are examples of such process. Any stationary germ-grain process with independent germs can also be also cast into this framework.

   Denote by Φ_f the corresponding set process (14.1.9) and its set-atoms by $F_i := X_i + Z_i$. Let Ξ be the corresponding coverage process (14.1.10). Denote the

   Prove the following relations between the stationary and Palm characteristics of the germ-grain model.

(a) Using the CLMM formula show for any $K \in \mathcal{K}$

$$\mathbf{E}\left[\Phi_f(\mathcal{F}_k)\right] = \lambda \mathbf{E}^0\left[|\check{Z}_0 \oplus K|\right] . \tag{14.3.1}$$

On the left-hand side we have the stationary expected number of set-atoms of Φ_f hitting K. On the right-hand side, under $\mathbf{P}^0$, the random set $Z_0 = Z \circ \theta_0$ is the grain of the germ located at the origin In case of independent grains (as in the Boolean germ-grain model) the distribution of Z_0 under $\mathbf{P}^0$ coincides with the distribution of any grain (we have called it "generic"grain in the description of the Boolean germ-grain model).

(b) Assuming $K = \{0\}$ in (14.3.1) derive the the following *Little's law* for the stationary germ-grain coverage processes

$$\mathbf{E}\left[N\right] = \lambda \mathbf{E}^0\left[|Z_0|\right], \tag{14.3.2}$$

where $\mathbf{E}\left[N\right]$ is the expected number of set-atoms F_i of the stationary model covering the origin, called the *mean coverage number*, and $\mathbf{E}^0\left[|Z_0|\right]$ is the volume of the typical grain.

(c) Assume that with probability $\mathbf{P}$ one the origin is covered by at most one set-atom F_i of Φ_f. (This is trivially true when Φ_f is a hard core set-process, i.e., F_i do not overlap. This is also true for the Voronoi tessellation and hence the Johnson-Mehl model since the stationary point processes do not have points equidistant to the origin.) Using (14.3.2) show that the volume fraction p of Ξ is equal to

$$p = \lambda \mathbf{E}^0\left[|Z_0|\right] .$$

(d) Assume that Φ_f is a hard-core set process with spherical set-atoms (F_i are possibly random balls). Using the relation given in Exercise 4 show that for any connected K, the full coverage functional admits the following expression

$$\mathbf{P}\{\, K \subset \Xi \,\} = \lambda \mathbf{P}^0\{\, K \subset Z_0 \}\,.$$

7. *Computer exercise.* Using *spatstat*:

(a) Generate a homogeneous Boolean germ-grain model with fixed spherical grains.

(b) Calculate and plot the theoretical spherical contact distribution function for this model.

(c) Using *spatsat* command `Hest` estimate the spherical contact distribution function on a reasonably large realization of the Boolean model.

(d) Consider the spatial mosaic of vegetation of the heather plant shown on Figure 14.3. The corresponding data set is available in *spatstat*, (cf Baddeley et al.). Estimate its spherical contact distribution function.

(e) Using the theoretical spherical contact distribution function of the homogeneous Boolean germ-grain model with fixed spherical grains try to fit it to this data playing with its two parameters: the density of germs and the radius of germs.

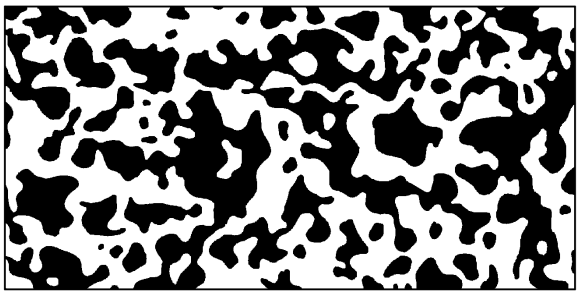


Figure 14.3: The spatial mosaic of vegetation of the heather plant; data set available in *spatstat*.

Lesson 15

Connectedness of random sets and continuum percolation

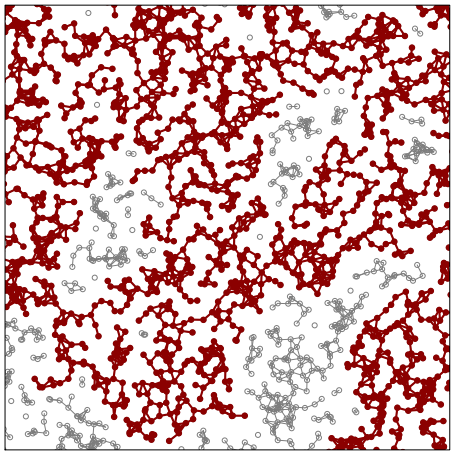


Figure 15.1: Random geometric (Gilbert) graph on Poisson process close to the phase transition related to its percolation, observed in a finite window. The largest component observed in the window is highlighted.

In the two previous lessons we have been mainly interested in coverage properties of random sets, with the fundamental characteristic, the capacity functional, expressing the probability of hitting some given test set. Now, we focus on the connectivity properties of random sets, in particular coverage processes in $\mathbb{R}^d$. We should say first that the term "connectivity", albeit often used, is not very precise. It embraces different characteristics allowing one to capture the structure of the fragmentation of the considered set. Two fundamental characteristics considered in this regard are the *connectedness*, telling whether the set is "in one piece" and *percolation*, which we understand here as the existence of a (topologically)

unbounded connected component, meaningful only if the considered set is itself unbounded. More fine descriptors of the connectivity structure can be given on the ground of algebraic topology Bobrowski and Kahle ((2014)).

Several mathematical and physical theories seek to tackle these problems and provide some answers, in particular: stochastic geometry, both mathematical and physical percolation theory, and the aforementioned algebraic topology. Many results are of numerical nature, obtained via theoretically backed up simulations.

So much scientific effort is devoted to these questions, because they have many applications. Traditionally, in geology and material sciences, where understanding the connectivity properties of porous media is crucial. Astronomy is perhaps a more surprising domain of applications, where one studies the distribution of matter in the Universe Kerscher et al. ((2001)). Recent massive deployment of wireless communication networks motivates the development and the study of new models in this matter Franceschetti and Meester ((2008)).

