

Poisson Point Measures & Stochastic Geometry

Nicolas CURIEN

October 5, 2026

Contents

I	Poisson Point Measures	5
1	Construction of PPMs	6
1.1	Review of Poisson random variables	6
1.1.1	Definition and Laplace transform	6
1.1.2	Infinite divisibility	7
1.2	Poisson point measures	9
1.2.1	Point measures	9
1.2.2	Construction of PPM	11
1.3	Basic operations	14
1.4	Recovering the points	15
2	Useful formulas	19
2.1	Exponential formula	19
2.2	The Mecke formula	22
2.2.1	Multi-point Mecke	25
2.3	Exercises	25
3	Random measures on Polish spaces	27
3.1	Radon measures	27
3.1.1	Weak and vague convergence	27
3.1.2	Random measures	28
3.2	Point measures	30
3.2.1	Convergence of atoms	30
3.2.2	An example	31

II	Poissonization	33
4	Temporal Poisson processes	34
4.1	Exponential representation	34
4.2	Direct applications	37
4.2.1	Markov properties	37
4.2.2	Choosing with clocks	38
5	Inhomogeneous PPMs	40
5.1	Inhomogeneous Poisson point measures	40
5.1.1	The Gumbel cloud	40
5.1.2	The polynomial cloud	41
5.1.3	The birthday process	43
5.2	Hawkes process	45
6	Poissonization	48
6.1	An example and a few tools	48
6.1.1	Order statistics	48
6.1.2	Two de-Poissonization tools	50
6.2	Examples without branching	50
6.2.1	A matchbox warm-up	51
6.2.2	Coupon collector	52
6.2.3	The coupon collector's earworm	54
6.2.4	Balls in bins	55
6.2.5	Pill problem	56
6.2.6	O.K. Corral	57
6.2.7	Kingman's coalescent: splitting at the root	59
6.3	Branching: Yule and binary search trees	61
6.3.1	From the discrete binary search tree to the Yule tree	61
6.3.2	Pólya urn and splitting at the root	64
III	Stochastic Geometry	66
7	Models based on points	67
7.1	Basic properties and intensities	68
7.1.1	Intensities	69
7.2	Neighbor models	70
7.2.1	Boolean percolation	70

7.2.2	The k -nearest-neighbor graph	73
7.2.3	The Gabriel graph	77
7.3	Poisson–Voronoi tessellations	78
7.3.1	Definition and elementary geometry	78
7.3.2	The Delaunay dual	80
7.3.3	Stationarity and scaling	80
7.3.4	The typical Poisson–Voronoi cell	81
7.3.5	The zero cell	82
7.3.6	Planar mean values	83
7.3.7	Power and Laguerre tessellations	84
7.4	Random parking from a Poisson rain	85
8	Models based on lines	88
8.1	The kinematic measure	88
8.1.1	Parametrizing the set of lines	88
8.1.2	The kinematic measure	90
8.2	Cauchy–Crofton formulas	92
8.2.1	The curve formula	92
8.2.2	Cauchy’s perimeter formula	93
8.2.3	Integration of chord lengths	94
8.3	Poisson line tessellation	94
8.3.1	General position and the induced tessellation	95
8.3.2	Mean geometric characteristics	98
8.3.3	The zero cell	100
8.3.4	Extension to higher dimensions	101
8.4	Bibliographical remarks	101

Introduction

Since Kolmogorov's axiomatization, probability theory can, in principle, be regarded as a branch of analysis: it amounts to studying probability measures, while random variables are merely measurable functions. Yet every probabilist knows that probabilistic intuition becomes much more powerful once we stop thinking of a random variable as a deterministic function and instead regard it as a genuinely random object. Structural properties of a probability measure μ often reveal themselves through the random behavior of a variable X with law μ , and the probabilistic language makes it possible to uncover and express deeper properties. As a simple illustration, try writing the law of large numbers purely in analytic language!

When μ is an infinite measure, however, this correspondence with random variables seems to break down: there is no way to normalize μ and sample a single point according to it. Yet infinite measures are often more natural than finite ones, precisely because they enjoy stronger symmetry, invariance, or scaling properties. Think, for instance, of Lebesgue measure on $\mathbb{R}^d$. Poisson random measures provide a beautiful way to restore a probabilistic interpretation of such measures and, in a sense, to sample from an infinite measure μ . The price to pay is that one no longer samples a single point, but rather a random collection—a cloud—of points. Developed in its modern general form during the 1950s and 1960s, this remarkable object now permeates much of modern probability, from stochastic geometry and stochastic processes to stochastic analysis and statistical physics.

In these lecture notes, intended as an advanced course in master, we introduce the basic theory of Poisson random measures, their construction, and some of their structural properties, with particular emphasis on time-layered spaces. This provides the foundations for the Poissonization trick, a powerful method that yields quick and elegant proofs of discrete probability results by embedding discrete models into continuous time. In the second part, we turn to stochastic geometry and present several beautiful constructions based on Poisson processes of points and lines in $\mathbb{R}^d$.

Part I:

Poisson Point Measures

Given a measure μ on a space E , one would usually like to sample a *random element* $X \in E$ according to μ . This is obviously possible if μ is a probability measure or, more generally, if μ has finite mass, so that we can consider the normalized probability measure $\mu(\cdot)/\mu(E)$. This seems hopeless when μ is genuinely infinite, as is, for example, Lebesgue measure on Euclidean spaces. The concept of a Poisson random cloud provides a way to sample not just a single point $X \in E$ but also a random collection of points $X_i \in E$ for $i \in I$ in a uniform manner even in the presence of an infinite measure. This beautiful object, called a **Poisson point measure**, is a central tool in modern probability theory. Its construction relies on the infinite divisibility properties of the Poisson law, which we first recall.



Figure 1: An AI-generated image of a Poisson point measure

I: Construction of PPMs

Remark 1.1 (Terminology). In the literature, Poisson point measures are also called Poisson point processes, Poisson random measures, or Poisson point clouds. The most common term is perhaps “Poisson point process”, but the word *process* can be misleading since it may suggest a time parametrization (which shall appear in Chapter 4). We shall mostly use *Poisson point measure* (PPM), which emphasizes the measure-valued nature of the object.

1.1 Review of Poisson random variables

In this section we recall the basic algebra of Poisson¹ random variables.

1.1.1 Definition and Laplace transform

Definition 1.1. Let $\lambda \geq 0$. A random variable N , with values in $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$, is said to have the Poisson distribution with parameter λ if

$$\mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq 0.$$

We write $N \sim \text{Poi}(\lambda)$.

When $\lambda = 0$, this means that $N = 0$ almost surely. We also use the convention $\text{Poi}(\infty) = \delta_{+\infty}$. The **Laplace transform** is given by straightforward computation: for $N \sim \text{Poi}(\lambda)$ and $t \geq 0$

$$\begin{aligned} \mathbb{E}[e^{-tN}] &= e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k e^{-tk}}{k!} \\ &= e^{-\lambda + \lambda e^{-t}} = \exp(-\lambda(1 - e^{-t})). \end{aligned} \tag{1.1}$$



¹ Siméon-Denis Poisson (1781–1840), French mathematician. The Poisson law appeared in his 1837 book *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, in connection with a limiting regime for binomial probabilities.

Differentiating this identity at $t = 0$ gives the familiar formulas $\mathbb{E}[N] = \lambda$ and $\text{Var}(N) = \lambda$. More generally, for every integer $r \geq 1$, the factorial moments (see Exercise 1.1) of the Poisson distribution are given by

$$\mathbb{E}[N(N-1) \cdots (N-r+1)] = \lambda^r.$$

It is easy to see that the Laplace transform $t \mapsto \mathbb{E}[e^{-tN}]$ characterizes the law of random variables N with values in $\{0, 1, 2, \dots\} \cup \{\infty\}$, where we use the convention that $0 \cdot \infty = 0$. In connection with this convention, let us collect here a lemma regarding Laplace transform and possibly infinite random variables:

Lemma 1.1. *If $X \in [0, \infty]$ a non-negative possibly infinite random variable and $\phi(t) = \mathbb{E}[e^{-tX}]$ for $t \geq 0$ is its Laplace transform. Then we have*

$$\mathbb{P}(X < \infty) = \lim_{t \downarrow 0} \phi(t).$$

Proof. For $x \in [0, \infty]$ we have $e^{-tx} \rightarrow \mathbf{1}_{x < \infty}$ as $t \downarrow 0$ and so the statement follows by dominated convergence theorem. $\square$

1.1.2 Infinite divisibility

Proposition 1.2 (Stability under addition). *Let $N_1, N_2, \dots$ be independent Poisson random variables with respective means $\lambda_1, \lambda_2, \dots \in [0, \infty]$. Then*

$$N = \sum_{i=1}^{\infty} N_i \sim \text{Poi}\left(\sum_{i=1}^{\infty} \lambda_i\right).$$

Proof. Let $\Lambda = \sum \lambda_i \in [0, \infty]$. We have the monotone almost sure convergence $\sum_{i=1}^n N_i \rightarrow N$ and so by the dominated convergence theorem we have for $s = e^{-t} \in [0, 1]$

$$\mathbb{E}[s^N] = \lim_{n \rightarrow \infty} \mathbb{E}[s^{N_1 + \dots + N_n}] = \lim_{n \rightarrow \infty} \prod_{i=1}^n \mathbb{E}[s^{N_i}] = \exp(-\Lambda(1-s)).$$

The right-hand side is the Laplace transform (or probability-generating function) of a Poisson variable with mean Λ . Since the Laplace transform characterizes the law of variables with values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ we are done. $\square$

We also have the reverse operation, which allows us to split a Poisson random variable.

Proposition 1.3 (Splitting property). *Fix $\lambda \in [0, \infty)$. Let $N \sim \text{Poi}(\lambda)$ and let $p_1, p_2, \dots \geq 0$ satisfy $\sum_{i \geq 1} p_i = 1$. Conditionally on N , assign each of the N objects independently to category i with probability p_i . Let N_i be the number assigned to category i . Then the variables $N_1, N_2, \dots$ are independent and*

$$N_i \sim \text{Poi}(\lambda p_i).$$

Proof. It is enough to prove the result for two categories, assigned with probabilities p and $1 - p$. The general case follows by iteration: for every $n \geq 1$, the $n + 1$ variables

$$N_1, \dots, N_n, \sum_{j>n} N_j$$

are independent, with respective Poisson parameters $\lambda p_1, \dots, \lambda p_n, \lambda \sum_{j>n} p_j$. This determines all finite-dimensional distributions of $(N_i)_{i \geq 1}$.

For two categories, conditionally on N the vector (N_1, N_2) is binomially split, hence

$$\begin{aligned} \mathbb{P}(N_1 = k_1, N_2 = k_2) &= \mathbb{P}(N = k_1 + k_2) \binom{k_1 + k_2}{k_1} p^{k_1} (1 - p)^{k_2} \\ &= e^{-\lambda} \frac{\lambda^{k_1 + k_2}}{(k_1 + k_2)!} \frac{(k_1 + k_2)!}{k_1! k_2!} p^{k_1} (1 - p)^{k_2} \\ &= e^{-\lambda p} \frac{(\lambda p)^{k_1}}{k_1!} \cdot e^{-\lambda(1-p)} \frac{(\lambda(1-p))^{k_2}}{k_2!}. \end{aligned}$$

Thus N_1 and N_2 are independent Poisson variables with respective means λp and $\lambda(1-p)$. $\square$

Exercise 1.1 (Factorial moments and the Poisson law). Let N be a non-negative integer-valued random variable. For $k \geq 0$, provided that $\mathbb{E}[N^k] < \infty$, the k th factorial moment of N is defined by

$$\mathbb{E}[(N)_k] = \mathbb{E}[N(N-1) \cdots (N-k+1)],$$

with the convention $(N)_0 = 1$.

1. Let $N \sim \text{Poi}(\lambda)$, with $\lambda > 0$. Compute the factorial moments of N and show that the Poisson distribution is determined by its factorial moments.

Hint: Prove that whenever the series is absolutely convergent we have

$$\mathbb{P}(N = r) = \frac{1}{r!} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \mathbb{E}[(N)_{r+j}],$$

2. Suppose that

$$N = \sum_{i \in I} \mathbf{1}_{A_i},$$

where I is finite or countable, $N < \infty$ almost surely, and $\mathbb{E}[N^k] < \infty$. Show that

$$\mathbb{E}[(N)_k] = \sum_{\substack{i_1, \dots, i_k \in I \\ \text{pairwise distinct}}} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}).$$

Hint: Prove that

$$\left(\sum_{i \in I} y_i\right)_k = \sum_{\substack{i_1, \dots, i_k \in I \\ \text{pairwise distinct}}} y_{i_1} \cdots y_{i_k}, \quad y_i \in \{0, 1\}.$$

3. Use the previous questions and the method of moments to prove the following:

(a) For every $\lambda > 0$,

$$\text{Bin}\left(n, \frac{\lambda}{n}\right) \xrightarrow[n \rightarrow \infty]{\text{law}} \text{Poi}(\lambda).$$

(b) Let σ_n be uniformly distributed on the symmetric group $\mathfrak{S}_n$. Then the number of fixed points of σ_n converges in distribution to $\text{Poi}(1)$ as $n \rightarrow \infty$.

(c) (*) Consider the complete graph on the vertex set $\{1, \dots, n\}$. Thus every pair of distinct vertices is joined by an edge, and the graph has $n(n-1)/2$ edges. For $p \in [0, 1]$, let $G(n, p)$ be the random graph obtained by retaining each edge independently with probability p . Let $T(n, p)$ be the number of triangles in $G(n, p)$. Show that, for every $\lambda > 0$,

$$T\left(n, \frac{\lambda}{n}\right) \xrightarrow[n \rightarrow \infty]{\text{law}} \text{Poi}\left(\frac{\lambda^3}{6}\right).$$

1.2 Poisson point measures

Let $(E, \mathcal{E})$ be a measurable space; no topology is needed. A natural way to describe a countable collection of points in E would be to index the points as $X_1, X_2, \dots$. Although this concrete approach works in many situations, it may be cumbersome on an abstract space, and repetitions must be allowed. A more symmetric point of view is to encode the collection as a counting measure.

1.2.1 Point measures

For $x \in E$, the Dirac measure at x is denoted by δ_x :

$$\delta_x(A) = \mathbf{1}_A(x), \quad A \in \mathcal{E}.$$

Definition 1.2. A point measure (or counting measure) on E is a measure of the form

$$\nu = \sum_{i \in I} \delta_{x_i},$$

where I is finite or countable and $x_i \in E$. Repetitions among the x_i are allowed. The zero measure corresponds to $I = \emptyset$. We denote the set of point measures on E by $\mathcal{N}(E)$.

The same point may occur several times. For example, $3\delta_x + \delta_y$ assigns mass 3 to x and mass 1 to y . In an abstract measurable space it is preferable to allow multiplicities from the start.

We endow $\mathcal{N}(E)$ with the smallest σ -field $\mathfrak{N}$ making all evaluation maps

$$\nu \mapsto \nu(A), \quad A \in \mathcal{E},$$

measurable, where the range is $\mathbb{Z}_{\geq 0} \cup \{\infty\}$. A *random point measure* is then a measurable map

$$\mathbf{N} : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathcal{N}(E), \mathfrak{N}). \quad (1.2)$$

Equivalently, $\mathbf{N}$ is a random point measure if it is a point measure and $\mathbf{N}(A)$ is a random variable for every $A \in \mathcal{E}$. Since $\mathfrak{N}$ is generated by the evaluation maps, the law of $\mathbf{N}$ is determined by the finite-dimensional laws of the family $(\mathbf{N}(A))_{A \in \mathcal{E}}$. The **mean intensity** of a random point measure is the measure (check it as an exercise) on $(E, \mathcal{E})$ defined by

$$\mu(A) = \mathbb{E}[\mathbf{N}(A)], \quad A \in \mathcal{E}.$$

Definition 1.3. A Poisson point measure on $(E, \mathcal{E})$ with intensity μ is a random point measure Π such that

(i) for every $A \in \mathcal{E}$,

$$\Pi(A) \sim \text{Poi}(\mu(A));$$

(ii) whenever $A_1, \dots, A_r \in \mathcal{E}$ are pairwise disjoint, the random variables $\Pi(A_1), \dots, \Pi(A_r)$ are independent.

We write $\Pi \sim \text{PPM}(E, \mathcal{E}, \mu)$, or simply $\text{PPM}(\mu)$ when the underlying space is clear.

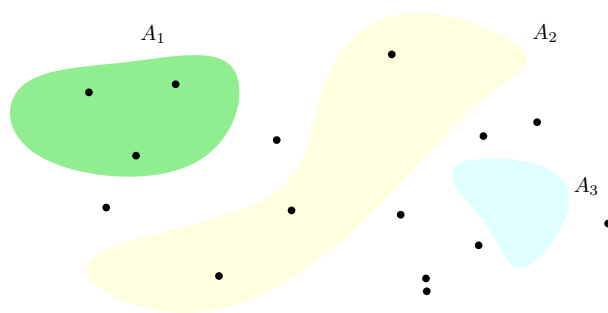


Figure 1.1: Illustration of the definition of a PPM: the numbers of atoms falling in disjoint regions are independent Poisson random variables.

A small point deserves to be checked: properties (i)–(ii) indeed determine the law among random point measures. For arbitrary $A_1, \dots, A_r \in \mathcal{E}$, let $B_1, \dots, B_m$ be the non-empty cells of the finite measurable partition generated by $A_1, \dots, A_r$. Each $\mathbf{N}(A_i)$ is a sum of some of the variables $\mathbf{N}(B_j)$. Since the B_j are pairwise disjoint, Definition 1.3 prescribes the joint law of $(\mathbf{N}(B_1), \dots, \mathbf{N}(B_m))$, and hence that of $(\mathbf{N}(A_1), \dots, \mathbf{N}(A_r))$. Thus all finite-dimensional laws of the evaluation family are determined.

Example 1.1. If μ has an atom at $x \in E$ then $\Pi(\{x\})$ is a non-trivial Poisson random variable, that is we may have repetitions in the atoms of a PPM.

Remark 1.2 (A subtlety). Let us stress a subtle point in Definition 1.3. A PPM is first of all a *random point measure* Π ; its evaluations $(\Pi(A) : A \in \mathcal{E})$ are then required to have the Poisson finite-dimensional laws. A family of random variables $(\Pi_A : A \in \mathcal{E})$ with the same finite-dimensional laws does not by itself produce a random measure. Indeed, on a general measurable space, Kolmogorov’s extension theorem ensures the existence of a family $(\Pi_A : A \in \mathcal{E})$ with the prescribed consistent finite-dimensional distributions. For each fixed disjoint sequence $A_1, A_2, \dots$, one then has

$$\sum_{i \geq 1} \Pi_{A_i} = \Pi_{\bigsqcup_{i \geq 1} A_i} \quad \text{almost surely,}$$

but the exceptional null set can depend on the chosen sequence. This does not yield, on one common event of probability one, a countably additive map $A \mapsto \Pi_A$ simultaneously for all sequences. The point-measure requirement in the definition therefore carries a true content.

We can now state the main existence result of this chapter.

Theorem 1.4 (Existence of PPMs)

If μ is a σ -finite measure on a measurable space $(E, \mathcal{E})$, then there exists a Poisson point measure on E with intensity μ .

Example 1.2. A basic example is a Poisson point measure on $\mathbb{R}^d$ with Lebesgue intensity.

Remark 1.3 (History). Although the Poisson distribution dates back to the nineteenth century, Poisson random measures became a systematic tool in probability in the mid-twentieth century. They play a central role in the theory of Lévy processes and, through the work of Itô, in modern stochastic calculus.² They are also fundamental objects in stochastic geometry.

1.2.2 Construction of PPM

We prove Theorem 1.4 in two steps.

 ² Paul Lévy (1886–1971), French mathematician.
mathematician.



Kiyoshi Itô (1915–2008), Japanese

Finite intensity. Assume first that $\mu(E) < \infty$. If $\mu(E) = 0$, simply take $\Pi = 0$. Otherwise, let $(X_i)_{i \geq 1}$ be i.i.d. E -valued random variables³ with law $\mu/\mu(E)$, and let $N \sim \text{Poi}(\mu(E))$ be independent of the sequence $(X_i)_{i \geq 1}$.

Define

$$\Pi = \sum_{i=1}^N \delta_{X_i}.$$

This is a random point measure, since the measurability of the X_i 's and N ensure that events of the form $\{\Pi(A) = k\}$ for $A \in \mathcal{E}$ and $k \in \mathbb{Z}_{\geq 0}$ are measurable. If $A_1, \dots, A_r$ are pairwise disjoint measurable sets, add $A_0 = E \setminus \bigcup_{j=1}^r A_j$ and apply the splitting property (Proposition 1.3) to the measurable partition $A_0, A_1, \dots, A_r$. We obtain that $\Pi(A_1), \dots, \Pi(A_r)$ are independent and

$$\Pi(A_i) \sim \text{Poi}\left(\mu(E) \frac{\mu(A_i)}{\mu(E)}\right) = \text{Poi}(\mu(A_i)).$$

Thus Π is a PPM with intensity μ .

σ -finite intensity. For a σ -finite intensity, partition $E = \bigsqcup_{n \geq 1} E_n$ with $\mu(E_n) < \infty$, construct independent PPMs (again using the countable product construction) Π_n with intensities $\mu(\cdot \cap E_n)$, and set

$$\Pi = \sum_{n \geq 1} \Pi_n.$$

The defining Poisson laws and independence follow immediately from the finite case and Proposition 1.2. $\square$

Corollary 1.5 (Avoidance probability). *If $\Pi \sim \text{PPM}(E, \mathcal{E}, \mu)$ then for every $A \in \mathcal{E}$,*

$$\mathbb{P}(\Pi(A) = 0) = e^{-\mu(A)},$$

where the right-hand side is interpreted as 0 when $\mu(A) = \infty$.

The following exercise gives a useful converse to the defining properties of a PPM: for a simple locally finite random point measure on $\mathbb{R}_+$, the void probabilities already determine the law.

Exercise 1.2 (Rényi's characterization of Poisson point measures). Let Π be an almost surely simple point measure

$$\Pi = \sum_{i \in I} \delta_{x_i}$$

³The required probability space can be obtained from the standard countable product-measure theorem, applied to $E^{\mathbb{Z}_{>0}}$ with its product σ -field, and then taking a product with a space carrying the Poisson variable N . The existence of such a probability space is in fact the most complicated part of the proof, but usually the most overlooked. However, this construction is done in undergraduate probability courses.

on $\mathbb{R}_+$, where I is finite or countable and the atoms x_i are pairwise distinct. Assume that Π is almost surely finite on compact sets and that, for every Borel set $A \subseteq \mathbb{R}_+$,

$$\mathbb{P}(\Pi(A) = 0) = e^{-\mu(A)},$$

where μ is a diffuse Borel measure that is finite on compact sets (with the convention $e^{-\infty} = 0$). The goal is to prove that

$$\Pi \sim \text{PPM}(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \mu).$$

1. Let $I_1, \dots, I_m$ be pairwise disjoint Borel sets. Show that the events

$$\{\Pi(I_j) = 0\}, \quad 1 \leq j \leq m,$$

are independent.

2. Fix $0 \leq a < b < \infty$, and for $0 \leq k \leq 2^n - 1$ set

$$D_{n,k} = \left[a + \frac{k}{2^n}(b-a), a + \frac{k+1}{2^n}(b-a) \right).$$

Show that, almost surely,

$$\Pi([a, b)) = \lim_{n \rightarrow \infty} \sum_{k=0}^{2^n-1} \mathbf{1}_{\{\Pi(D_{n,k}) \neq 0\}}.$$

3. Deduce that, for every $z \in \mathbb{C}$ with $|z| < 1$,

$$\mathbb{E} \left[z^{\Pi([a, b))} \right] = \lim_{n \rightarrow \infty} \prod_{k=0}^{2^n-1} \left(z + (1-z)e^{-\mu(D_{n,k})} \right).$$

4. Show that

$$\max_{0 \leq k \leq 2^n-1} \mu(D_{n,k}) \xrightarrow{n \rightarrow \infty} 0,$$

and use the product in the previous question to determine the law of $\Pi([a, b))$.

5. Show that, for pairwise disjoint half-open intervals $[a_i, b_i)$, the random variables

$$\Pi([a_i, b_i))$$

are independent.

6. Deduce that Π is a Poisson point measure on $\mathbb{R}_+$ with intensity μ .

1.3 Basic operations

The stability properties of Poisson random variables under addition and splitting translate into analogous properties for Poisson point measures. The proofs are left as exercises.

Proposition 1.6 (Additivity). *Let $\Pi_1, \Pi_2, \dots$ be independent PPMs on $(E, \mathcal{E})$ with respective intensities $\mu_1, \mu_2, \dots$. Then*

$$\Pi = \sum_{n \geq 1} \Pi_n$$

is a Poisson point measure with intensity $\mu = \sum_{n \geq 1} \mu_n$.

Proposition 1.7 (Restriction). *Let $\Pi \sim \text{PPM}(E, \mathcal{E}, \mu)$ and let $B \in \mathcal{E}$. Then*

$$\Pi_B = \Pi(\cdot \cap B), \quad \Pi_{B^c} = \Pi(\cdot \cap B^c)$$

are independent PPMs with respective intensities $\mu(\cdot \cap B)$ and $\mu(\cdot \cap B^c)$.

Proposition 1.8 (Mapping). *Let $\Pi \sim \text{PPM}(E, \mathcal{E}, \mu)$ and let $\phi : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ be measurable. Define, for $B \in \mathcal{F}$,*

$$(\phi_{\#}\Pi)(B) = \Pi(\phi^{-1}(B)), \quad (\phi_{\#}\mu)(B) = \mu(\phi^{-1}(B)).$$

Then $\phi_{\#}\Pi$ is a PPM on $(F, \mathcal{F})$ with intensity $\phi_{\#}\mu$.

Even when μ is σ -finite, its image $\phi_{\#}\mu$ need not be σ -finite. For example, the projection $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\phi(x, y) = x$, sends two-dimensional Lebesgue measure to the measure

$$(\phi_{\#}\text{Leb}_2)(A) = \begin{cases} 0, & \text{Leb}_1(A) = 0, \\ +\infty, & \text{Leb}_1(A) > 0, \end{cases}$$

which is not σ -finite. Thus the mapping proposition also produces a PPM on $\mathbb{R}$ with this non- σ -finite intensity.

Remark 1.4 (σ -finite and s -finite measures). A measure which is a countable sum of finite measures is called *s-finite*. This condition is strictly weaker than σ -finiteness and is preserved under measurable mappings. The construction above extends to s -finite intensities by superposing independent finite-intensity PPMs. We shall nevertheless mostly assume σ -finiteness, since it is the standard hypothesis and it simplifies several later arguments, including Theorem 1.11.

Proposition 1.9 (Marking). *Let $\Pi \sim \text{PPM}(E, \mathcal{E}, \mu)$ and suppose that we are given a measurable enumeration*

$$\Pi = \sum_{i=1}^N \delta_{X_i}, \quad N \in \mathbb{Z}_{\geq 0} \cup \{\infty\}.$$

Let $(Y_i)_{i \geq 1}$ be i.i.d. random variables with values in $(F, \mathcal{F})$ and common law ν , independent of N and of the sequence $(X_i)_{i \geq 1}$. Then

$$\Pi^F := \sum_{i=1}^N \delta_{(X_i, Y_i)} \sim \text{PPM}(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu).$$

The proof of the above proposition is slightly more complicated than the additivity, mapping and restriction properties. We will see a slick proof using the exponential formula in the next chapter, but in the mean time we encourage the reader to find their own proof.

1.4 Recovering the points

Theorem 1.4 is very general: it only assumes σ -finiteness of the intensity. On an arbitrary measurable space, however, it is not automatic that the atoms of a PPM can be recovered as measurable random variables. The following example shows why an additional measurability assumption is needed. Let $E = [0, 1]$ and let $\mathcal{E}$ be the countable-cocountable σ -field. Define

$$\mu(A) = \begin{cases} 0, & A \text{ countable,} \\ 1, & A^c \text{ countable.} \end{cases}$$

Let $(X_i)_{i \geq 1}$ be i.i.d. E -valued variables with law μ , let $N \sim \text{Poi}(1)$ be independent of them, and set

$$\Pi = \sum_{i=1}^N \delta_{X_i}.$$

Also define the random measure $\tilde{\Pi} = N\mu$, i.e. $\tilde{\Pi}(A) = N\mu(A)$. The random measures Π and $\tilde{\Pi}$ have the same finite-dimensional laws for all their evaluations on $\mathcal{E}$, and these laws satisfy the Poisson count and independence formulas. However, Π is a point measure whereas, on $\{N > 0\}$, $\tilde{\Pi}$ is not a point measure. Thus the Poisson formulas alone do not characterize a point measure among all random measures; the point-measure requirement in Definition 1.3 is genuine.

To recover individual points we shall use the following assumption:

$$(\Delta) \quad \Delta_E := \{(x, x) : x \in E\} \in \mathcal{E} \otimes \mathcal{E}. \quad (1.3)$$

If $A \in \mathcal{E} \otimes \mathcal{E}$, then every section $A_x = \{y : (x, y) \in A\}$ belongs to $\mathcal{E}$. In particular, under (Δ) every singleton $\{x\}$ is measurable. The assumption holds in all standard Borel spaces, and in particular for the Borel σ -field of a Polish space.

Exercise 1.3. Show that the countable-cocountable σ -field in the preceding example does not satisfy (Δ) .

A useful equivalent formulation of (Δ) is the existence of a countable family $(C_n)_{n \geq 1} \subset \mathcal{E}$ separating points⁴. Indeed, if such a family exists, then

$$\Delta_E = \bigcap_{n \geq 1} ((C_n \times C_n) \cup (C_n^c \times C_n^c)).$$

Conversely, every element of the σ -field generated by measurable rectangles belongs to the σ -field generated by countably many of those rectangles. Apply this to Δ_E : the countably many sides of these rectangles must separate points, otherwise (x, x) and (x, y) could not be distinguished for some $x \neq y$.

We start with a simpler question than recovering the atoms, namely deciding whether two atoms coincide. By Example 1.1, this is the case with positive probability as soon as μ has an atom, the following proposition proves the converse:

Proposition 1.10 (Simplicity). *Suppose that Π is a PPM on $(E, \mathcal{E})$ satisfying (Δ) with a σ -finite intensity μ . If μ is diffuse, then Π is almost surely simple.*

Proof. We first treat the case $\mu(E) < \infty$. Let $(C_j)_{j \geq 1}$ be a countable family separating points, and let $\mathcal{P}_n$ be the finite partition generated by $C_1, \dots, C_n$. We claim that

$$\delta_n := \max_{D \in \mathcal{P}_n} \mu(D) \xrightarrow{n \rightarrow \infty} 0. \quad (1.4)$$

If not, there exists $\varepsilon > 0$ and cells of arbitrarily fine partitions with mass at least ε . Since the partitions are finite and refining, König's lemma⁵ yields a nested sequence

$$D_1 \supset D_2 \supset \dots, \quad D_n \in \mathcal{P}_n, \quad \mu(D_n) \geq \varepsilon.$$

Because (C_j) separates points, $\bigcap_n D_n$ contains at most one point. Continuity from above, using $\mu(D_1) < \infty$, gives

$$\mu\left(\bigcap_{n \geq 1} D_n\right) = \lim_{n \rightarrow \infty} \mu(D_n) \geq \varepsilon,$$

contradicting diffuseness. Thus $\delta_n \rightarrow 0$.

Now let

$$A_n = \{\exists D \in \mathcal{P}_n : \Pi(D) \geq 2\},$$

and notice that $\{\Pi \text{ has an atom of multiplicity at least two}\} \subset \bigcap_n A_n$ so it suffices to prove that $\mathbb{P}(A_n) \rightarrow 0$ as $n \rightarrow \infty$. Before doing that by a variation of the first-moment method, if $N \sim \text{Poi}(\lambda)$, then

$$\mathbb{P}(N \geq 2) \leq \frac{1}{2} \mathbb{E}[N(N-1)] = \frac{\lambda^2}{2}.$$

⁴Equivalently, $(E, \mathcal{E})$ admits an injective measurable map into $[0, 1]$. This is strictly weaker than being a standard Borel space.