Interestingly enough, the study of the connectivity properties of geometric graphs was initiated by Edgar Nelson Gilbert [1] in his paper Gilbert ((1959)), with wireless communication networks as a suggested application: nodes represent wireless devices distributed in the plane according to some point process and connected to each other if their euclidean distance is less than some critical transmission range. Since the publication of this seminal paper, this model with points distributed according to a homogeneous Poisson process, became the fundamental model in the theory of continuum percolation. In the literature it is called random geometric graph (RGG) or Gilbert graph. It captures in an equivalent way the connectivity of the Boolean model with fixed spherical grains.

In the remaining part of this section we briefly introduce some general notions and comment on existing general results. In two subsequent sections we consider homogeneous Boolean models with spherical grains in $\mathbb{R}^d$ and present two fundamental results for them, regarding local full connectedness and percolation in the whole space. For more reading on mathematical theory of continuum percolation see Meester and Roy ((1996)).

15.1 Connectedness and percolation

15.1.1 General random closed sets

Informally a set is considered as connected if it is in one piece. This intuitive notion requires some mathematical clarification, especially when applied to subsets of abstract spaces.

[1] An American mathematicians with wide interest in communication theory; he also proposed the "version" of the Erdős-Rényi graph that we have studied in a previous lesson.

Topological connectedness

A subset $B \subset \mathbb{E}$ of a topological space $\mathbb{E}$ is said to be *connected* if it cannot be represented as the union of two or more disjoint nonempty open (equivalently, closed) subsets [2]. B is called *disconnected* if it is not connected. Some, perhaps surprising, examples of connected and disconnected sets are given in Exercise 1.

For $B \subset \mathbb{E}$ and $x, y \in B$, we say x and y are connected in B, denoted $x \overset{B}{\sim} y$, if there exists a connected subset $B' \subset B$ containing x and y, $x, y \in B$. This is an equivalence relation. *Connected components* of B are the equivalence classes of the relation $\overset{B}{\sim}$.

A connected component $C_x(B)$ of $x \in B$ is the union of all connected subsets of B containing x; i.e., the maximal connected subset of Ξ containing x. Equivalently $C_x(B) = \{y \in B : x \overset{B}{\sim} y\}$.

Connectedness of random closed sets

In order to be able to speak about the probability that a random closed set is connected or about the distribution of the number of its connected components one needs a few measurability results, which we summarize in what follwos; cf ((Tsirelson, 2013, Chapter 5)).

Remark 15.1.1. Let $\mathbb{E}$ be a CSCH space, (the whole space $\mathbb{E}$ is assumed to be compact). The following subsets are measurable sets of the respective spaces.

1. $\{F \in \mathcal{F} : F \text{ is connected}\} \in \mathcal{B}_{\mathcal{F}}$.
2. $\{(F, F') : F \text{ is a connected component of } F\} \in \mathcal{B}_{\mathcal{F}} \otimes \mathcal{B}_{\mathcal{F}}$.

Let Ξ be a random closed set on the space $\mathcal{F}$ of $\mathbb{E}$. Then

1. $\{\Xi \text{ is connected}\}$ is a measurable event.
2. For any $x \in \mathbb{E}$, the connected component $C_x(\Xi)$ of x in Ξ, is a random closed set.
3. The number of connected components of Ξ hitting a given closed or open $B \subset \mathbb{E}$ is a random variable taking values in $1, \dots, \infty$.
4. The set of connected components of Ξ is a set-process (point process on $\mathcal{F}$) provided the number of connected components hitting any compact $K \subset \mathbb{E}$ is finite. [3]

The facts presented in Remark 15.1.1 allow one to study the connectedness of a random closed set *in a compact window* $K \subset \mathbb{E}$ for any LCSCH space. The compactness assumption cannot be relaxed in general; cf ((Tsirelson, 2013, a remark after 5a9 Core exercise)). Whether the statements of Fact 15.1.1 hold true in a specific case of $\mathbb{E} = \mathbb{R}^d$ (which is not compact) seems to be an open question. As a palliative solution for a random closed set in a LCSCH space $\mathbb{E}$, as e.g. $\mathbb{R}^d$, ((Tsirelson, 2013, 5a10 Remark)) suggests to consider one-point compactifiation $\mathbb{E} \cup \{\boldsymbol{\infty}\}$ of $\mathbb{E}$; cf Exercise 1. When doing this one merges all unbounded connected components of $\mathbb{E}$ into one connected component of $\mathbb{E} \cup \{\boldsymbol{\infty}\}$, thus making meaningless the study of their number. [4]

[2] Equivalently, if it cannot be partitioned into two nonempty subsets, such that each subset has no points in common with the set closure of the other.

[3] If $\mathbb{E}$ is a compact, this means the total number of connected components needs to be finite; cf Remark 15.1.1.

[4] Note that this does not rule out a priori the possibility to study the connectedness of some specifics random

Local connectedness

Remark 15.1.1 allows one to consider the following *local* probabilistic connectivity characteristics of a general random closed set Ξ in any compact subset $K \subset \mathbb{E}$ of a LCSCH space $\mathbb{E}$:

- $\mathbf{P}\{\,\Xi \cap K; \text{is connected}\,\}$,
- Distribution of the connected component of $C_x(\Xi \cap K)$,
- Number of the connected components of $\Xi \cap K$.
- The set process of connected components of $\Xi \cap K$ can be seen as a point process on $\mathcal{F}$, provided the total number of components is finite.

None of these objects can be explicitly related to the capacity functional of Ξ. Most of the existing results regard some specific germ-grain models generated by Poisson process. An example will be given in Section 15.2.1.

Full connectedness and percolation

Many interesting unbounded (not compact) random closed sets, including several coverage process, are disconnected sets in the whole space, often with probability one. An example that we have already seen in the first lesson, is the independent bond percolation model, cf Exercise 3. The same is true for homogeneous Boolean models; cf Exercise 6.

For such sets, a more pertinent question regarding their connectivity properties, is the existence of a non-compact (equivalently unbounded [5]) component. This event is called *percolation* of the corresponding random closed set.