⁵if you have never seen this lemma, make the proof yourself!

Consequently,

$$\begin{aligned}\mathbb{P}(A_n) &\leq \sum_{D \in \mathcal{P}_n} \mathbb{P}(\Pi(D) \geq 2) \\ &\leq \frac{1}{2} \sum_{D \in \mathcal{P}_n} \mu(D)^2 \\ &\leq \frac{1}{2} \delta_n \sum_{D \in \mathcal{P}_n} \mu(D) = \frac{1}{2} \delta_n \mu(E) \xrightarrow[n \rightarrow \infty]{(1.4)} 0.\end{aligned}$$

This proves the finite-intensity case.

For a general σ -finite diffuse intensity, partition E into countably many sets of finite μ -mass and apply the finite case to each restriction of Π . A countable intersection of the resulting probability-one events gives the conclusion. $\square$

Theorem 1.11 (Recovering the points)

Assume that $E \neq \emptyset$, and suppose that Π is a PPM on $(E, \mathcal{E})$ satisfying (Δ) and having a σ -finite diffuse intensity μ . Then the atoms of Π can be enumerated measurably: there exist E -valued random variables $X_1, X_2, \dots$ (with arbitrary values after the last atom when $\Pi(E) < \infty$) such that, with $N = \Pi(E) \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$,

$$\Pi = \sum_{i=1}^N \delta_{X_i}.$$

Moreover, each X_i may be chosen as a measurable function of the point measure Π , i.e. using only the evaluation maps $\nu \mapsto \nu(A)$, $A \in \mathcal{E}$. By Proposition 1.10, these atoms are almost surely distinct.

Proof. Fix a point $x_* \in E$, to be used as a dummy value when necessary, and let $(C_n)_{n \geq 1}$ be a countable family separating points. The binary codes $(\mathbf{1}_{C_n}(x))_{n \geq 1}$ induce a lexicographic order $<$ on E . The idea is simply to enumerate the atoms according to this order.

There is one small nuisance: an arbitrary countable subset of E need not have a $<$ -smallest element. We therefore begin with the finite-intensity case. Then $\Pi(E) < \infty$ almost surely and, by Proposition 1.10, we may write its atoms as

$$X_1 < X_2 < \dots < X_N, \quad N = \Pi(E).$$

The only point to check is that these atoms can be chosen measurably from Π . We explain this only for X_1 , which contains the main idea.

For $n \geq 1$, let

$$\mathcal{P}_n = \{D_1^n, \dots, D_{r_n}^n\}$$

be the finite partition generated by $C_1, \dots, C_n$, with its cells listed in lexicographic order of their binary labels. The partitions refine one another, and since Π has only finitely many atoms, for all sufficiently large n distinct atoms belong to distinct cells. Hence, on $\{N \geq 1\}$ and up to a null event, for every $A \in \mathcal{E}$,

$$\{X_1 \in A\} = \bigcup_{n \geq 1} \bigcap_{m \geq n} \bigcup_{1 \leq i \leq r_m} \left(\{\Pi(A \cap D_i^m) = 1\} \right. \\ \left. \cap \{\Pi(D_1^m \cup \dots \cup D_{i-1}^m) = 0\} \right).$$

The right-hand side is measurable in terms of the evaluation maps. Extending X_1 by x_* on $\{N = 0\}$ therefore makes X_1 measurable. The other selectors are obtained in the same way by looking successively for the second, third, *etc.* occupied cell; the notation is heavier but no new idea is involved.

Finally, in the σ -finite case, partition $E = \bigsqcup_{m \geq 1} E_m$ with $\mu(E_m) < \infty$, apply the finite construction to each restriction of Π to E_m , and concatenate the resulting finite lists. The counts $\Pi(E_m)$ are measurable, so this concatenation is measurable as well. This gives the desired enumeration. $\square$

The take-away message of this section is that PPMs on well-behaved spaces, such as Polish spaces, with diffuse intensities can be viewed as finite or countable collections of distinct random points. There is then no issue with measurability, and one can manipulate these objects without undue concern. We shall be less strict in the coming chapters on measurability checks.

II: Useful formulas

Throughout this chapter, unless explicitly stated otherwise, we consider a Poisson point measure Π on a measurable space $(E, \mathcal{E})$ with a σ -finite intensity measure μ . Recall that if $f : E \rightarrow [0, \infty]$ is measurable and ν is a positive measure on E , we use the notation

$$\nu(f) = \int_E f \, d\nu = \int_E f(x) \, \nu(dx).$$

In particular, if $\nu = \sum_i \delta_{x_i}$ is a point measure, then

$$\nu(f) = \sum_i f(x_i) \quad \text{and takes values in } [0, \infty].$$

2.1 Exponential formula

Theorem 2.1 (*Exponential formula*)

For every nonnegative measurable function $f : E \rightarrow [0, \infty]$,

$$\mathbb{E}[\exp(-\Pi(f))] = \exp\left(-\int_E (1 - e^{-f}) \, d\mu\right), \quad (2.1)$$

where as usual we use the convention $e^{-\infty} = 0$.

A mnemonic way to remember the formula is to notice that there are only minus signs.

Proof. First suppose that f is a simple function

$$f = \sum_{j=1}^r t_j \mathbf{1}_{A_j},$$

where $t_j \geq 0$, the sets A_j are pairwise disjoint (not necessarily of finite μ -measure). Then

$$\Pi(f) = \sum_{j=1}^r t_j \Pi(A_j).$$

By the defining properties of a Poisson point measure,

$$\begin{aligned}\mathbb{E}[e^{-\Pi(f)}] &= \prod_{j=1}^r \mathbb{E}[e^{-t_j \Pi(A_j)}] \\ &= \prod_{j=1}^r \exp(\mu(A_j)(e^{-t_j} - 1)) \\ &= \exp\left(-\sum_{j=1}^r \mu(A_j)(1 - e^{-t_j})\right) \\ &= \exp\left(-\int_E (1 - e^{-f}) d\mu\right).\end{aligned}$$

Thus the formula holds for such simple functions.

Now let $f : E \rightarrow [0, \infty]$ be arbitrary. There exists a sequence of simple positive functions f_n such that $f_n \uparrow f$. We apply the dominated convergence theorem on the left-hand side and the monotone convergence theorem inside the integral on the right-hand side to complete the proof. $\square$

The exponential formula determines the law of a random point measure in the sense of Section 1.2.1 (and even in the sense of a random measure when (E, d) is Polish, see Chapter 3). Indeed, taking

$$f = \sum_{j=1}^r t_j \mathbf{1}_{A_j}, \quad t_j \geq 0,$$

gives the joint Laplace transform, hence the law of every finite family $\Pi(A_1), \dots, \Pi(A_r)$. It is easy to see that the Laplace transform $(t_1, \dots, t_n) \mapsto \mathbb{E}[\exp(-\sum t_i N_i)]$ characterizes the law of a random integer vector $\mathbf{N} = (N_1, \dots, N_n) \in (\mathbb{Z}_{\geq 0} \cup \{\infty\})^n$. In particular, we can now prove the marking property stated in Proposition 1.9.

Proof of Proposition 1.9. Use the notation of that proposition and set

$$\mathcal{G} = \sigma(N, X_1, X_2, \dots).$$

Let $f : E \times F \rightarrow [0, \infty]$ be measurable, and introduce the notation

$$q(x) = \int_F e^{-f(x,y)} \nu(dy), \quad g(x) = -\log q(x),$$

with the convention $-\log 0 = +\infty$. Conditionally on $\mathcal{G}$, the marks are independent with law ν , and therefore

$$\mathbb{E}[e^{-\Pi^F(f)} \mid \mathcal{G}] = \prod_{i=1}^N q(X_i) = \exp\left(-\sum_{i=1}^N g(X_i)\right) = e^{-\Pi(g)}.$$

Taking expectations and applying the usual exponential formula to Π gives

$$\begin{aligned}\mathbb{E}\left[e^{-\Pi^F(f)}\right] &= \exp\left(-\int_E (1 - e^{-g(x)}) \mu(dx)\right) \\ &= \exp\left(-\int_E \int_F (1 - e^{-f(x,y)}) \nu(dy) \mu(dx)\right) \\ &= \exp\left(-\int_{E \times F} (1 - e^{-f}) d(\mu \otimes \nu)\right).\end{aligned}$$

This is the exponential formula for a Poisson point measure on $E \times F$ with intensity $\mu \otimes \nu$. Since the exponential formula determines the law of a random point measure, the conclusion of Proposition 1.9 follows. $\square$

The exponential formula gives access to the law of $\Pi(f)$ for every nonnegative measurable f . In particular, it gives its first moment; this identity is a special case of Campbell's¹ formula:

Theorem 2.2 (Campbell formula)

For every nonnegative measurable function $f : E \rightarrow [0, \infty]$,

$$\mathbb{E}\left[\int_E f(x) \Pi(dx)\right] = \int_E f(x) \mu(dx).$$

In particular, if $\int d\mu f < \infty$ then $\Pi(f)$ is almost surely finite. Conversely, if $\mu(E) = \infty$ and

$$\int_E \mu(dx) 1 \wedge f(x) = \infty,$$

then $\Pi(f) = \infty$ almost surely.

Proof. This could be deduced from the exponential formula, but the proof from scratch is quick: For $f = \mathbf{1}_A$ with $\mu(A) < \infty$, this is simply

$$\mathbb{E}[\Pi(A)] = \mu(A).$$

By linearity, the formula follows for nonnegative simple functions supported on a set of finite μ -mass. For arbitrary nonnegative measurable f , choose $f_m \uparrow f$ as in the proof of Theorem 2.1 and apply monotone convergence on both sides. This proves the Campbell formula and we deduce that $\Pi(f) < \infty$ as soon as f is μ -integrable. Conversely, recall from Lemma 1.1 that $\mathbb{P}(\Pi(f) < \infty) = \lim_{t \downarrow 0} \mathbb{E}[e^{-t\Pi(f)}]$. We can then use the exponential formula



¹ Norman Robert Campbell (1880-1949), a British physicist.

to get

$$\mathbb{E}[e^{-t\Pi(f)}] = \exp\left(-\int \mu(dx)(1 - e^{-tf(x)})\right).$$

When $t \rightarrow 0$ we have $(1 - e^{-tf(x)}) \rightarrow \mathbf{1}_{f(x)=\infty}$ so in case of finite measure, by dominated convergence theorem we have

$$\mathbb{P}(\Pi(f) < \infty) = \exp\left(-\int \mu(dx)\mathbf{1}_{f(x)=\infty}\right).$$

However, if μ is infinite and $\int d\mu (1 \wedge f) = \infty$, using $1 - e^{-u} \geq (1 - e^{-1})(1 \wedge u)$ we have, for $0 < t \leq 1$,

$$\int \mu(dx)(1 - e^{-tf(x)}) \geq (1 - e^{-1}) \int \mu(dx) (1 \wedge (tf(x))) \geq (1 - e^{-1})t \int \mu(dx) (1 \wedge f(x)) = \infty,$$

and so even when letting $t \rightarrow 0$ we get $\mathbb{P}(\Pi(f) < \infty) = 0$. $\square$

If $f : E \rightarrow \mathbb{R}$ is measurable and

$$\int_E |f| d\mu < \infty,$$

then Campbell's formula applied to $|f|$ gives

$$\mathbb{E}[\Pi(|f|)] = \mu(|f|) < \infty.$$

Hence $\Pi(|f|) < \infty$ almost surely, so the signed integral $\Pi(f)$ is well defined almost surely, belongs to L^1 , and

$$\mathbb{E}[\Pi(f)] = \int_E f d\mu.$$

2.2 The Mecke formula

Campbell's formula computes the expectation of a sum whose contribution at a point depends only on that point. The Mecke formula allows the contribution to depend on the entire configuration. The key of the formula sits on the size-biasing property of Poisson random variables:

Exercise 2.1. Let $\lambda > 0$ and $N \sim \text{Poi}(\lambda)$. For every $f : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}_+$ we have

$$\frac{1}{\lambda} \mathbb{E}[Nf(N)] = \mathbb{E}[f(N+1)].$$

In words, the variable N biased by N itself has the same distribution as $N+1$.

Theorem 2.3 (Mecke formula)

Let $\Pi \sim \text{PPM}(\mu)$, where μ is σ -finite. Let $H : E \times \mathcal{N}(E) \rightarrow [0, \infty]$ be $\mathcal{E} \otimes \mathfrak{N}$ -measurable.

Then

$$\mathbb{E} \left[\int_E H(x, \Pi) \Pi(dx) \right] = \int_E \mathbb{E}[H(x, \Pi + \delta_x)] \mu(dx).$$

Both sides may be infinite.

If H is real-valued rather than nonnegative, the same identity holds as soon as one side is absolutely integrable; equivalently it is enough to apply the nonnegative formula to $|H|$ and then subtract positive and negative parts.

Remark 2.1. The above formula has several names in the literature, sometimes depending on the field: Mecke² formula, Slivnyak–Mecke formula, refined Campbell formula, or even Palm formula. This is surely another application of Stigler’s law of eponymy.

Proof. We leave to the reader the technical exercise of checking that all quantities are measurable and well-defined. We first assume that $\mu(E) = \lambda < \infty$. If $\lambda = 0$ the result is immediate, so suppose $\lambda > 0$. Since both sides depend measurably only on the point measure Π , whose law is fixed by its intensity, it is enough to verify the formula for the construction of Theorem 1.4. Let $N \sim \text{Poi}(\lambda)$ and, conditionally on N , let $X_1, \dots, X_N$ be i.i.d. with law μ/λ . Then

$$\Pi = \sum_{j=1}^N \delta_{X_j}.$$

For brevity, write $\mathbf{x}_n = (x_1, \dots, x_n)$. By Tonelli’s theorem,

$$\begin{aligned} \mathbb{E} \left[\int_E H(x, \Pi) \Pi(dx) \right] \\ = e^{-\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{E^n} \sum_{i=1}^n H \left(x_i, \sum_{j=1}^n \delta_{x_j} \right) \mu^{\otimes n}(d\mathbf{x}_n). \end{aligned}$$

The product measure $\mu^{\otimes n}$ is invariant under permutations of the coordinates. Therefore all n terms in the inner sum have the same integral, and

$$\begin{aligned} \mathbb{E} \left[\int_E H(x, \Pi) \Pi(dx) \right] \\ = e^{-\lambda} \sum_{n=1}^{\infty} \frac{n}{n!} \int_{E^n} H \left(x_1, \sum_{j=1}^n \delta_{x_j} \right) \mu^{\otimes n}(d\mathbf{x}_n). \end{aligned}$$



² Joseph Mecke (1938-2014), a German mathematician.



Izrail Mikhailovich Slivnyak

(????- <2012), a Soviet mathematician.

Writing $m = n - 1$, separating the first coordinate, and applying Tonelli again, we obtain

$$\begin{aligned} \mathbb{E} \left[\int_E H(x, \Pi) \Pi(dx) \right] \\ = \int_E \left[e^{-\lambda} \sum_{m=0}^{\infty} \frac{1}{m!} \int_{E^m} H \left(x, \delta_x + \sum_{j=1}^m \delta_{y_j} \right) \mu^{\otimes m}(d(y_1, \dots, y_m)) \right] \mu(dx). \end{aligned}$$

The expression in square brackets is precisely $\mathbb{E}[H(x, \Pi + \delta_x)]$. This proves the formula when $\mu(E) < \infty$.