The following general result follows from the ergodic theorem (cf the lesson on the ergodicity of point processes).

Proposition 15.1.2. *Let Ξ be a stationary ergodic germ-grain model (considered in a suitable stationary framework, which is ergodic). Then Ξ is connected (respectively percolates) with probability 0 or 1. Moreover, the number of connected components (respectively unbounded components) of Ξ is almost surely constant.*

Proof. Recall from the aforementioned lesson that ergodicity is equivalent to the metrical transitivity. This latter property means that all events which are $\mathbf{P}$-invariant have probability 0 or 1. Note that the events $\{\Xi \text{ is connected}\}$ and $\{\Xi \text{ percolates}\}$ are invariant w.r.t. all translations, hence $\mathbf{P}$-invariant. Hence the first statement. For the second statement note that the number of connected components and the number of unbounded connected components are random variables measurable with respect to the σ-field of invariant sets. Hence they are almost surely constants. □

closes sets in $\mathbb{R}^d$, in particular the coverage processes (to be proved...). In fact, in what follows we *will* consider the question of the number of connected components of some Boolean models on $\mathbb{R}^d$.

[5] Note that a connected component of a closed set is a closed set. Thus a component is compact iff it is topologically bounded. In the case of germ-gran models with compact grains a connected component is unbounded iff an infinite number of grains has a non-empty intersection with it (*infinite component*).

Example 15.1.3. Examples of stationary, ergodic random closed sets are homogeneous Boolean germ-grain models on $\mathbb{R}^d$. The case of deterministic grains follows immediately form the mixing property (implying ergodicity) of the homogeneous Poisson point process. The general case of (i.i.d.) compact grains requires some additional argument; cf ((Meester and Roy, 1996, Proposition 2.8)).

The fact that (ergodic) random closed sets can either percolate or not with probability one is called the *phase transition* related to percolation. A random closed set is said to be *close to its phase transition* (or close to the *critical state*) when a small modification (e.g. up- or downscaling of grains or another change of one of its parameters) can make it go from the non-percolation regime to the percolation or other way around. We present this phenomenon for the Boolean model in Section 15.2.2.

Percolation and Euler characteristic

The *Euler* or (*Euler-Poincaré*) characteristic χ can be defined for quite general (random) closed sets Ξ on the plane $\mathbb{R}^2$ as the difference between the number of connected components in Ξ and the number holes in Ξ

$$\chi(\Xi) := \#\{\text{connected components in } \Xi\} - \#\{\text{holes in } \Xi\}$$

and for closed sets Ξ in $\mathbb{R}^3$

$$\chi(\Xi) := \#\{\text{connected components in } \Xi\} - \#\{\text{holes in } \Xi\} + \#\{\text{cavities in } \Xi\},$$

provided the respective numbers are finite. These definitions can be made formal and generalized to higher dimensions with the *Betti numbers*, on the ground of algebraic topology Bobrowski and Kahle ((2014)) and with *Minkowski functionals* in integral geometry Mecke ((1997)). They can be extended to unbounded sets, by appropriately normalizing the Euler characteristic of the set observed in the increasing window.

It has been observed many times, without exactly knowing why, that the topological information about Ξ carried by the Euler characteristic is linked to percolation of Ξ, in the sense that the critical behaviour regarding its percolation takes place when the Euler characteristic changes its sign; cf Okun ((1990)); Mecke ((1997)); Roubin and Colliat ((2015)). The relation is not exact, the difference between the theoretically known critical probability of percolation and its topological estimate based on the root of the Euler characteristic can be observed for example for some lattice percolation models, cf Neher et al. ((2008)).

A heuristic explanation for planar sets can be as follows: When the set is very fragmented and far from percolation, there are many connected components and few holes (closed loops containing empty space) making χ positive. "Increasing" the set connectivity one observes merging connected components, which reduces their number, and creation of holes (loops), thus decreasing χ that eventually becomes negative, when there are few components and many holes (loops). Experiments show that $\chi \approx 0$ corresponds to a nearly critical behaviour of the random set. This offers a rough, bur relatively simple way of verifying whether an random closed set observed in a compact window percolates in the whole space.

15.2 Connectedness and percolation of Boolean models

15.2.1 Local connectedness of homogeneous Boolean models

In this section we consider homogeneous Boolean germ-grain models with intensity λ on $\mathbb{R}^d$ and study their full connectivity in a given compact window. As for the full coverage, only asymptotic results are known in the regime when $\lambda \to \infty$ and the size of the grains goes to zero. Here we consider only the case of spherical grains of deterministic radius r. The main result is best formulated in terms of some spanning tree.

Euclidean Poisson minimum spanning tree

Let Φ be a homogeneous Poisson process on $\mathbb{R}^d$ with intensity λ. Let $K \subset \mathbb{R}^d$ be a compact set and denote by Φ_K the points of Φ in K. We define the following tree on Φ_K called the *(Euclidean) minimum spanning tree* (MST) of Φ_K: Connect any two points $X, Y \in \Phi_K$ by an edge if and only if there is no sequence $X_1 = X$, $X_2, \ldots, X_n = Y$, $X_i \in \Phi_K$, for some $n > 2$ such that $|X_{i+1} - X_i| < |X - Y|$ for all $i = 1, \ldots, n-1$. [6]

Denote by $M = M_K(\lambda)$ the *longest edge in the MST* of Φ_K.

Remark 15.2.1. The key relation between the MST and the Boolean model with germs in Φ_K and spherical grains of radius r is that the latter is fully connected iff $M_K \leq 2r$. The result is true for arbitrary point process Φ_K. Also, note that this in not equivalent to, but only implies the connectedness of the Boolean model construed on the whole process Φ and then truncated to the observation window. In this latter situation some grains of the points outside K may contribute to the connectedness of the model inside K.

The following result was proved in Penrose ((1997)) in $\mathbb{R}^2$ (and in higher dimension on the torus).

Proposition 15.2.2. *Given a unit square $K = [-\frac{1}{2}, \frac{1}{2}]^2 \subset \mathbb{R}^2$ and a homogeneous Poisson process Φ with intensity λ on the plane $\mathbb{R}^2$, denote by $M = M(\lambda)$ the longest edge of the MST of Φ_K. Then*

$$\lim_{\lambda \to \infty} \mathbf{P}\left\{ \lambda \pi M^2 - \log \lambda \leq u \right\} = \exp[-e^{-u}] \quad u \in \mathbb{R}\,. \tag{15.2.1}$$

Proof. The proof consists of two main steps. [7]

- Consider first the *nearest neighbor graph* (NNG) on Φ_K. This is a graph in which each nodes $X \in \Phi$ is connected to its nearest neighbour in Φ_K. Denote by $\tilde{M} = \tilde{M}(\lambda)$ the longest edges of the NNG of Φ_K. Ignoring the boundary effects, the random variables S_i denoting lengths of the edges of the NNG are distributed as a generic S

[6] It can be shown that the constructed graph is indeed a tree spanning all points of Φ_K and the total length of its edges is minimum among all tress spanning Φ_K. But this is not important for us.

[7] These steps are analogous to the steps of the proof of the asymptotic connectivity probability of the Erdős-Rényi graph with edge probability $p_n := \frac{\log n + c}{n}$ for some for some constant c, $-\infty < c < \infty$; cf the corresponding lesson.

with $\mathbf{P}\{ S > r \} = e^{-\lambda\pi r^2}$. (The typical node and Slivnyak's theorem should be used to state and prove this observation formally; cf Exercise 5.) Conditioning on the number of points of Φ_K being equal to n and ignoring the local dependence of S_i, for large n and λ

$$\begin{aligned}
\mathbf{P}\Big\{ \lambda\pi\tilde{M}^2 - \log\lambda \le u \Big\} &= \mathbf{P}\left\{ \tilde{M} \le \sqrt{\frac{u+\log\lambda}{\lambda\pi}} \right\} \\
&= \mathbf{P}\left\{ \max_{i=1,\dots,n} S_i \le \sqrt{\frac{u+\log\lambda}{\lambda\pi}} \right\} \\
\boxed{\text{ignoring dependence}} \quad &\approx \left(\mathbf{P}\left\{ S \le \sqrt{\frac{u+\log\lambda}{\lambda\pi}} \right\} \right)^n \\
&= (1 - e^{-(u+\log\lambda)})^n \\
\boxed{(1-1/a)^a \to e^{-1} \text{ for } a\to\infty} \quad &\approx \exp[-ne^{-(u+\log\lambda)}] \\
\boxed{\text{taking } n \approx \lambda} \quad &\approx \exp[-e^{-u}],
\end{aligned}$$

where the dependence can be formally treated using the Chain-Stein method and the last approximation is due to the fact that the distribution of the Poisson random variable concentrates around its mean λ, for large λ. More specifically, one can use a technique called *depoissonization*, which allows one to prove asymptotic results for Poisson random variable by replacing it with the constant equal to its mean; cf e.g. Jacquet and Szpankowski ((1998)).

- Note that the NNG is a sub-graph of the MST. Then $\tilde{M} \le M$ and hence

$$\begin{aligned}
\mathbf{P}\Big\{ \lambda\pi M^2 - \log\lambda \le u \Big\} &\le \mathbf{P}\Big\{ \lambda\pi\tilde{M}^2 - \log\lambda \le u \Big\} \\
&\le \mathbf{P}\Big\{ \lambda\pi M^2 - \log\lambda \le u \Big\} \\
&\quad + \mathbf{P}\{ \exists \text{ edge } u\text{-long in MST that is not in NNG} \},
\end{aligned}$$

where an edge is called u-long if its length l satisfies $\lambda\pi l^2 - \log\lambda > u$. Using some more involved arguments one can show that when $\lambda \to \infty$, all u-long edges appear as NNG edges with probability tending to one [8].

□

We can use Proposition 15.2.2 to approximated the local connectedness probability of the Boolean model with fixed spherical grains in the following way.

[8] In other words, it is hardly likely that the geometric Gilbert graph on Φ_K with the edge-length threshold defining the u-long edges, if disconnected, will have connected components of size larger than 2.

Corollary 15.2.3. *Let K be a square in $\mathbb{R}^2$ and consider the parametric family of Boolean models $\Xi_K(r)$ with germs in Φ_K and spherical grains of radius r on $\mathbb{R}^2$. Assume that $\lambda \to \infty$, $r \to 0$ so that*

$$\phi(\lambda, r) = 4\pi r^2 \lambda - \log \lambda - \log |K| \to u \quad \text{for some } u \in [-\infty, \infty]. \tag{15.2.2}$$

Then

$$\mathbf{P}\{\, \Xi_K(r) \text{ is connected} \,\} = \exp[-e^{-u}] + o(1). \tag{15.2.3}$$

In particular, when $\phi(\lambda, r) \to u = \infty$ then $\Xi_K(r)$ is connected with probability approaching 1.

Proof. By Remark 15.2.1 $\Xi_K(r)$ is connected iff the longest edge M_K of the minimum spanning tree of Φ_K is not longer than $2r$; $M_K \le 2r$. Scaling down the picture by the factor $\sqrt{|K|}$, by the translation invariance property of the homogeneous Poisson process and its scaling property, the distribution of M_K is equal to the distribution of $M'\sqrt{|K|}$, where M' is the longest edge of the MST of the homogeneous Poisson process of intensity $\lambda' := \lambda|K|$ observed in the unit window $[-1/2, 1/2]$. Hence

$$\begin{aligned}
\mathbf{P}\{\, \Xi_K(r) \text{ is connected} \,\} &= \mathbf{P}\{\, M_K \le 2r \,\} \\
&= \mathbf{P}\Big\{\, M'\sqrt{|K|} \le 2r \,\Big\} \\
&= \mathbf{P}\{\, \lambda' \pi M'^2 - \log \lambda' \le \lambda' \pi 4r^2/|K| - \log \lambda' \,\} \\
\boxed{\text{Proposition 15.2.2}} &= \exp[-e^{-u}] = o(1)
\end{aligned}$$

provided $\lambda' \pi 4r^2/|K| - \log \lambda' \to u$, which is equivalent to (15.2.2). □

15.2.2 Percolation of homogeneous Boolean model

We continue to consider homogeneous germ-grains Boolean models in $\mathbb{R}^d$ with spherical grains, but admit their random radii, distributed as the generic random variable R, for which we assume $\mathbf{E}\left[R^d\right] < \infty$. Under this condition the volume fraction of the Boolean model is smaller than one. Moreover, one can show that the entire Boolean model on $\mathbb{R}^d$ is disconnected with probability 1; cf Exercise 6.