We now treat the σ -finite case. Choose $F_m \uparrow E$ with $\mu(F_m) < \infty$, and set

$$\Pi_m^{\text{in}} = \Pi|_{F_m}, \quad \Pi_m^{\text{out}} = \Pi|_{F_m^c}.$$

By Proposition 1.7, these two random point measures are independent. Conditioning on Π_m^{out} , we apply the finite-intensity Mecke formula to Π_m^{in} with the measurable function

$$H_m(x, \Pi) = H_m(x, \Pi_m^{\text{out}} + \Pi_m^{\text{in}}) = \mathbf{1}_{F_m}(x) \cdot H(x, \Pi_m^{\text{out}} + \Pi_m^{\text{in}})$$

and obtain

$$\begin{aligned} \mathbb{E} \left[\int_{F_m} H(x, \Pi) \Pi(dx) \right] &= \mathbb{E} \left[\int_{F_m} H_m(x, \Pi_m^{\text{in}} + \Pi_m^{\text{out}}) \Pi_m^{\text{in}}(dx) \right] \\ &= \mathbb{E}^{\text{out}} \left[\int_{F_m} \mathbb{E}^{\text{in}} [H_m(x, \Pi_m^{\text{in}} + \Pi_m^{\text{out}} + \delta_x)] \mu(dx) \right] \\ &= \int_{F_m} \mathbb{E}^{\text{in+out}} [H_m(x, \Pi_m^{\text{in}} + \Pi_m^{\text{out}} + \delta_x)] \mu(dx). \end{aligned}$$

Finally, both sides increase as $m \rightarrow \infty$, so monotone convergence proves the formula on E . $\square$

Remark 2.2. Taking $H(x, \nu) = f(x)$ in the Mecke formula gives Campbell's formula. Thus Mecke's identity is a genuine extension of Campbell's identity.

A common interpretation of the Mecke formula is as follows: a typical point X of Π is distributed according to μ , and, conditionally on X , the remaining point measure has the same distribution as Π . The proper meaning of this statement is precisely the Mecke formula. One can also consider " Π conditioned to have an atom at x ". Even if this event has zero probability, the resulting point measure is obtained simply by adding δ_x to an independent sample of Π . For a general random point measure, the corresponding conditional law is called the Palm distribution³.



³ Conny Palm (1907-1951), a Swedish teletraffic engineer

2.2.1 Multi-point Mecke

The one-point Mecke formula has a very useful multi-point version. To get nice formulas, the correct object on the left-hand side is not the ordinary product measure $\Pi^{\otimes k}$ and one must exclude the possibility of selecting the same *occurrence* of a point more than once. The proof is left to the reader, and can be obtained by recursing from Theorem 2.3:

Theorem 2.4 (Multi-point Mecke formula)

Let $k \geq 1$ and let $H : E^k \times \mathcal{N}(E) \rightarrow [0, \infty]$ be $\mathcal{E}^{\otimes k} \otimes \mathfrak{N}$ -measurable. Let $\Pi \sim \text{PPM}(E, \mu)$ written as

$$\Pi = \sum_{i \in I} \delta_{X_i}.$$

Then

$$\mathbb{E} \left[\sum_{\substack{i_1, \dots, i_k \in I \\ \text{pairwise distinct}}} H(X_{i_1}, \dots, X_{i_k}, \Pi) \right] = \int_{E^k} \mathbb{E} \left[H \left(x_1, \dots, x_k, \Pi + \sum_{r=1}^k \delta_{x_r} \right) \right] \mu(dx_1) \mu(dx_2) \dots \mu(dx_k).$$

Specializing to functions $H(\mathbf{x}, \Pi) = h(\mathbf{x})$, we obtain the multi-point Campbell formula:

Corollary 2.5 (Factorial Campbell formula). For every $k \geq 1$ and every nonnegative measurable $h : E^k \rightarrow [0, \infty]$,

$$\mathbb{E} \left[\sum_{\substack{i_1, \dots, i_k \in I \\ \text{pairwise distinct}}} h(X_{i_1}, \dots, X_{i_k}) \right] = \int_{E^k} h(\mathbf{x}) \mu^{\otimes k}(d\mathbf{x}).$$

2.3 Exercises

Exercise 2.2 (Simplicity from the second factorial measure). Assume (Δ) and suppose that μ is diffuse. Use the multi-point Mecke formula to prove (again) that Π is almost surely simple.

Exercise 2.3 (Mecke's identity characterizes the Poisson law). Let $\mathbf{N}$ be a random point measure with σ -finite intensity measure μ and assume that, for every nonnegative measurable H ,

$$\mathbb{E} \left[\int_E H(x, \mathbf{N}) \mathbf{N}(dx) \right] = \int_E \mathbb{E} [H(x, \mathbf{N} + \delta_x)] \mu(dx).$$

First suppose that f is bounded, nonnegative, and vanishes outside a set of finite μ -mass. Set

$$L(t) = \mathbb{E}[e^{-t\mathbf{N}(f)}], \quad t \geq 0.$$

1. Apply Mecke's identity with

$$H(x, \nu) = f(x) e^{-t\nu(f)}$$

to derive a differential equation for L .

2. Deduce that $L(t) = \exp\left(-\int_E (1 - e^{-tf}) d\mu\right)$.
3. Extend the result by monotone approximation and deduce that $\mathbf{N}$ is a Poisson point measure with intensity μ .

Exercise 2.4 (Dickman distribution). Let Π be a PPM on $(0,1]$ with intensity

$$\mu(dx) = \frac{dx}{x},$$

and set

$$X = \int_{(0,1]} x \Pi(dx).$$

1. Show that $X < \infty$ almost surely and that $\mathbb{E}[X] = 1$.
2. Show that, for $\lambda \geq 0$,

$$\mathbb{E}[e^{-\lambda X}] = \exp\left(\int_0^1 \frac{e^{-\lambda x} - 1}{x} dx\right) = \exp(I(-\lambda)),$$

where

$$I(s) = \int_0^s \frac{e^t - 1}{t} dt.$$

This law is called the Dickman distribution.

3. Under the change of variables $x = e^{-t}$, identify the image of Π with a homogeneous Poisson process of rate 1 on $\mathbb{R}_+$. If T_1 denotes its first point, show that $U = e^{-T_1}$ is uniform on $(0,1)$ and deduce the distributional identity

$$X \stackrel{d}{=} U(1 + X'),$$

where X' is an independent copy of X , independent of U .

III: Random measures on Polish spaces

While the preceding chapters on PPMs required no topology, we shall now suppose that the underlying space E carries a Polish topology. This will enable us to regard random measures as random variables taking values in a Polish space and to discuss their convergence. The takeaway message is that a random measure is observed through the random integrals

$$\mu(f) = \int_E f d\mu,$$

for continuous functions (bounded or compactly supported, depending on the context). Throughout the chapter, E is a *locally compact Polish space*: it is separable and completely metrizable, and every point has a compact neighbourhood. We then specialize the convergence criteria to random point measures and Poisson random measures.

3.1 Radon measures

3.1.1 Weak and vague convergence

Let $\mathcal{M}(E)$ (resp. $\mathcal{M}_1(E)$) denote the space of non-negative Radon measures on E , that is, Borel measures which are finite on compact sets (resp. probability measures). Write $C_c(E)$ (resp. $C_b(E)$) for the continuous real-valued functions with compact support (resp. bounded continuous functions).

Definition 3.1 (Vague and weak convergence). *A sequence $\mu_n \in \mathcal{M}(E)$ converges vaguely (resp. weakly) to $\mu \in \mathcal{M}(E)$ if*

$$\mu_n(f) \longrightarrow \mu(f) \quad \text{for every } f \in C_c(E) \quad (\text{resp. for every } f \in C_b(E)). \quad (3.1)$$

We write $\mu_n \xrightarrow{(v)} \mu$ and $\mu_n \xrightarrow{(d)} \mu$.

Weak convergence is implicitly restricted to finite measures. Thus the vague and weak topologies are precisely the weakest topologies for which the maps

$$\mu \longmapsto \mu(f)$$

are continuous, with $f \in C_c(E)$ for the vague topology and $f \in C_b(E)$ for the weak topology, respectively.

Example 3.1 (Mass may escape to infinity). On $\mathbb{R}$,

$$\delta_n \xrightarrow{v} 0,$$

so δ_n converges vaguely but not weakly. Vague convergence is therefore a *local* notion: it sees what happens on bounded regions and ignores mass which escapes to infinity.

When E is compact, Prokhorov's theorem tells us that $\mathcal{M}_1(E)$ endowed with weak convergence is itself a Polish space (e.g., under the Prokhorov metric). By localization, this implies that $\mathcal{M}(E)$ is also a Polish space. In particular, the Borel σ -field of $\mathcal{M}(E)$ is generated by the coordinate maps

$$\mu \mapsto \mu(f), \quad f \in C_c(E). \quad (3.2)$$

A countable convergence-determining family $\{f_j\}$ already generates the same σ -field. More concretely, one may choose a countable family $\{f_j : j \geq 1\} \subset C_c(E)$ which determines vague convergence and build a distance from it. Via this identification, $\mathcal{M}(E)$ is embedded in $\mathbb{R}^{\mathbb{N}}$ with the topology of coordinatewise convergence.

3.1.2 Random measures

Once $\mathcal{M}(E)$ is seen as a Polish space, one can make the following definition:

Definition 3.2. A random measure Ξ on E is a random element of $\mathcal{M}(E)$ endowed with the Borel σ -field of the vague topology. Equivalently, $\omega \mapsto \Xi_\omega$ is measurable as a map into $\mathcal{M}(E)$. Because of (3.2), this is equivalent to requiring that

$$\Xi(f) = \int_E f \, d\Xi$$

be a real random variable for every $f \in C_c(E)$.

The notion of convergence in law is now the ordinary one for random elements of the Polish space $\mathcal{M}(E)$. We write

$$\Xi_n \xrightarrow[n \rightarrow \infty]{(d)} \Xi \quad \text{if} \quad \mathbb{E}[F(\Xi_n)] \longrightarrow \mathbb{E}[F(\Xi)] \quad (3.3)$$

for every bounded continuous functional $F : \mathcal{M}(E) \rightarrow \mathbb{R}$, where continuity is understood for the vague topology. At first sight, (3.3) seems difficult to verify: it would require testing every bounded continuous functional on an infinite-dimensional space. In practice, one needs only to test the convergence in law of the integrals $\Xi(f)$.

Proposition 3.1 (Testing random measures by functions). *Let Ξ_n and Ξ be random measures on E . The following are equivalent.*

1. $\Xi_n \xrightarrow[n \rightarrow \infty]{(d)} \Xi$ in $\mathcal{M}(E)$;

2. for every $f \in C_c(E)$,

$$\Xi_n(f) \xrightarrow[n \rightarrow \infty]{(d)} \Xi(f);$$

3. for every $f \in C_c^+(E)$,

$$\mathbb{E}[e^{-\Xi_n(f)}] \longrightarrow \mathbb{E}[e^{-\Xi(f)}]. \quad (3.4)$$

4. for every finite family $f_1, \dots, f_k \in C_c(E)$, we have the joint convergence

$$(\Xi_n(f_1), \dots, \Xi_n(f_k)) \Rightarrow (\Xi(f_1), \dots, \Xi(f_k)).$$

Proof. Since the map $\mu \mapsto \mu(f)$ is continuous for the vague topology, the continuous mapping theorem shows that 1) $\Rightarrow$ 2). Because $x \mapsto e^{-x}$ is bounded and continuous on $[0, \infty)$, 2) $\Rightarrow$ 3).

Let us show that 3) $\Rightarrow$ 4). Fix $f_1, \dots, f_k \in C_c(E)$ and write $f_i = f_i^+ - f_i^-$. For every $t_1, \dots, t_{2k} \geq 0$, condition 3), applied to the non-negative function

$$t_1 f_1^+ + t_2 f_1^- + \dots + t_{2k-1} f_k^+ + t_{2k} f_k^-,$$

shows the convergence of the corresponding joint Laplace transforms. By the continuity theorem for multivariate Laplace transforms,

$$(\Xi_n(f_1^+), \Xi_n(f_1^-), \dots, \Xi_n(f_k^+), \Xi_n(f_k^-)) \xrightarrow{(d)} (\Xi(f_1^+), \Xi(f_1^-), \dots, \Xi(f_k^+), \Xi(f_k^-)).$$

The continuous mapping theorem then gives 4).

It remains to show that 4) $\Rightarrow$ 1). Let $\{f_j : j \geq 1\} \subset C_c(E)$ be the countable convergence-determining family introduced above and consider

$$T : \mathcal{M}(E) \longrightarrow \mathbb{R}^{\mathbb{N}}, \quad T(\mu) = (\mu(f_j))_{j \geq 1}.$$

By construction, T is a topological embedding. Condition 4) gives convergence of all finite-dimensional marginals of $T(\Xi_n)$. Moreover, the convergence of each coordinate implies tightness of that coordinate; in the countable product $\mathbb{R}^{\mathbb{N}}$, this implies tightness of the whole sequence. Consequently,

$$T(\Xi_n) \xrightarrow{(d)} T(\Xi) \quad \text{in } \mathbb{R}^{\mathbb{N}}.$$

By the Skorokhod representation theorem, we may therefore work on a probability space on which

$$\Xi_n(f_j) \longrightarrow \Xi(f_j) \quad \text{almost surely for every } j \geq 1.$$

Since $\{f_j\}$ determines vague convergence, it follows that

$$\Xi_n \xrightarrow{(v)} \Xi \quad \text{almost surely,}$$

and hence $\Xi_n \xrightarrow{(d)} \Xi$. $\square$

3.2 Point measures

We now specialize to *point measures* $\nu \in \mathcal{M}(E)$ that can be written

$$\nu = \sum_i \delta_{x_i}$$

with only finitely many atoms in each compact set. Multiplicities are allowed. Recall from the first chapter that a random point measure Π is measurable if and only if $\Pi(A)$ is a random variable for every $A \in \mathcal{E}$, the σ -field of E . When E has the additional topology required in this chapter and when $\mathcal{E}$ is its associated Borel σ -field, this automatically implies that Π is a random measure in the sense of Definition 3.2. Indeed, for any $f \in C_c(E)$, approximating f by step functions and taking limits shows that $\Pi(f)$ is measurable. The exponential formula (2.1) characterizes its distribution by Proposition 3.1.

3.2.1 Convergence of atoms

For random point measures, convergence in distribution in the sense of (3.3) is in fact equivalent to the simultaneous convergence of atoms. To state this precisely, one must be able to enumerate the atoms in a measurable way, so we restrict to $\mathbb{R}_+$ for simplicity:

Proposition 3.2 (Vague convergence and ordered atoms). *Let*

$$\mathbf{N}_n = \sum_{k \geq 1} \delta_{T_{n,k}}, \quad \mathbf{N} = \sum_{k \geq 1} \delta_{T_k},$$

be random locally finite point measures on $\mathbb{R}_+$, with all atoms strictly positive and ordered in increasing order. Since $\mathbf{N}$ is locally finite, we in particular have $T_k \rightarrow \infty$ a.s. as $k \rightarrow \infty$. Then

$$\mathbf{N}_n \xrightarrow{(d)} \mathbf{N} \iff (T_{n,k})_{k \geq 1} \xrightarrow[n \rightarrow \infty]{(d)} (T_k)_{k \geq 1}$$

that is, $(T_{n,k})_{1 \leq k \leq K} \rightarrow (T_k)_{1 \leq k \leq K}$ in law for each fixed K as $n \rightarrow \infty$.

Proof. Using the Skorokhod representation theorem, we can assume that $\mathbf{N}_n \rightarrow \mathbf{N}$ almost surely in the vague sense. It is then an easy deterministic exercise to see that $T_{n,k} \xrightarrow{T_k} T_k$ almost

surely. This proves the first implication. Conversely, assume that $(T_{n,k})_{k \geq 1} \xrightarrow{(d)} (T_k)_{k \geq 1}$. Using the Skorokhod representation theorem again, we can assume that $T_{n,k} \rightarrow T_k$ almost surely for every k . Let $f \in C_c(\mathbb{R}_+)$ and choose R such that $\text{supp}(f) \subset [0, R]$. Since $T_k \rightarrow \infty$, choose K such that $T_K > R$. For n large enough, $T_{n,K} > R$, and since the atoms are ordered, $f(T_{n,k}) = 0$ for every $k \geq K$. Hence

$$\mathbf{N}_n(f) = \sum_{k=1}^{K-1} f(T_{n,k}) \rightarrow \sum_{k=1}^{K-1} f(T_k) = \mathbf{N}(f).$$

Thus $\mathbf{N}_n \rightarrow \mathbf{N}$ almost surely vaguely, which proves the converse implication. $\square$

Remark 3.1 (Why the state space matters). On $(0, \infty)$, an atom may disappear through the boundary:

$$\delta_{1/n} \xrightarrow{v} 0.$$

This is why the ordered-atom statement is cleanest on $\mathbb{R}_+ = [0, \infty)$ with a limiting process having no atom at 0. On the whole line $\mathbb{R}$, a stationary Poisson process has infinitely many atoms in both directions and there is no first atom; one instead orders points relative to an origin or works on bounded windows.

3.2.2 An example

We end this chapter by illustrating the convergence of a discrete counting measure to a Poisson random measure. Fix $d \geq 1$ and consider the counting measure Π_p obtained by placing an atom independently at each point of $\mathbb{Z}^d$ with probability p . We then consider its rescaled version $\alpha \cdot \Pi_p$, formally defined by

$$\alpha \cdot \Pi_p(A) = \Pi_p(\alpha \cdot A), \quad \text{for all Borel sets } A \subset \mathbb{R}^d.$$

Proposition 3.3. *The following convergence holds:*

$$n \cdot \Pi_{c/n^d} \xrightarrow[n \rightarrow \infty]{(d)} \text{PPM}(c \cdot \text{Leb}_{\mathbb{R}^d}),$$

in law for the vague topology on random measures.

Proof. Let $f \in C_c^+(\mathbb{R}^d)$ and set $g = 1 - e^{-f}$. By independence,

$$\mathbb{E} \left[e^{-(n \cdot \Pi_{c/n^d})(f)} \right] = \prod_{z \in \mathbb{Z}^d} \left(1 - \frac{c}{n^d} g(z/n) \right),$$

where only finitely many factors are different from 1. Hence, for n large enough,

$$\log \mathbb{E} \left[e^{-(n \cdot \Pi_{c/n^d})(f)} \right] = -\frac{c}{n^d} \sum_{z \in \mathbb{Z}^d} g(z/n) + O(n^{-d}).$$

Indeed, this follows from $\log(1-u) = -u + O(u^2)$, since there are $O(n^d)$ non-zero terms and each quadratic term is $O(n^{-2d})$. By Riemann approximation,

$$\frac{1}{n^d} \sum_{z \in \mathbb{Z}^d} g(z/n) \longrightarrow \int_{\mathbb{R}^d} g(x) dx.$$

We thus obtain

$$\mathbb{E} \left[e^{-(n \cdot \Pi_{c/n^d})(f)} \right] \longrightarrow \exp \left\{ -c \int_{\mathbb{R}^d} (1 - e^{-f(x)}) dx \right\}.$$

By the exponential formula, the right-hand side is the Laplace functional of $\text{PPM}(c \cdot \text{Leb}_{\mathbb{R}^d})$.

The result follows from Proposition 3.1. $\square$

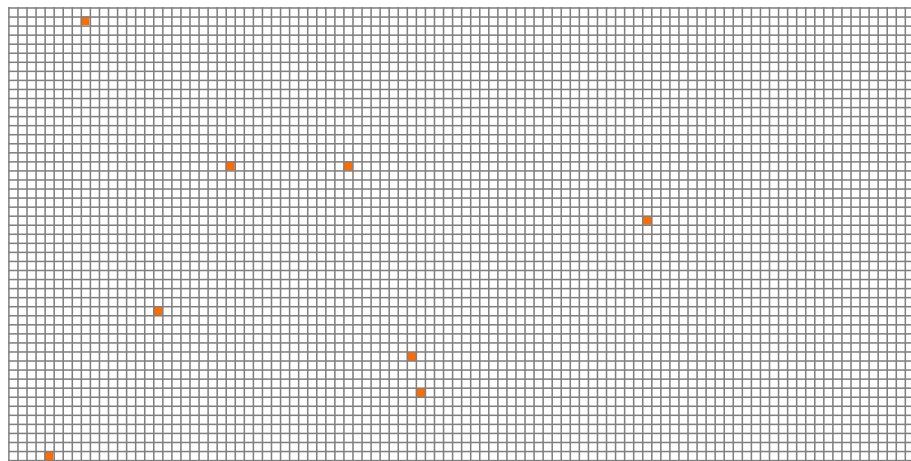


Figure 3.1: Illustration of face percolation on a fine grid of mesh size n^{-1} and percolation parameter c/n^2 ; here $n = 50$ and $c = 2$.

Part II:

Temporal processes

Poissonization

The preceding chapters developed the theory of Poisson random measures on an arbitrary measurable space. We now add a time coordinate and specialize the general theory to the product space

$$\mathbb{R}_+ \times E, \quad \mathbb{R}_+ = [0, \infty).$$

We shall describe such measures using exponential variables in the time parametrization. This forms the basis of the “Poissonization trick”, which embeds many discrete problems into continuous time and allows for easy solutions.



Figure 3.2: An exponential clock

IV: Temporal Poisson processes

The first two chapters developed the theory of Poisson random measures on an arbitrary measurable space. We now add a time coordinate and specialize the general theory to the product space

$$\mathbb{R}_+ \times E, \quad \mathbb{R}_+ = [0, \infty),$$

endowed with the product σ -algebra $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$, where $(E, \mathcal{E})$ is still an arbitrary measurable space. Elements of E will be called *marks*. We write dt for Lebesgue measure on $\mathbb{R}_+$ and let μ be a σ -finite measure on E . We consider $\Pi \sim \text{PPM}(dt \otimes \mu)$. An atom (t, x) of the random measure Π is interpreted as an event occurring at time t and carrying mark x . The time coordinate provides an order, which was absent in the abstract construction. This order leads to arrival times and waiting times, which we shall now describe.

More precisely, if $\mu(E) = \infty$, then for every $0 \leq a < b < \infty$, $\Pi((a, b] \times E) = \infty$ almost surely (Exercise!). If instead $0 < \mu(E) < \infty$, then, under (Δ) , the atoms of Π can almost surely be ranked in increasing chronological order as

$$\Pi = \sum_{i=1}^{\infty} \delta_{(T_i, X_i)}, \quad 0 < T_1 < T_2 < T_3 < \cdots, \quad X_i \in E.$$

For the remainder of this chapter, we suppose that $\Pi \sim \text{PPM}(\mathbb{R}_+ \times E, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}, dt \otimes \mu)$ with $0 < \mu(E) < \infty$ and write

$$\lambda := \mu(E), \quad q := \frac{\mu}{\lambda}.$$

Thus q is a probability measure on $(E, \mathcal{E})$.

4.1 Exponential representation

By the mapping theorem, the projection Π^* of Π onto the time axis is a Poisson random measure with intensity λdt . Since λdt is diffuse, the measure Π^* is simple and so, almost

surely, no atoms of Π share the same time coordinate. Since $\mu(E) < \infty$, Π is almost surely finite on every bounded time slab, meaning that $\Pi([0, t] \times E) < \infty$ for any $t > 0$. It follows from both remarks that we can enumerate the atoms of Π as (T_i, X_i) with strictly increasing time coordinate. We shall describe those times via iid exponential variables. For $\alpha > 0$, we denote by $\mathcal{E}(\alpha)$ the exponential distribution of mean $1/\alpha$, i.e.,

$$\mathcal{E}(\alpha) = \alpha e^{-\alpha x} \mathbf{1}_{x>0} dx.$$

We call α its **rate** or **intensity**. It corresponds to the inverse of its mean. It is easy to see that if $X \sim \mathcal{E}(\alpha)$ and $\beta > 0$ then $\beta X \sim \mathcal{E}(\alpha/\beta)$. If $X \sim \mathcal{E}(\alpha)$, then for every $t \geq 0$, conditionally on $\{X \geq t\}$, the variable $X - t$ again has distribution $\mathcal{E}(\alpha)$. Equivalently,

$$\mathbb{P}(X \in [x, x+h] \mid X \geq x) = 1 - e^{-\alpha h} = \alpha h + o(h), \quad h \downarrow 0. \quad (4.1)$$

Recall also that this property characterizes exponential and geometric laws.

Exercise 4.1 (Memorylessness). Let X be a non-degenerate non-negative random variable.

- (i) Suppose that $\mathbb{P}(X > b) > 0$ for every $b \geq 0$ and that, for every $a, b \geq 0$,

$$\mathbb{P}(X > a + b \mid X > b) = \mathbb{P}(X > a).$$

Show that X has an exponential distribution.

- (ii) Suppose that X takes values in $c\mathbb{N}_{>0}$ for some $c > 0$, that $\mathbb{P}(X > cn) > 0$ for every $n \geq 0$, and that, for every $m, n \geq 0$,

$$\mathbb{P}(X > c(m+n) \mid X > cn) = \mathbb{P}(X > cm).$$

Show that X/c has a geometric distribution on $\mathbb{N}_{>0}$.

Theorem 4.1 (Exponential waiting-time and mark representation)

Write the atoms of Π in chronological order as

$$\Pi = \sum_{i=1}^{\infty} \delta_{(T_i, X_i)}, \quad T_0 := 0,$$

and define the waiting times

$$S_i := T_i - T_{i-1}, \quad i \geq 1.$$

Then $(S_i : i \geq 1)$ are i.i.d. exponential random variables with rate λ , while $(X_i : i \geq 1)$ are i.i.d. with law $q = \mu/\lambda$, and the two sequences are independent.

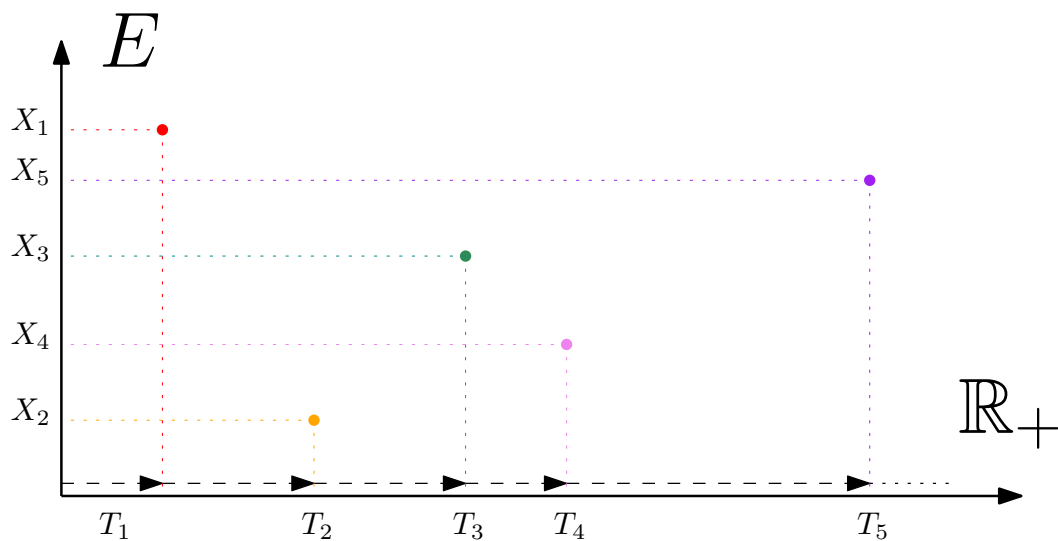


Figure 4.1: Illustration of Theorem 4.1. The interarrival times are independent and exponentially distributed with mean $1/\mu(E)$, whereas the marks are i.i.d. with law $\mu/\mu(E)$ and are independent of the interarrival times.

Remark 4.1 (If one cannot recover the points...). On a completely arbitrary mark space E , one may not be able to recover the marks X_i (see Theorem 1.11 and the (Δ) assumption). Thus, in general, it is safest to formulate the chronological representation as the *existence* of a version carrying an ordered sequence of marks, rather than as a canonical measurable enumeration of the atoms of every element of $\mathcal{N}(\mathbb{R}_+ \times E)$.

Proof. We first identify the law of the chronological time coordinates of Π . By the mapping theorem, the projection Π^* of Π onto the time axis is a Poisson random measure with intensity λdt . Fix $m \geq 1$ and $T > 0$. Conditionally on $\Pi^*([0, T]) = m$ which happens with probability $e^{-\lambda T} (\lambda T)^m / m!$, by the construction of Theorem 1.4 the ordered time coordinates in $[0, T]$ of the atoms of Π^* are the order statistics of m independent uniform random variables on $[0, T]$. Their full joint density is

$$\frac{m!}{T^m} \mathbf{1}_{\{0 < t_1 < \dots < t_m < T\}} dt_1 \dots dt_m.$$

The density of the (random number) of increasing time coordinates on $[0, T]$ is then

$$\sum_{m \geq 0} \frac{m!}{T^m} \mathbf{1}_{\{0 < t_1 < \dots < t_m < T\}} \times e^{-\lambda T} \frac{(\lambda T)^m}{m!} dt_1 \dots dt_m = \sum_{m \geq 0} \lambda^m e^{-\lambda T} \mathbf{1}_{\{0 < t_1 < \dots < t_m < T\}} dt_1 \dots dt_m.$$

On the other hand, let us compute the law of the points $X_1, X_1 + X_2, X_1 + X_2 + X_3, \dots$ obtained by summing iid exponential variables of parameter λ until the sum exceeds T . On the event

that there is m such points their joint density is given by

$$\mathbf{1}_{\{0 < t_1 < \dots < t_m < T\}} \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \dots \lambda e^{-\lambda(t_m - t_{m-1})} \times e^{-\lambda(T - t_m)} dt_1 \dots dt_m.$$

Notice in particular the absence of λ is the last factor accounting for the fact that $X_1 + \dots + X_{m+1} > T$. We recover precisely the above density and so we deduce equality of the two laws. This proves that the waiting times are independent exponential random variables with rate λ .

It remains to identify the marks. By the construction of Theorem 1.4, conditionally on $\Pi^*([0, T]) = m$, the m points of Π in $(0, T] \times E$, before chronological ordering, may be represented by i.i.d. pairs (U_i, Z_i) with law

$$\frac{du}{T} \otimes q(dx).$$

Ordering the pairs by their time coordinates applies a random permutation that depends only on $(U_1, \dots, U_m)$. Since the marks are independent of the times and are i.i.d. with law q , the reordered marks are still i.i.d. with law q and are independent of the ordered time coordinates. Hence, on $\{T_n < T\}$, the first n chronological marks have law $q^{\otimes n}$ and are independent of $(T_1, \dots, T_n)$. $\square$

4.2 Direct applications

Theorem 4.1 gives a process representation of PPMs on $\mathbb{R}_+ \times E$ and is at the core of many important results. Let us first list a few direct applications. The first one is about so-called Markov properties whereas the second one spells out the link between PPM and exponential variables and forms the basis of the Poissonization technique of Chapter 6.

4.2.1 Markov properties

The temporal parametrization enables us to interpret Poisson point measures as random processes. One of the key tools is then the Markov property, which roughly states that the future is independent of the past conditionally on the present. In the case of Poisson point measures, let us give two simple examples of Markov properties without going into detail, since this is not the primary scope of these lecture notes.

Simple Markov property. For $t \geq 0$, define the shifted future random measure

$$\Theta_t \Pi := \sum_{i: T_i > t} \delta_{(T_i - t, X_i)}.$$

Proposition 4.2 (Restart at a deterministic time). *For every deterministic $t \geq 0$, the shifted future $\Theta_t \Pi$ is independent of $\Pi|_{[0,t]}$ and*

$$\Theta_t \Pi \stackrel{d}{=} \Pi.$$

Proof. The restrictions of Π to the disjoint sets

$$[0, t] \times E \quad \text{and} \quad (t, \infty) \times E$$

are independent by Proposition 1.7. The map $(s, x) \mapsto (s - t, x)$ from $(t, \infty) \times E$ to $(0, \infty) \times E$ preserves the intensity measure up to the time translation. The mapping theorem therefore shows that the shifted future is again a Poisson random measure with intensity $ds \otimes \mu$. $\square$

Strong Markov at jump times. The exponential representation gives a particularly transparent form of the Markov property at jump times (which are random).