Let us consider now the intensity of germs λ as the parameter of the family Ξ_λ of the considered Boolean models and define the *percolation function*

$$\Theta(\lambda) := \mathbf{P}\{\, \Xi_\lambda \text{ percolates} \,\}.$$

Using the results on thinning and superposition of Poisson processes one can see that Θ is an increasing function of λ. Let us define the *critical intensity* λ_c as the

$$\lambda_c := \inf\Big\{ \lambda \ge 0 : \Theta(\lambda) > 0 \Big\}. \tag{15.2.4}$$

We observe the following simple facts:

Remark 15.2.4.
- A priori λ_c can be equal to 0 or ∞.
- For $0 < \lambda < \lambda_c$ (if any) $\Theta(\lambda) = 0$.
- For $\lambda_c < \lambda < \infty$ (if any) $\Theta(\lambda) > 0$ and hence by Proposition 15.1.2 $\Theta(\lambda) = 1$.

Define also the *local percolation function* $\theta(\lambda)$ as the Palm probability that the connected component C_0 containing the typical grain $B_0(R_0)$ is unbounded (infinite)

$$\theta(\lambda) := \mathbf{P}^0\{C_0 \text{ is unbounded}\}.$$

Note that θ is also increasing function of λ. The following result says that λ_c is also critical for $\theta(\lambda)$; cf Exercise 7.

Proposition 15.2.5.

$$\lambda_c = \inf\Big\{\lambda \geq 0 : \theta(\lambda) > 0\Big\}. \tag{15.2.5}$$

Note that, $\theta(\lambda) = 0$ for $\lambda < \lambda_c$ but, unlike Θ, $\theta(\lambda) > 0$ does *not* imply $\theta(\lambda) = 1$.

One of the central questions of percolation theory concerns the non-degeneracy of λ_c, which here means $0 < \lambda_c < \infty$. One also calls this result the *not-trivial phase transition* in the continuum percolation.

Theorem 15.2.6. *Let λ_c be the critical intensity (15.2.4) of the family of the homogeneous Boolean models with spherical grains of random radius R.*

1. *If $d \geq 2$ and $\mathbf{P}R = 0 < 1$ (i.e. if R is not almost surely equal to 0), then $\lambda_c < \infty$.*
2. *If $\mathbf{E}\left[R^{2d-1}\right] < \infty$, then $\lambda_c > 0$.*

Proof. In order to prove the two parts of the result, one considers two previously studied models: a lattice percolation and a Galton-Watson process, to obtain appropriate bounds on the percolation probability. These are in fact some extensions the models studied in two first lessons on this lecture.

Part 1, using a 1-dependent bond percolation model. The proof of the fact that $\lambda_c < \infty$ is based on a discrete approximating of Ξ_λ by some *1-dependent percolation model* on the square lattice (cf Exercise 6 in the lesson on bond percolation).

Namely, consider some constants $\eta > 0$ and $p_0 > 0$, such that $\mathbf{P}\{\, R \geq \eta \,\} = p_0 > 0$; such positive constants exist under the assumption $\mathbf{P}\{\, R_0 = 0 \,\} < 1$. Consider a square lattice (in $\mathbb{R}^d$) with side of length $\Delta := \eta/(2d\sqrt{d})$. Note that this value is chosen so that any two balls of radius not less than η, centered at some points of any two adjacent sites of the lattice, are not disjoint. We declare a site of the lattice open if there is a germ of in with the grain of radius $R \geq \eta$. Otherwise, we declare the site closed. Note that the probability $p = p(\lambda)$ for a given site to be open is positive and tends to 1 when $\lambda \to \infty$. Moreover, the sites are declared open independently.

It can be shown that this *independent site percolation model* percolates for some $p < 1$. Indeed, the independent site percolation model can be presented as a 1-dependent bond percolation model where an edge connecting (say the centers) of the two adjacent sites is open iff the both sites are open. This one-dependent bond percolation model percolates for

the edge probability (which is equal to $p^2(\lambda)$) sufficiently close to 1; (cf Exercise 6 in the lesson on bond percolation).

Part 2, using a multi-type Galton-Watson tree. In order to prove that $\lambda_c > 0$, we use its equivalent expression given in Proposition 15.2.5. We consider the following *generations* of grains connected to $B_0(R_0)$. The first generation consists of all the grains directly connected to it. Given $n \geq 1$ generations, the $(n+1)$-st generation consists of all grains directly connected to some grain of the n-th generation and which do not intersect any grain of generation $1, \ldots, n-1$. We say that any grain $x_i + B(R_i)$ is of type k if $k-1 \leq R_i < k$ $(k = 1, 2 \ldots)$. Note that the number of grains of type k of the $(n+1)$ th generation, directly connected to a given grain of type i of the n th generation, but not totally contained in it, is not larger than the number of *all* grains of radius R, $k \leq R < k+1$ intersecting this given grain and not totally contained in it, which is in turn dominated by a Poisson random variable of parameter

$$\mu(i,k) = \lambda \nu_d \Big((i+k)^d - (i-k)_+^d \Big) \mathbf{P}\{\, k \leq R < k+1 \,\} \; .$$

The process of generations of grains connected to $B_0(R_0)$ is not a branching process due to the dependence between generations; however it can be stochastically bounded by a multi-type branching (Galton-Watson) process with a Poisson number of children of type k born to a parent of type i; this Poisson number has mean $\mu(i,k)$. It is not difficult to see that the *expected number of all individuals in all generations* of this branching process, given the root is of type i, is equal to $1 + \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} m_{ik}^n$, where m_{jk}^n is the jk th entry of the n th power of the matrix $\{m_{ik} = \mu(i,k)\}$. It is a matter of a direct calculation (see the details in ((Meester and Roy, 1996, proof of Theorem 3.3))) that the (unconditional) expectation of the total number of individuals is finite for sufficiently small $\lambda > 0$ provided $\mathbf{E}\left[R^{2d-1}\right] < \infty$. □

The exact value of the critical intensity λ_c does not admit any explicit expression even in the simplest case of fixed spherical grains. The following invariance property of the two percolation functions can be proved using the scaling properties of Poisson process; cf Exercise 9.