Proposition 4.3 (Restart after the n -th arrival). *Under the representation of Theorem 4.1, for $n \geq 0$, the randomly shifted measure $\Theta_{T_n} \Pi$ is independent of $\sigma(S_1, \dots, S_n, X_1, \dots, X_n)$ and has the same law as Π .*

Proof. If $E_1, E_2, \dots$ are the iid exponential waiting times and $X_1, \dots, X_n, \dots$ the iid marks from which the measure Π is build as in Theorem 4.1 then the PPM $\Theta_{T_n} \Pi$ is similarly build but from the waiting times $E_{n+1}, E_{n+2}, \dots$ and the marks $X_{n+1}, X_{n+2}, \dots$. Since those remain iid and are independent of $(E_1, \dots, E_n)$ and $(X_1, \dots, X_n)$, the statement of the proposition follows. $\square$

4.2.2 Choosing with clocks

Consider independent variables $X_i \sim \mathcal{E}(\alpha_i)$, $1 \leq i \leq k$, and put

$$\mathcal{M} = \min_{1 \leq i \leq k} X_i, \quad \mathcal{J} = \operatorname{argmin}_{1 \leq i \leq k} X_i.$$

Proposition 4.4 (Choosing with clocks). *The index $\mathcal{J}$ is almost surely well-defined and*

$$\mathcal{M} \sim \mathcal{E}(\alpha_1 + \dots + \alpha_k), \quad \mathbb{P}(\mathcal{J} = i) = \frac{\alpha_i}{\alpha_1 + \dots + \alpha_k}.$$

Moreover, conditionally on $(\mathcal{J}, \mathcal{M})$, the residual clocks

$$(X_i - \mathcal{M} : i \neq \mathcal{J})$$

are independent, with respective laws $\mathcal{E}(\alpha_i)$.

Proof. Let Π be a PPM on $\mathbb{R}_+ \times [0, \alpha_1 + \dots + \alpha_k)$ with intensity $dt \otimes dx$, and split the second coordinate into consecutive intervals I_i of lengths α_i . By restriction, the point measures $\Pi|_{\mathbb{R}_+ \times I_i}$ are independent. By Theorem 4.1, their first arrival times are therefore independent with laws $\mathcal{E}(\alpha_i)$. By Theorem 4.1 again, the first atom of Π has time coordinate $\mathcal{E}(\sum_i \alpha_i)$, while its mark falls in I_i with probability $\alpha_i / \sum_j \alpha_j$. Finally, by Theorem 4.1, after the first atom, the shifted future process, is independent of both the time and the mark of the first atom and has the same law as the original process (see Proposition 4.3). This implies the statement about the residual clocks. $\square$

V: Inhomogeneous PPMs

In this chapter, we focus on a few important constructions with time-inhomogeneous intensities. They arise, in particular, in extreme-value theory and in important biological models.

5.1 Inhomogeneous Poisson point measures

Let $r : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be locally integrable. A useful way to study an inhomogeneous PPM $\Pi \sim \text{PPM}(r(s)ds)$ is to time-change it by the cumulative function

$$A(t) = \int_0^t r(s) ds.$$

The map $t \mapsto A(t)$ transforms it into a vanilla unit-intensity PPM on $[0, A(\infty))$. In particular, the transformed process lives on the whole of $\mathbb{R}_+$ when $A(\infty) = \int_0^\infty r(s) ds = \infty$. Let us list a few examples of such processes that naturally appear in probability:

5.1.1 The Gumbel cloud

The standard Gumbel¹ cloud is the Poisson random measure on $\mathbb{R}$ with intensity

$$e^{-x} dx. \tag{5.1}$$

It has infinitely many points to the left but only finitely many above every fixed level. The

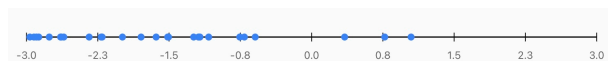


Figure 5.1: A simulation of $\Pi \sim \text{PPM}(e^{-x}dx)$.

change of variable $y = e^{-x}$ transforms (5.1) into Lebesgue measure on $\mathbb{R}_+$. Hence, if its points are written in decreasing order,

$$G_1 > G_2 > \cdots, \quad G_k = -\log \Gamma_k. \tag{5.2}$$



¹ Emil Julius Gumbel (1891–1966), a German mathematician.

In particular

$$\mathbb{P}(G_1 \leq x) = \mathbb{P}(\Pi((x, \infty)) = 0) = \exp(-e^{-x}).$$

Thus the Gumbel distribution is simply the law of the largest point of this inhomogeneous Poisson cloud. It appears in the description of the maxima of large samples of exponentially decaying random variables:

Proposition 5.1 (Maximum of exponentials and the Gumbel distribution). *Let $Y_1, Y_2, \dots$ be i.i.d. $\text{Exp}(1)$ and $M_n = \max_{i \leq n} Y_i$. Then*

$$\sum_{i=1}^n \delta_{Y_i - \log n} \Rightarrow \text{PRM}(e^{-x} dx)$$

on $\mathbb{R}$, and therefore

$$M_n - \log n \Rightarrow G_1, \quad \mathbb{P}(G_1 \leq x) = e^{-e^{-x}}. \quad (5.3)$$

Proof. If $f \in C_c^+(\mathbb{R})$, then for all large n ,

$$\begin{aligned} n \mathbb{E}[1 - e^{-f(Y_1 - \log n)}] &= \int_{-\log n}^{\infty} (1 - e^{-f(x)}) e^{-x} dx \\ &\longrightarrow \int_{\mathbb{R}} (1 - e^{-f(x)}) e^{-x} dx, \end{aligned}$$

and the same Laplace-functional argument applies. Finally, for every $x \in \mathbb{R}$ and all sufficiently large n ,

$$\mathbb{P}(M_n - \log n \leq x) = \left(1 - \frac{e^{-x}}{n}\right)^n \longrightarrow e^{-e^{-x}},$$

which proves (5.3). $\square$

5.1.2 The polynomial cloud

Fix $\alpha \in \mathbb{R}$. The polynomial cloud of parameter α is the PPM on $\mathbb{R}_+$ with intensity

$$t^\alpha dt.$$

If $\alpha > -1$, it is locally finite around 0 and has infinitely many particles towards ∞ . If $\alpha < -1$, it has infinitely many particles accumulating at 0 but only finitely many above every positive level. For $\alpha = -1$, both regions have infinite mass.

As for the Gumbel cloud, it appears in extreme-value theory:

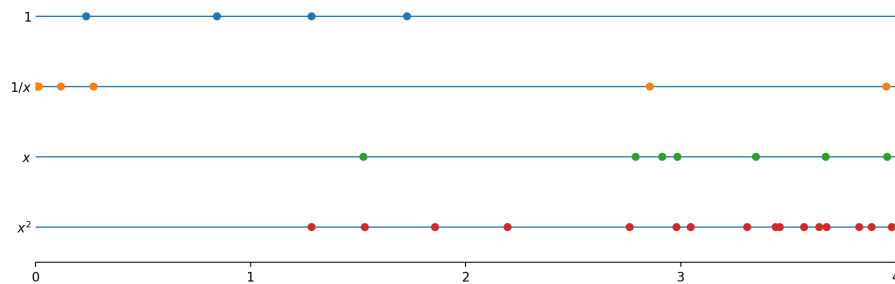


Figure 5.2: A few simulations of polynomial clouds.

Proposition 5.2 (Minima). *Let $X_1, X_2, \dots$ be i.i.d. non-negative variables such that, for some $c, \alpha > 0$,*

$$F(x) = \mathbb{P}(X_1 \leq x) \sim cx^\alpha, \quad x \downarrow 0.$$

With $a_n = (cn)^{1/\alpha}$,

$$\sum_{i=1}^n \delta_{a_n X_i} \Rightarrow \text{PRM}(\alpha x^{\alpha-1} dx).$$

Consequently, if $X_{(1)} = \min_{i \leq n} X_i$,

$$a_n X_{(1)} \Rightarrow W_\alpha, \quad \mathbb{P}(W_\alpha > x) = e^{-x^\alpha}. \quad (5.4)$$

In particular, when $c = \alpha = 1$, $nX_{(1)} \Rightarrow \mathcal{E}(1)$.

Proof. For the minimum, let

$$\nu_n = n \mathcal{L}(a_n X_1).$$

For every $x \geq 0$,

$$\nu_n([0, x]) = nF(x/a_n) \longrightarrow x^\alpha,$$

so ν_n converges vaguely to $\alpha x^{\alpha-1} dx$. Hence, for $f \in C_c^+(\mathbb{R}_+)$,

$$n \mathbb{E}[1 - e^{-f(a_n X_1)}] \longrightarrow \int_0^\infty (1 - e^{-f(x)}) \alpha x^{\alpha-1} dx.$$

Since

$$\mathbb{E}\left[e^{-\sum_{i=1}^n f(a_n X_i)}\right] = \left(1 - \mathbb{E}[1 - e^{-f(a_n X_1)}]\right)^n,$$

the first point-process convergence follows. $\square$

The case $\alpha = 1$ gives in particular

$$cnX_{(1)} \Rightarrow \text{Exp}(1)$$

whenever the density is continuous at 0 with value $c > 0$. This cloud also appears in connection with collisions in the birthday paradox.

5.1.3 The birthday process

We now focus on the polynomial cloud with intensity $t dt$. Let $0 < C_1 < C_2 < \dots$ be the points of a PRM($t dt$) on $\mathbb{R}_+$. Since $A(t) = t^2/2$,

$$C_k = \sqrt{2\Gamma_k}. \quad (5.5)$$

In particular

$$\mathbb{P}(C_1 > t) = e^{-t^2/2},$$

which is known as the Rayleigh law. This process appears in the scaling limit of the birthday paradox: for every n , let $B_1^{(n)}, B_2^{(n)}, \dots$ be i.i.d. uniform on $\{1, \dots, n\}$. Think of a year with n possible dates and label the people in a room by $1, 2, \dots$; their birthdays $B_i^{(n)}$ are independent and uniformly distributed over the year. At time j , a collision occurs when the date $B_j^{(n)}$ has already appeared. Put

$$K_j^{(n)} = \#\{i < j : B_i^{(n)} = B_j^{(n)}\}$$

and define the collision point measure

$$\Xi_n = \sum_{j \geq 2} \mathbf{1}_{\{K_j^{(n)} \geq 1\}} \delta_{j/\sqrt{n}}. \quad (5.6)$$

Proposition 5.3 (Poisson limit for birthday collisions). *As $n \rightarrow \infty$,*

$$\Xi_n \Rightarrow \text{PRM}(t dt)$$

on $\mathbb{R}_+$. In particular, if

$$\tau_n = \min\{j \geq 2 : K_j^{(n)} \geq 1\}$$

is the first collision time, then

$$\frac{\tau_n}{\sqrt{n}} \Rightarrow C_1, \quad \mathbb{P}(C_1 > t) = e^{-t^2/2}. \quad (5.7)$$

Proof. It is slightly cleaner to count colliding pairs and first consider

$$\widehat{\Xi}_n = \sum_{1 \leq i < j} \mathbf{1}_{\{B_i^{(n)} = B_j^{(n)}\}} \delta_{j/\sqrt{n}} = \sum_{j \geq 2} K_j^{(n)} \delta_{j/\sqrt{n}}.$$

For a bounded interval $I = (a, b]$, let

$$\mathcal{E}_n(I) = \{(i, j) : 1 \leq i < j, j/\sqrt{n} \in I\}.$$

Then

$$\frac{|\mathcal{E}_n(I)|}{n} = \frac{1}{n} \sum_{a\sqrt{n} < j \leq b\sqrt{n}} (j-1) \longrightarrow \int_a^b t dt. \quad (5.8)$$

Fix disjoint bounded intervals $I_1, \dots, I_\ell$ and integers $r_1, \dots, r_\ell \geq 0$, and put $r = r_1 + \dots + r_\ell$. Expand the joint factorial moment

$$\mathbb{E} \left[\prod_{q=1}^{\ell} (\widehat{\Xi}_n(I_q))_{r_q} \right]$$

as a sum over r distinct pairs. If the pairs have disjoint endpoints, the r equalities involve disjoint birthdays, have probability n^{-r} , and by (5.8) their contribution tends to

$$\prod_{q=1}^{\ell} \left(\int_{I_q} t \, dt \right)^{r_q}.$$

The other configurations are negligible. Indeed, if their graph has v vertices and c connected components, there are $O(n^{v/2})$ such configurations and the required birthday equalities have probability $n^{-(v-c)}$. A non-matching has $v \geq 2c + 1$, so its total contribution is $O(n^{c-v/2}) = o(1)$. Thus the joint factorial moments converge to those of independent Poisson variables with means $\int_{I_q} t \, dt$, which gives

$$\widehat{\Xi}_n \Rightarrow \text{PRM}(t \, dt).$$

It remains only to forget the multiplicity of a collision. On $[0, T]$,

$$\begin{aligned} 0 \leq \widehat{\Xi}_n([0, T]) - \Xi_n([0, T]) &= \sum_{j \leq T\sqrt{n}} (K_j^{(n)} - 1)_+ \\ &\leq \sum_{1 \leq i < k < j \leq T\sqrt{n}} \mathbf{1}_{\{B_i^{(n)} = B_k^{(n)} = B_j^{(n)}\}}. \end{aligned}$$

The expectation of the last term is

$$\binom{\lfloor T\sqrt{n} \rfloor}{3} \frac{1}{n^2} = O(n^{-1/2}),$$

so Ξ_n and $\widehat{\Xi}_n$ have the same limit on every compact set. This proves the point-process convergence and hence (5.7). Equivalently, for $m = \lfloor t\sqrt{n} \rfloor$,

$$\mathbb{P}(\tau_n > m) = \prod_{k=0}^{m-1} \left(1 - \frac{k}{n} \right) \longrightarrow e^{-t^2/2}.$$

□

5.2 Hawkes process

We end this chapter with the self-exciting point processes introduced by Hawkes². They were initially motivated by event sequences in which the occurrence of an event temporarily increases the probability of observing further events. Applications include neuronal spikes, earthquakes, epidemics and financial transactions.

The goal is to build a simple point process $\mathbf{N} = \sum_{i \geq 1} \delta_{T_i}$ whose intensity, informally for the moment, varies randomly with time and is given by

$$\lambda_t = \left(\nu + \sum_{T_i < t} h(t - T_i) \right)_+, \quad \nu > 0, \quad (5.9)$$

where $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ is locally bounded and càdlàg. A positive value of $h(s)$ describes excitation s units of time after a point, whereas a negative value describes inhibition. The positive part in (5.9) prevents a negative rate.

Construction one point at a time. It is not clear at first sight how to define a process whose rate depends on its own past. To do this, we use one planar Poisson point measure

$$\Pi \sim \text{PRM}(dt dz) \quad \text{on } \mathbb{R}_+^2.$$

Think of its atoms as proposals: an atom (t, z) is accepted when it lies below the current intensity. Starting with no point, set

$$T_1 = \inf\{t > 0 : \Pi(\{t\} \times (0, \nu]) = 1\}.$$

Once $T_1, \dots, T_n$ have been constructed, put, for $t > T_n$,

$$\lambda_t^{[n]} = \left(\nu + \sum_{i=1}^n h(t - T_i) \right)_+$$

and define

$$T_{n+1} = \inf\{t > T_n : \Pi(\{t\} \times (0, \lambda_t^{[n]}) = 1\}. \quad (5.10)$$

In words, after the n th accepted point we continue scanning the same cloud and retain the first proposal below the new graph. The convention $\inf \emptyset = \infty$ is understood.

Proposition 5.4 (The construction works for all times). *When h is locally bounded and càdlàg, the recursive construction either stops after finitely many points, or satisfies*

$$T_n \longrightarrow \infty \quad \text{almost surely.}$$

Consequently $\mathbf{N} = \sum_{n \geq 1} \delta_{T_n}$ is a simple, locally finite point process on the whole of $\mathbb{R}_+$.



² Alan Hawkes (1938–2023), British statistician.

Proof. There is only one possible difficulty: the construction might select infinitely many points before some finite time. Suppose that infinitely many points can accumulate before time $A < \infty$ with positive probability, we will show that this is excluded. Set

$$B = \sup_{0 \leq s \leq A} h(s)_+ < \infty,$$

and note that if n points have already been selected before A , then for any $t \leq A$ the intensity of selection satisfies

$$\lambda_t^{[n]} \leq \nu + \sum_{i=1}^n h(t - T_i)_+ \leq \nu + nB.$$

Thus the next accepted point cannot arrive faster than a point of a Poisson process of rate $\nu + nB$, or equivalently the random time $T_{n+1} - T_n$ stochastically dominates an exponential law of mean $(\nu + nB)^{-1}$. Moreover, the successive waiting times can be coupled simultaneously with independent exponential variables $\mathcal{E}_1, \mathcal{E}_2, \dots$, with respective means $(\nu + iB)^{-1}$, so that

$$T_{i+1} - T_i \geq \mathcal{E}_i, \quad i \geq 1.$$

But

$$\sum_{i \geq 1} \mathcal{E}_i = \infty, \quad a.s.$$

When $B = 0$, this follows directly from the law of large numbers. When $B > 0$, it follows for example from the three-series theorem: the sum of the means diverges as the harmonic series, whereas the sum of the variances is finite. Hence we cannot accumulate infinitely many atoms before time A with positive probability. Since this holds for any $A > 0$, this proves the result. A planar Poisson cloud almost surely has no two points with the same time coordinate, so the resulting point process is simple. $\square$

Excitation and inhibition in pictures. Figure 5.3 uses exactly the preceding restriction rule. With a positive kernel, every event pushes the intensity up and events tend to arrive in bursts. With a negative kernel, every event pushes it down, producing a short refractory period and more regularly spaced events. A kernel with both signs first excites and then inhibits: events may form small bursts, followed by quieter periods. The last row illustrates the growing kernel $h(x) = x$, for which the events become increasingly dense.

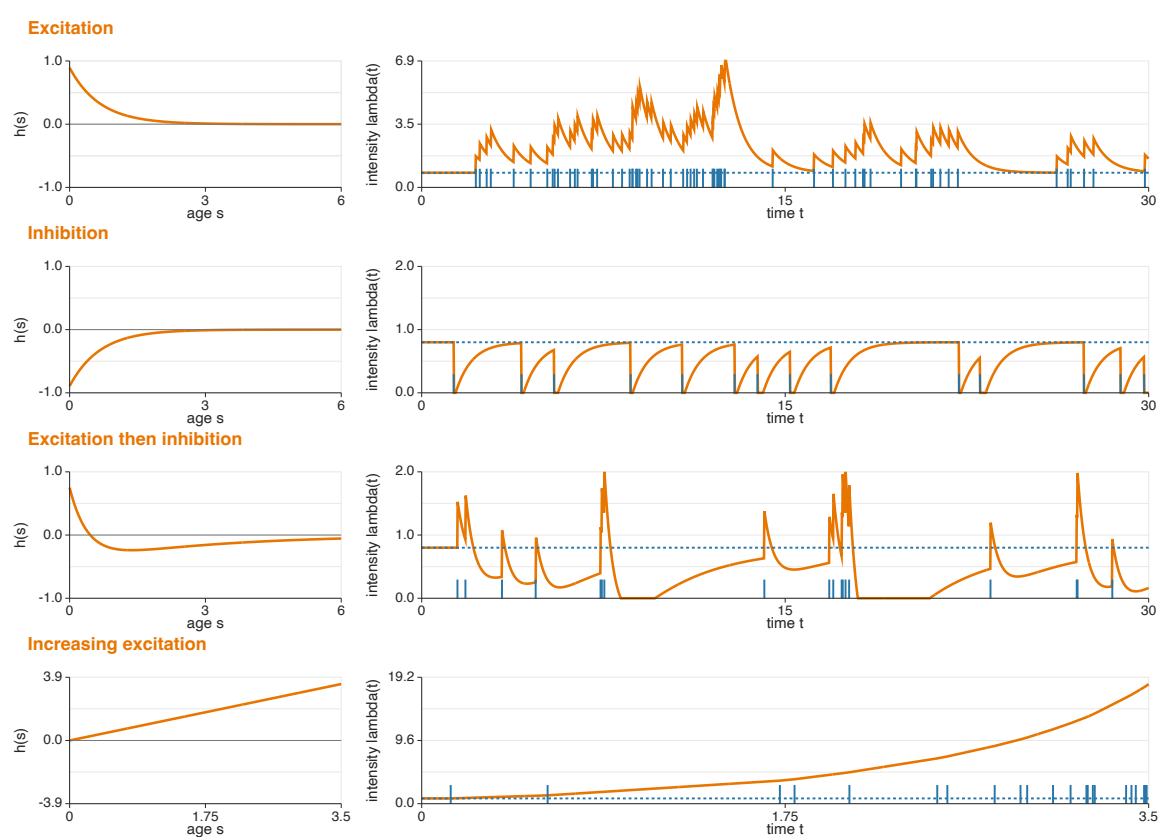


Figure 5.3: The kernel h (left) and one simulated intensity path with its event times (right): pure excitation, pure inhibition, excitation followed by inhibition, and the growing kernel $h(x) = x$. Here $\nu = 0.8$; the first three simulations use $T = 30$, while the rapidly growing last one is shown on $[0, 3.5]$. The horizontal dashed line is the baseline ν .

VI: Poissonization

Randomize to make it simpler!

In this chapter, we develop the Poissonization technique: a discrete-time process is embedded into continuous time by attaching independent exponential clocks to the possible moves, or Poisson streams when the same move can occur repeatedly. The jump chain is unchanged, but the continuous-time picture often reveals useful independence. We shall see this principle at work first without branching, and then for urns and random growing trees.

6.1 An example and a few tools

We start with an example of the idea where we introduce additional randomness which turns a difficult problem into a simpler one about independent variables, to the cost of a global rescaling which is handle using Slutsky's type argument.

6.1.1 Order statistics

The following result was established in the proof of Theorem 4.1; we record it here:

Proposition 6.1 (Order-statistic description). *Let $U_1, U_2, \dots, U_n$ be independent uniform variables on $[0, 1]$ and denote their order statistics by*

$$0 < V_1^{(n)} < \dots < V_n^{(n)} < 1.$$

Also let $\mathcal{E}_1, \dots, \mathcal{E}_{n+1}$ be independent exponential variables of mean 1 and put $S_i = \mathcal{E}_1 + \dots + \mathcal{E}_i$ for $1 \leq i \leq n+1$. Then we have the following equality in distribution:

$$(V_i^{(n)})_{1 \leq i \leq n} \stackrel{(d)}{=} \left(\frac{S_i}{S_{n+1}} \right)_{1 \leq i \leq n}.$$

Proof. We compute the law of $(\frac{S_i}{S_{n+1}})_{1 \leq i \leq n}$ by noticing that

$$\begin{aligned} & \mathbb{E} \left[\phi \left(\left(\frac{S_i}{S_{n+1}} \right)_{1 \leq i \leq n} \right) \right] \\ &= \int dx_1 e^{-x_1} \int dx_2 e^{-x_2} \dots \int dx_{n+1} e^{-x_{n+1}} \phi \left(\frac{x_1}{x_1 + \dots + x_{n+1}}, \frac{x_1 + x_2}{x_1 + \dots + x_{n+1}}, \dots, \frac{x_1 + \dots + x_n}{x_1 + \dots + x_{n+1}} \right). \end{aligned}$$

We first perform the change of variable $s = x_1 + \dots + x_{n+1}$ and then the changes $x_1 \rightarrow v_1, x_1 + x_2 \rightarrow v_2, \dots, x_1 + \dots + x_n \rightarrow v_n$ all of Jacobian 1 to arrive at

$$\int ds e^{-s} s^n \iint dv_1 \dots dv_n \mathbf{1}_{0 < v_1 < \dots < v_n < 1} \phi((v_i)_{1 \leq i \leq n}),$$

which proves the claim. $\square$

This proposition is useful because it transforms intricate random variables into essentially independent ones. For example, one easily deduces interesting information about spacings between uniform variables:

Corollary 6.2. *With the above notation, set*

$$V_0^{(n)} = 0, \quad V_{n+1}^{(n)} = 1, \quad D_i^{(n)} = V_i^{(n)} - V_{i-1}^{(n)}, \quad 1 \leq i \leq n+1.$$

Then

$$n^2 \min_{1 \leq i \leq n+1} D_i^{(n)} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{E},$$

and

$$n \max_{1 \leq i \leq n+1} D_i^{(n)} - \log n \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{G},$$

where $\mathcal{E}$ is an exponential variable of mean one and $\mathcal{G}$ is a Gumbel random variable.

Proof. The preceding representation gives

$$(D_i^{(n)})_{1 \leq i \leq n+1} \stackrel{(d)}{=} \left(\frac{\mathcal{E}_i}{S_{n+1}} \right)_{1 \leq i \leq n+1}.$$

Now, Proposition 4.4 implies $(n+1) \min_{i \leq n+1} \mathcal{E}_i$ has exactly the exponential law. On the other hand, the law of large number entails that $S_{n+1}/n \rightarrow 1$ almost surely, hence in probability. By Slutsky's theorem, it follows that

$$n^2 \min_{1 \leq i \leq n+1} D_i^{(n)} = n^2 \min \left(\frac{\mathcal{E}_i}{S_{n+1}} \right) = \underbrace{\frac{n}{S_{n+1}}}_{\xrightarrow[n \rightarrow \infty]{(\mathbb{P})} 1} \cdot \underbrace{\left((n+1) \min_{i \leq n+1} \mathcal{E}_i \right)}_{\xrightarrow[n \rightarrow \infty]{(d)} \mathcal{E}(1)} \underbrace{\frac{n}{n+1}}_{\xrightarrow[n \rightarrow \infty]{} 1} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{E}(1).$$

For the maximum, for $x \in \mathbb{R}$, if $X \sim \mathcal{E}(1)$ we have

$$\mathbb{P}(\max_{1 \leq i \leq n+1} \mathcal{E}_i \leq \log n + x) = (\mathbb{P}(\mathcal{E} \leq \log n + x))^{n+1} = \left(1 - \frac{e^{-x}}{n} \right)^{n+1} \xrightarrow[n \rightarrow \infty]{} e^{-e^{-x}}.$$

So it follows that

$$\max_{i \leq n+1} \mathcal{E}_i - \log(n+1) \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{G} \stackrel{(d)}{=} -\log \mathcal{E}.$$

We conclude similarly using Slutsky's lemma. $\square$

6.1.2 Two de-Poissonization tools

We record two elementary tools for returning from continuous to discrete time. The first is a probabilistic version of the elementary fact that convergence of monotone functions towards a continuous limit is automatically locally uniform.

Lemma 6.3 (Dini). *Let $(D_t^{(n)} : t \in \mathbb{R})$ be random non-decreasing càdlàg functions. Suppose that for every $t \in \mathbb{R}$,*

$$D_t^{(n)} \xrightarrow[n \rightarrow \infty]{(\mathbb{P})} f(t),$$

where f is continuous. Then $D^{(n)} \rightarrow f$ in probability for the topology of uniform convergence on compact subsets of $\mathbb{R}$.

Proof. Fix $[a, b]$ and $\varepsilon > 0$. Choose a finite partition $a = t_0 < \dots < t_m = b$ such that $|f(t) - f(s)| \leq \varepsilon$ whenever s, t belong to the same subinterval. With probability tending to one, $|D_{t_j}^{(n)} - f(t_j)| \leq \varepsilon$ for all j . Monotonicity then sandwiches $D_t^{(n)}$ between its values at the two neighboring grid points, and yields

$$\sup_{t \in [a, b]} |D_t^{(n)} - f(t)| \leq 2\varepsilon.$$

The same statement of course holds for non-increasing processes. $\square$

Lemma 6.4 (Slutsky). *Suppose that $D^{(n)} \Rightarrow F$ for the topology of uniform convergence on compact sets, and $T_n \Rightarrow \theta$. Assume that F has continuous paths and that either F is deterministic or θ is constant. Then*

$$D_{T_n}^{(n)} \xrightarrow[n \rightarrow \infty]{(d)} F(\theta).$$

Proof. Since one of the two limits is deterministic, Slutsky's lemma gives the joint convergence $(D^{(n)}, T_n) \Rightarrow (F, \theta)$. For the topology of uniform convergence on compact sets, the map $(f, t) \mapsto f(t)$ is continuous at every pair (f, t) for which f is continuous. The conclusion follows from the continuous mapping theorem. $\square$

6.2 Examples without branching

Recall from Section 4.2.2 that, to sample from $\{1, \dots, k\}$ in proportion to weights $\alpha_1, \dots, \alpha_k$, one may attach independent exponential clocks of corresponding rates and choose the first clock that rings. The point is that the clocks which have not rung can be used again after subtracting the elapsed time. We shall repeatedly use this trick to embed discrete-time ran-

dom processes into continuous time. This idea is usually attributed to Athreya & Karlin¹, who systematized it.

The common feature of the next examples is that the collection of objects is fixed from the beginning. Continuous time is only used to make their clocks independent. Branching, where a clock creates new clocks, will appear only in the last section.

6.2.1 A matchbox warm-up

We start with a size-biased version of Banach's² matchbox problem, tailor-made for exponential clocks. One box contains n matches and the other m matches. At each step one of the remaining matches is chosen uniformly and burnt. We ask how many matches are left in the second box when the first one becomes empty.

Attach to every match an independent variable of law $\mathcal{E}(1)$ and burn the matches in increasing order of their clocks. Since the clocks are i.i.d. and almost surely distinct, this produces exactly a uniform ordering of the $n + m$ matches. Write

$$X_1, \dots, X_n, \quad Y_1, \dots, Y_m$$

for the clocks in the two boxes and let

$$M_n = \max_{1 \leq i \leq n} X_i, \quad R_{n,m} = \sum_{j=1}^m \mathbf{1}_{\{Y_j > M_n\}}.$$

Thus $R_{n,m}$ is precisely the number of matches in the second box at the instant when the first box becomes empty; the value 0 simply means that the second box had emptied earlier.

Proposition 6.5 (Matchbox residue). *Suppose that $m/n \rightarrow \rho \in (0, \infty)$. Then*

$$R_{n,m} \xrightarrow[n \rightarrow \infty]{(d)} G_\rho, \quad \mathbb{P}(G_\rho = k) = \frac{1}{1+\rho} \left(\frac{\rho}{1+\rho} \right)^k, \quad k \in \mathbb{Z}_{\geq 0}.$$

In other words, the limiting residue is geometric, with mean ρ .

Proof. Conditionally on M_n ,

$$R_{n,m} \sim \text{Bin}(m, e^{-M_n}).$$



¹ Krishna B. Athreya (1939–2023) Indian



Samuel Karlin (1924–2007) American



² Stefan Banach (1892–1945) Polish. Died from lung cancer.

Moreover e^{-X_1} is uniform on $[0,1]$, and therefore the minimum result for the polynomial cloud, Proposition 5.2 applied with $c = \alpha = 1$, gives

$$ne^{-M_n} = n \min_{1 \leq i \leq n} e^{-X_i} \xrightarrow[n \rightarrow \infty]{(d)} E, \quad E \sim \mathcal{E}(1).$$

Hence, for $s \in [0,1]$,

$$\begin{aligned} \mathbb{E}[s^{R_{n,m}} \mid M_n] &= (1 - (1-s)e^{-M_n})^m \\ &\xrightarrow[n \rightarrow \infty]{(d)} \exp(-(1-s)\rho E). \end{aligned}$$

Taking expectations gives

$$\mathbb{E}[s^{R_{n,m}}] \longrightarrow \mathbb{E}[e^{-(1-s)\rho E}] = \frac{1}{1 + \rho(1-s)},$$

which is the generating function displayed above. $\square$

If instead we stop when the *first* box becomes empty and look at the residue in the other one, let

$$p = \frac{1}{1+\rho}, \quad q = \frac{\rho}{1+\rho}.$$

Conditioning on which box empties first gives the limiting distribution

$$\mathbb{P}(R = k) = qp^k + pq^k, \quad k \geq 1.$$

In particular, in the symmetric case $m = n$ this reduces to the geometric law

$$\mathbb{P}(R = k) = 2^{-k}, \quad k \geq 1.$$

6.2.2 Coupon collector

The famous coupon collector problem is the following. Let $(X_k)_{k \geq 1}$ be i.i.d. uniform on $\{1, \dots, n\}$ and set

$$\mathcal{T}_n := \inf\{k \geq 1 : \{X_1, \dots, X_k\} = \{1, \dots, n\}\}.$$

We ask how many coupons have to be bought before the full collection is obtained.