Fact 15.2.7. *Let the distribution of the generic radius R be given and consider a two-parameter family $\Xi_\lambda(r)$ of homogeneous Boolean models with intensity of germs λ and spherical grains of radius rR on $\mathbb{R}^d$. Denote by $\Theta(\lambda, r)$ and $\theta(\lambda, r)$ the respective percolation functions. Given the distribution of R, these functions depend only on λr^d:*

$$\Theta(\lambda, r) = \Theta(\lambda r^d, 1) = \quad \theta(\lambda, r) = \theta(\lambda r^d, 1) \, . \tag{15.2.6}$$

Consequently, the critical intensity $\lambda_c(r)$ for a given r satisfies

$$\lambda_c(r) = r^{-d} \lambda_c(1) = \frac{\lambda_c^*}{r^d \mathbf{E}\left[R\right]^d |B_0(1)|}, \tag{15.2.7}$$

where λ_c^* denotes the critical intensity for the Boolean model with random grains having unite expected d-dimensional volume $|B_0(1)| r^d \mathbf{E}\left[R\right]^d = 1$ in the family $\Xi_\lambda(r)$.

For the fixed radii (R constant), $\lambda_c^* \approx 1.1281$ for $d = 2$ and $\lambda_c^* \approx 0.3419$ for $d = 3$; cf ((Chiu et al., 2013, Section 3.3.4)) and the references cited therein.

Recall that the volume fraction of $\Xi_\lambda(r)$ is also a function of λr^d: $p_{\Xi_\lambda(r)} = 1 - e^{-\lambda r^d \mathbf{E}[R]^d |B_0(1)|}$ and the volume fractions p_c^* corresponding to the critical models in the family $\Xi_\lambda(r)$ satisfy

$$p_c^* = 1 - e^{-\lambda_c^*} .$$

Thus one can say that the spherical Boolean model with fixed grains is critical when it has the following volume fraction: $p_c^* \approx 0.6763$ for $d = 2$ and $p_c^* \approx 0.2896$ for $d = 3$.

Remark 15.2.8. Phase transition has been studied for various extension of the fundamentally Boolean model. Some models exhibit more than one phase transition. For example the shot-noise coverage model generated by Poisson process, mentioned in the lesson on coverage processes, percolates for some moderate values of λ and does not percolate when λ is too small or too large, cf Dousse et al. ((2006)).

Extensions to non-Poisson models have also been considered. In general, when a Poisson process is replaced be some more regular processes (exhibiting less clustering of points) than the corresponding model is expected to exhibit similar phase transition Błaszczyszyn and Yogeshwaran ((2014)).

15.3 Exercises

1. Consider the following two subsets of the plane $\mathbb{R}^2$:
   - Let $F_1 = \{(x, y) \in \mathbb{R}^2 : y = \sin(1/x), x \neq 0\} \sup\{(0, y) : y \in [-1, 1]\}$. Prove that F_1 is a closed and connected set.
   - Let $F_2 = \{(x, y) \in \mathbb{R}^2 : y = 1/|x|, x \neq 0\}$. Prove that F_2 is a closed and disconnected set.
   - $F_2 \cup \{\infty\}$ is closed and connected in the one-point (Alexandroff) compactifiation $\mathbb{R}^2 \cup \{\infty\}$ of $\mathbb{R}^2$ where the extra point ∞ is considered to be the limit of all sequences x_n on $\mathbb{R}^2$ such that $|x_n| \to \infty$.

   The above two examples show that the notion of connectedness of unbounded sets needs to be treated with some care.

2. *The Cantor set.* The Cantor ternary set $\mathcal{C}$ is a subset of the interval $[0, 1]$ created by iteratively deleting the open middle thirds from the remaining set of line segments starting from $[0, 1]$; $\mathcal{C} := \bigcap_{n=1}^{\infty} C_n$, where $C_0 = [0, 1]$ and

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3} \right) \text{ for } n \geq 1.$$

   Prove that $\mathcal{C}$ is a closed subset of $[0, 1]$, which is *totally disconnected* i.e., all its elements are different connected components. Hence $\mathcal{C}$ has an uncountable number of connected components.

3. Prove that the independent bond percolation model on the square lattice in $\mathbb{R}^d$ considered in the first lesson is a disconnected set with probability one.

4. Prove the relation between the MST and the Boolean model formulated in Remark 15.2.1.

5. Using Slinvyak's theorem show that the distance S of the typical point of a homogeneous Poisson point process on intensity λ on $\mathbb{R}^d$ to its nearest neighbor (which is unique!) has the distribution $\mathbf{P}\{ S \leq r \} = 1 - e^{-\lambda |B_0(1)| r^d}$.

6. Consider a homogeneous Boolean germ-grain model Ξ with on $\mathbb{R}^d$ with intensity λ of germs $\Phi = \sum_i \delta_{X_i}$ and random spherical grains R_i of generic radius R such that $\mathbb{E}[R^d] < \infty$. Prove the following statements:

   (a) The probability that a typical grain is isolated is equal to

$$\mathbf{P}^0 \left\{ B_0(R_0) \cap \bigcup_{\Phi \ni X_i \neq 0} \Big(X_i + B_0(R_i) \Big) = \emptyset \right\} = \mathbf{E} \left[e^{-\lambda \nu_d \sum_{k=0}^{d} \binom{d}{k} R^{d-k} \mathbb{E}[R'^k]} \right], \tag{15.3.1}$$

where R' is independent of R ans has the same distribution. Hit: Use Slivnyak's theorem. Given the radius R_0 of the typical grain (which is independent of the all other germs and their grains) represent the desired probability in terms of the capacity functional of the union of all other grains.

(b) The number of isolated grains of Ξ is infinite with probability 1. Hint: Use the Campbell-Mecke-Matthes theorem to show that the number of isolated grains N satisfies

$$\mathbf{E}\,[N] = \lambda \int_{\mathbb{R}^d} \mathbf{P}^0\{\, B_0(R_0) \text{ is isolated set}\,\}\,\mathrm{d}x = \infty\,.$$

Conclude using the ergodicity of the Boolean germ-grain process.

7. Prove Proposition 15.2.5. Hint: Denote by I_i the indicator of the event that the grain $X_i + B_{X_i}(R_i)$ belongs to an unbounded component. Use the Campbell-Mecke-Matthes theorem.

8. Prove for a one-dimensional homogeneous Boolean model that $\mathbf{E}\,[R] < \infty$ implies $\lambda_c = \infty$ and $\mathbf{E}\,[R] = \infty$ implies $\lambda_c = 0$. (Strictly speaking in this latter case it is no longer a Boolean model, for which $\mathbf{E}\,[R] < \infty$ is required).

9. Prove Fact 15.2.7.

10. *Computer exercise.* Estimate the normalized critical intensity λ_c^* for planar Boolean model with fixed spherical grains. (Recall, $\lambda_c^* \approx 1.1281$). Use *R* with *spatstat* and *igraph* and *SGCS*.

(a) Simulate Poisson process with unit intensity in a reasonably large window. Construct the matrix of distances between the simulated points. Use the command `graph.adjacency` from *igraph* to define the corresponding Gilbert graph. Use the command `clusters` to obtain the connected components of the graph. For a given realization of germs, plot the fraction of nodes in the largest and second largest cluster in function of the grain radius. Estimate the critical radius as the radius when second largest component attains its maximal value. It approximately corresponds to the inflection point of the curve of the largest component. Use the relation (15.2.7) to derive λ_c^*.

(b) Try an alternative approach, based on the estimation of the Euler characteristic, cf Section 15.1.1. Use `morphoEuler` function from the package *SGCS*.

Bibliography

D. Aldous and R. Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12(54):1454–1508, 2007.

D. Aldous and J. M. Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

K. B. Athreya and P. Ney. *Branching processes*. Dover Publications, 2004. ISBN 9780486434742. Originally published: Springer-Verlag, Berlin 1972.

F. Baccelli and B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks, Volume I — Theory*, volume 3, No 3–4 of *Foundations and Trends in Networking*. NoW Publishers, 2009.

F. Baccelli, B. Błaszczyszyn, and M. Karry. *Random Measures, Point Processes and Stochastic Geometry*. 2017. working project available as the lecture supplementary material.

A. Baddeley, R. Turner, and E. Rubak. Datasets provided for spatstat. https://cran.r-project.org/web/packages/spatstat/vignettes/datasets.pdf.

I. Benjamini. *Coarse geometry and randomness–École d'Été de Probabilités de Saint-Flour 2011*, volume 2100 of *Lecture Notes in Mathematics*. Springer, 2013.

N. Berglund. La probabilité d'extinction d'une espèce menacée. http://images.math.cnrs.fr/La-probabilite-d-extinction-d-une.html, 2013.

B. Błaszczyszyn and D. Yogeshwaran. Clustering comparison of point processes with applications to random geometric models. In V. Schmidt, editor, *Stochastic Geometry, Spatial Statistics and Random Fields: Models and Algorithms*, volume 2120 of *Lecture Notes in Mathematics*, chapter 2, pages 31–71. Springer, 2014.

B. Błaszczyszyn, M. Haenggi, P. Keeler, and S. Mukherjee. *Stochastic Geometry Analysis of Cellular Networks*. Cambridge University Press, 2018.

O. Bobrowski and M. Kahle. Topology of random geometric complexes: a survey. arXiv:1409.4734, 2014.

B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.

S. R. Broadbent and J. M. Hammersley. Percolation processes: I. crystals and mazes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 53(3):629–641, 1957.

S. N. Chiu, D. Stoyan, W. Kendall, and J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 3 edition, 2013.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, vol. I*. Probability and Its Applications. Springer New York, New York, NY, 2003. doi: 10.1007/978-0-387-49835-5. URL http://link.springer.com/10.1007/978-0-387-49835-5.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

O. Dousse, M. Franceschetti, N. Macris, R. Meester, and P. Thiran. Percolation in the signal to interference ratio graph. *Journal of Applied Probability*, 43(2):552–562, 2006.

M. Draief and L. Massoulié. *Epidemics and Rumours in Complex Networks*. Cambridge University Press, Jan. 2010. ISBN 0521734436, 9780521734431. URL http://dl.acm.org/citation.cfm?id=1824103.

H. Duminil-Copin and V. Tassion. A new proof of the sharpness of the phase transition for bernoulli percolation on $\mathbb{Z}^d$. *arXiv preprint arXiv:1502.03051*, 2015.

P. Erds and A. Rényi. On random graphs I. *Publicationes Mathematicae*, 6:290–297, 1959.

B. Forghani and K. Mallahi-Karai. Amenability of trees. *Groups, Graphs and Random Walks*, 436:176, 2017.

M. Franceschetti and R. Meester. *Random networks for communication: from statistical physics to information systems*, volume 24. Cambridge University Press, 2008.

E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

G. Grimmett. Random labelled trees and their branching networks. *Journal of the Australian Mathematical Society*, 30(2):229–237, 1980.

G. Grimmett. *Percolation*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2013. ISBN 9783662039816. URL https://books.google.fr/books?id=Vnz1CAAAQBAJ.

P. Hall. *Introduction to the Theory of Coverage Processes*. Wiley & Sons, New York, 1988.

H. Hermann and A. Elsner. Geometric models for isotropic random porous media: a review. *Advances in Materials Science and Engineering*, 2014.

M. Heveling, G. Last, et al. Characterization of palm measures via bijective point-shifts. *The Annals of Probability*, 33(5):1698–1715, 2005.

C. Hirsch and G. Last. On maximal hard-core thinnings of stationary particle processes. arXiv:1707.00494, 2017.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

P. Jacquet and W. Szpankowski. Analytical depoissonization and its applications. *Theoretical Computer Science*, 201(1):1 – 62, 1998. ISSN 0304-3975. doi: https://doi.org/10.1016/S0304-3975(97)00167-9. URL http://www.sciencedirect.com/science/article/pii/S0304397597001679.

S. Janson. Random coverings in several dimensions. *Acta Math.*, 156:83–118, 1986.

S. Janson and M. J. Luczak. A new approach to the giant component problem. *Random Structures and Algorithms*, 34(2):197–216, Mar. 2009. ISSN 10429832. doi: 10.1002/rsa.20231. URL http://doi.wiley.com/10.1002/rsa.20231.

O. Kallenberg. *Random measures*. Academic Pr, 1983.

O. Kallenberg. *Foundations of modern probability*. springer, 2002.

J. L. Kelley. *General Topology*. Springer-Verlag, New York, NY, USA, 1955.

M. Kerscher, K. Mecke, P. Schücker, H. Bohringer, L. Guzzo, C. A. Collins, S. Schindler, S. D. Grandi, and R. Cruddace. Non-gaussian morphology on large scales: Minkowski functionals of the reflex cluster catalogue. *Astronomy & Astrophysics*, 377(1):1–16, Oct. 2001. doi: 10.1051/0004-6361:20011063. URL http://dx.doi.org/10.1051/0004-6361:20011063.

G. Last and M. Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, oct 2017. ISBN 9781107088016. doi: 10.1017/9781316104477. URL https://www.cambridge.org/core/product/identifier/9781316104477/type/book.

Laszlo Gadar. Generate (random) graphs with igraph. https://rpubs.com/lgadar/generate-graphs. Accessed: 2017-10-16.

J.-Y. Le Boudec. Understanding the simulation of mobility models with palm calculus. *Performance Evaluation*, 64(2):126 – 147, 2007. ISSN 0166-5316. doi: https://doi.org/10.1016/j.peva.2006.03.001. URL http://www.sciencedirect.com/science/article/pii/S0166531606000150.

G. Matheron. *Random sets and integral geometry*. John Wiley & Sons, London, 1975.

J. Mecke. Stationäre zufällige maße auf lokalkompakten abelschen gruppen. *Probability Theory and Related Fields*, 9(1):36–58, 1967.

K. R. Mecke. Morphology of spatial patterns-porous media, spinodal decomposition and dissipative structures. *Acta Physica Polonica. Series B*, 28(8):1747–1782, 1997.

R. Meester and R. Roy. *Continuum percolation*, volume 119. Cambridge University Press, 1996.

I. Molchanov. *Theory of random sets*. Springer Science & Business Media, 2005.

M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing*, 7(3):295–305, 1998.

R. A. Neher, K. Mecke, and H. Wagner. Topological estimation of percolation thresholds. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(01):P01011, 2008. doi: 10.1088/1742-5468/2008/01/P01011.

B. L. Okun. Euler characteristic in percolation theory. *Journal of Statistical Physics*, 59(1): 523–527, apr 1990. ISSN 1572-9613. doi: 10.1007/BF01015581. URL https://doi.org/10.1007/BF01015581.

T. S. Package. Spatstat quick reference guide. http://spatstat.org/resources/spatstatQuickref.pdf, 2017.

M. Penrose. The longest edge of the random minimal spanning tree. *The Annals of Applied Probability*, 7(2):340–361, 1997.

E. Roubin and J.-B. Colliat. Critical probability of percolation over bounded region in N-dimensional Euclidean space. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(3):033306, 2015. ISSN 1742-5468. doi: 10.1088/1742-5468/2016/03/033306. URL http://stacks.iop.org/1742-5468/2016/i=3/a=033306?key=crossref.044bcc75d116e3206b48d6edafee14a7http://arxiv.org/abs/1510.08252http://dx.doi.org/10.1088/1742-5468/2016/03/033306.

Y. A. Rozanov. *Markov Random Fields*. Springer New York, New York, NY, 1982. ISBN 978-1-4613-8192-1. doi: 10.1007/978-1-4613-8190-7. URL http://www.springerlink.com/index/10.1007/978-1-4613-8190-7.

R. Schneider and W. Weil. *Stochastic and integral geometry*. Spronger-Verlag, 2008.

C. Song, P. Wang, and H. A. Makse. A phase diagram for jammed matter. *Nature*, 453(7195): 629–632, 2008. arXiv:0808.2196.

D. Stauffer and A. Aharony. *Introduction To Percolation Theory*. Taylor & Francis, 2003. ISBN 9781135747831. URL https://books.google.fr/books?id=Dph5AgAAQBAJ.

J. M. Stoyanov. *Counterexamples in probability*. Courier Corporation, 2013.

S. Torquato, O. Uche, and F. Stillinger. Random sequential addition of hard spheres in high euclidean dimensions. *Physical Review E*, 74(6):061308, 2006.

B. Tsirelson. Measurability and Continuity,, 2013. URL http://www.tau.ac.il/{~}tsirel/Courses/MeasContin/main.html. Lecture at the School of Mathematical Sciences, Tel Aviv University, available at http://www.tau.ac.il/{~}tsirel/Courses/MeasContin/main.html.

R. Turner. Package deldir. https://cran.r-project.org/web/packages/deldir/deldir.pdf, 2017.

R. Van Der Hofstad. *Random graphs and complex networks:: Volume 2*. 2014. In preparation, preliminary version available on the author's web page http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf.

R. Van Der Hofstad. *Random graphs and complex networks:: Volume 1*, volume 43 of *Cambridge Series in Statistical and probabilistic Mathematics*. Cambridge university press, 2017. Available on the author's web page http://www.win.tue.nl/~rhofstad/.

H. W. Watson and F. Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.