For this example, and for balls in bins below, it is useful to use the same continuous-time object. Let $(\Pi_i : 1 \leq i \leq n)$ be independent unit-rate PPMs on $\mathbb{R}_+$. An atom of Π_i means that coupon i is selected. Their superposition

$$N_n(t) := \sum_{i=1}^n \Pi_i([0, t])$$

is a Poisson counting process of rate n , and, conditionally on the ordered atom times, the corresponding labels are i.i.d. uniform on $\{1, \dots, n\}$. Hence the labels of the atoms reproduce the original coupon sequence.

Proposition 6.6 (Coupon collector). *We have*

$$\frac{\mathcal{T}_n - n \log n}{n} \xrightarrow[n \rightarrow \infty]{(d)} G,$$

where G has the Gumbel distribution.

Proof. For each coupon let

$$S_i := \inf\{t : \Pi_i([0, t]) \geq 1\}$$

be its first appearance. The S_i are independent $\mathcal{E}(1)$ variables, so the time at which the collection becomes complete is

$$C_n := \max_{1 \leq i \leq n} S_i.$$

By Proposition 5.1 of the preceding chapter,

$$C_n - \log n \Rightarrow G.$$

We now have to come back to discrete time. By construction,

$$\mathcal{T}_n \stackrel{(d)}{=} N_n(C_n).$$

Since C_n and N_n are not independent, we use the de-Poissonization tools above. For convenience, set $N_n(s) = 0$ when $s < 0$ and, for $t \in \mathbb{R}$, put

$$D_t^{(n)} := \frac{N_n(\log n + t) - n \log n}{n}.$$

Lemma 6.7 (Coupon-collector profile). *The processes $D^{(n)}$ converge in probability, locally uniformly on $\mathbb{R}$, towards the identity function $t \mapsto t$.*

Proof. For every fixed t , and all sufficiently large n ,

$$\mathbb{E}[D_t^{(n)}] = t, \quad \text{Var}(D_t^{(n)}) = \frac{\log n + t}{n} \longrightarrow 0.$$

Thus $D_t^{(n)} \rightarrow t$ in probability for every t . Since the processes $D^{(n)}$ are non-decreasing, Lemma 6.3 gives the local uniform convergence. $\square$

We have

$$\frac{\mathcal{T}_n - n \log n}{n} \stackrel{(d)}{=} D_{C_n - \log n}^{(n)}.$$

Lemma 6.7 and the de-Poissonization Lemma 6.4 now give the desired convergence to G . $\square$

6.2.3 The coupon collector's earworm

There is a pleasant variant of the preceding problem. At the instant when the last new coupon is collected, how many times has the most frequent coupon already appeared? This question is sometimes called the *coupon collector's earworm*³. Put

$$\mathcal{A}_n := \max_{1 \leq i \leq n} \#\{1 \leq k \leq \mathcal{T}_n : X_k = i\}.$$

Proposition 6.8 (The earworm). *We have*

$$\frac{\mathcal{A}_n}{\log n} \xrightarrow[n \rightarrow \infty]{(\mathbb{P})} e.$$

Proof. In the preceding embedding, for $t \in \mathbb{R}$, set

$$D_t^{(n)} := \frac{1}{\log n} \max_{1 \leq i \leq n} \Pi_i([0, (\log n + t)_+]).$$

Then

$$\frac{\mathcal{A}_n}{\log n} \stackrel{(d)}{=} D_{C_n - \log n}^{(n)}.$$

Lemma 6.9 (Earworm profile). *The processes $D^{(n)}$ converge in probability, locally uniformly on $\mathbb{R}$, towards the constant function $t \mapsto e$.*

Proof. Fix $t \in \mathbb{R}$ and let $Z_n \sim \text{Poi}(\log n + t)$ for all sufficiently large n . For $\alpha > 1$, Stirling's formula, together with the fact that beyond $\alpha \log n$ the Poisson tail is comparable on the logarithmic scale to its first term, gives

$$\mathbb{P}(Z_n \geq \alpha \log n) = n^{-I(\alpha) + o(1)}, \quad I(\alpha) := \alpha \log \alpha - \alpha + 1.$$

Now I is increasing on $(1, \infty)$ and $I(e) = 1$. Hence, if $\alpha > e$,

$$\mathbb{P}(D_t^{(n)} \geq \alpha) \leq n \mathbb{P}(Z_n \geq \alpha \log n) \longrightarrow 0,$$

whereas if $1 < \alpha < e$, independence gives

$$\mathbb{P}(D_t^{(n)} < \alpha) = (1 - \mathbb{P}(Z_n \geq \alpha \log n))^n \longrightarrow 0.$$

Thus $D_t^{(n)} \rightarrow e$ in probability for every fixed t . Since $t \mapsto D_t^{(n)}$ is non-decreasing, Lemma 6.3 gives the local uniform convergence. $\square$

Proposition 5.1 gives $C_n - \log n \Rightarrow G$. Lemma 6.9 and Lemma 6.4 now give the desired convergence to e . $\square$

³See MathOverflow, Question 215664, *The coupon collector's earworm*.

6.2.4 Balls in bins

Again let $(X_k)_{k \geq 1}$ be i.i.d. uniform on $\{1, \dots, n\}$, now interpreted as balls thrown into n bins. After n throws, write

$$B_i = \#\{1 \leq k \leq n : X_k = i\}, \quad \text{ML}_n = \max_{1 \leq i \leq n} B_i.$$

The question is the size of the most loaded bin.

Proposition 6.10 (Balls in bins). *We have*

$$\text{ML}_n \sim_{\mathbb{P}} \frac{\log n}{\log \log n}.$$

Proof. Keep the same Poisson streams as for the coupon collector and let

$$\tau_n := \inf\{t : N_n(t) = n\}.$$

Since the labels of the first n atoms are i.i.d. uniform,

$$\text{ML}_n \stackrel{(d)}{=} \max_{1 \leq i \leq n} \Pi_i([0, \tau_n]).$$

The total process N_n has rate n , so $\tau_n \rightarrow 1$ in probability.

At a deterministic time $t > 0$, set

$$D_t^{(n)} := \frac{1}{\log n / \log \log n} \max_{1 \leq i \leq n} \Pi_i([0, t]).$$

Lemma 6.11 (Balls-in-bins profile). *The processes $D^{(n)}$ converge in probability, locally uniformly on $(0, \infty)$, towards the constant function $t \mapsto 1$.*

Proof. For fixed $t > 0$, the variables $\Pi_1([0, t]), \dots, \Pi_n([0, t])$ are independent with law $\text{Poi}(t)$. For every integer $r \geq 1$,

$$\mathbb{P} \left(\max_{1 \leq i \leq n} \Pi_i([0, t]) < r \right) = (1 - \mathbb{P}(\text{Poi}(t) \geq r))^n.$$

For fixed t and $r \rightarrow \infty$,

$$\mathbb{P}(\text{Poi}(t) \geq r) \sim e^{-t} \frac{t^r}{r!}.$$

Stirling's formula then shows that, for every $\varepsilon > 0$,

$$\mathbb{P} \left(\max_{1 \leq i \leq n} \Pi_i([0, t]) \leq (1 - \varepsilon) \frac{\log n}{\log \log n} \right) \rightarrow 0$$

and

$$\mathbb{P} \left(\max_{1 \leq i \leq n} \Pi_i([0, t]) \geq (1 + \varepsilon) \frac{\log n}{\log \log n} \right) \rightarrow 0.$$

Thus, at every fixed $t > 0$,

$$D_t^{(n)} \xrightarrow[n \rightarrow \infty]{(\mathbb{P})} 1.$$

Since $t \mapsto D_t^{(n)}$ is non-decreasing, Lemma 6.3 gives the local uniform convergence on $(0, \infty)$. $\square$

We may now evaluate at the dependent random time τ_n :

$$\frac{ML_n}{\log n / \log \log n} \stackrel{(d)}{=} D_{\tau_n}^{(n)}.$$

Lemma 6.11 and Lemma 6.4 now give the desired convergence to 1. $\square$

6.2.5 Pill problem

The pill jar puzzle starts with n whole pills. Each time one pill is chosen uniformly. If a whole pill is chosen, it is broken into two halves, one half is consumed and the other one is returned to the jar. If a half-pill is chosen, it is simply consumed. Let L_n be the number of half-pills remaining just after the last whole pill is selected.

Proposition 6.12 (Pill problem). *We have*

$$\frac{L_n}{\log n} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{E}(1).$$

The continuous-time embedding is particularly simple here and does not require a Poisson point measure. For each initial pill i , take two independent variables

$$X_2^{(i)}, X_1^{(i)} \sim \mathcal{E}(1).$$

The pill stays whole for time $X_2^{(i)}$, then stays as a half-pill for an additional time $X_1^{(i)}$, and finally disappears. By memorylessness, at every instant all pills currently present have independent residual $\mathcal{E}(1)$ clocks. Hence the next pill to change state is uniform among all pills present, exactly as in the discrete puzzle.

Proof. The last whole pill is selected at time

$$C_n := \max_{1 \leq i \leq n} X_2^{(i)}.$$

By Proposition 5.1,

$$T_n := C_n - \log n \xrightarrow[n \rightarrow \infty]{(d)} G.$$

At that instant, pill i is still present as a half-pill precisely when $X_2^{(i)} + X_1^{(i)} > C_n$. Thus, for $t \in \mathbb{R}$, set

$$D_t^{(n)} := \frac{1}{\log n} \sum_{i=1}^n \mathbf{1}_{\{X_2^{(i)} + X_1^{(i)} > \log n + t\}},$$

so that

$$\frac{L_n}{\log n} = D_{T_n}^{(n)}.$$

Lemma 6.13 (Pill profile). *The processes $D^{(n)}$ converge in probability, locally uniformly on $\mathbb{R}$, towards $t \mapsto e^{-t}$.*

Proof. The sum of two independent $\mathcal{E}(1)$ variables has tail

$$\mathbb{P}(X_1 + X_2 > s) = e^{-s}(1 + s).$$

Consequently, for every fixed $t \in \mathbb{R}$,

$$\mathbb{E}[D_t^{(n)}] \longrightarrow e^{-t}, \quad \text{Var}(D_t^{(n)}) \longrightarrow 0.$$

Thus $D_t^{(n)} \rightarrow e^{-t}$ in probability. Since the processes $D^{(n)}$ are non-increasing, Lemma 6.3 gives the local uniform convergence. $\square$

Lemma 6.13 and Lemma 6.4 give convergence to e^{-G} , which has law $\mathcal{E}(1)$. $\square$

6.2.6 O.K. Corral

Imagine two groups of n people facing each other. At each step, one individual is chosen uniformly and shoots one person in the opposite group. The question is how many people remain when one group dies out. This is usually called the O.K. Corral⁴ problem.

Formally, let $(O_1(k), O_2(k))_{k \geq 0}$ start from (n, n) and satisfy

$$\mathbb{P}(O_1(k+1) = i-1, O_2(k+1) = j \mid O_1(k) = i, O_2(k) = j) = \frac{j}{i+j},$$

with the complementary probability $i/(i+j)$ for the second coordinate to decrease. Put

$$\theta_n := \inf\{k \geq 0 : O_1(k) = 0 \text{ or } O_2(k) = 0\}.$$

Theorem 6.14

We have

$$n^{-3/4}(O_1(\theta_n) + O_2(\theta_n)) \xrightarrow[n \rightarrow \infty]{(d)} \left(\frac{8}{3}\right)^{1/4} \sqrt{|\mathcal{N}|},$$

where $\mathcal{N}$ is a standard normal variable.

Proof. Here again exponential variables are enough. For $r \in \{1, 2\}$ and $1 \leq i \leq n$, let

$$X_i^{(r)} \sim \mathcal{E}(1/i)$$

⁴The gunfight at the O.K. Corral took place on October 26, 1881. Four lawmen were facing five outlaws. During that brief battle, three men were killed, three were wounded, two ran away, and one was unharmed.

be independent. Think of lineage r as successively passing through the states

$$n, n-1, \dots, 2, 1, 0,$$

spending time $X_i^{(r)}$ in state i . Thus a particle in state i carries a clock of rate $1/i$ and, when this clock rings, its state decreases by one.

Suppose that the two current states are (i, j) . The active clocks have rates $1/i$ and $1/j$. By Proposition ??, the first clock rings on the first lineage with probability

$$\frac{i^{-1}}{i^{-1} + j^{-1}} = \frac{j}{i+j},$$

which is exactly the transition probability of the O.K.-Corral chain. Thus the discrete chain is the jump chain of these two independent continuous-time lineages.

Let

$$\mathcal{L}_r := \sum_{i=1}^n X_i^{(r)}$$

be the extinction time of lineage r and put $\mathcal{L} = \mathcal{L}_1 \wedge \mathcal{L}_2$. When the first lineage dies, the remaining lifetime of the survivor is

$$|\mathcal{L}_1 - \mathcal{L}_2|.$$

Now

$$\mathbb{E}[\mathcal{L}_1 - \mathcal{L}_2] = 0, \quad \text{Var}(\mathcal{L}_1 - \mathcal{L}_2) = 2 \sum_{i=1}^n i^2 \sim \frac{2}{3} n^3.$$

The Lindeberg central limit theorem therefore gives

$$R_n := n^{-3/2} |\mathcal{L}_1 - \mathcal{L}_2| \xrightarrow[n \rightarrow \infty]{(d)} \sqrt{\frac{2}{3}} |\mathcal{N}|.$$

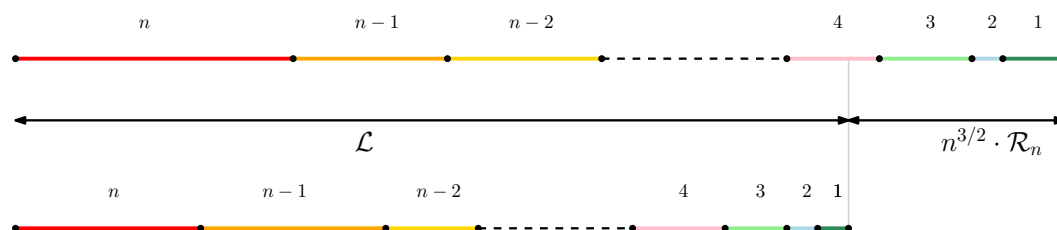


Figure 6.1: The two independent lineages $n \rightarrow n-1 \rightarrow \dots \rightarrow 1 \rightarrow 0$. The state of the lineage still standing when the other dies is read from its remaining lifetime.

It remains to convert a remaining lifetime of order $n^{3/2}$ into a state of order $n^{3/4}$. For one lineage define, for $x \geq 0$,

$$H_x^{(n)} := n^{-3/2} \sum_{i=1}^{\lfloor xn^{3/4} \rfloor} X_i.$$

Lemma 6.15 (O.K.-Corral lifetime profile). *The processes $H^{(n)}$ converge in probability, locally uniformly on $\mathbb{R}_+$, towards $x \mapsto x^2/2$. Consequently, their generalized inverses converge in probability, locally uniformly, towards $t \mapsto \sqrt{2t}$.*

Proof. The variable $H_x^{(n)}$ is the rescaled lifetime of a lineage started from state $\lfloor xn^{3/4} \rfloor$. We have

$$\mathbb{E}[H_x^{(n)}] = n^{-3/2} \sum_{i \leq xn^{3/4}} i \longrightarrow \frac{x^2}{2},$$

and

$$\text{Var}(H_x^{(n)}) = n^{-3} \sum_{i \leq xn^{3/4}} i^2 \longrightarrow 0.$$

Hence $H_x^{(n)} \rightarrow x^2/2$ in probability for every $x \geq 0$, and monotonicity together with Lemma 6.3 upgrades this to local uniform convergence.

Taking generalized inverses, the state of a lineage whose remaining lifetime is $n^{3/2}t$, divided by $n^{3/4}$, converges locally uniformly in probability towards

$$t \mapsto \sqrt{2t}.$$

□

Doing this for the two lineages simultaneously, their two inverse profiles converge in probability towards the same deterministic function $t \mapsto \sqrt{2t}$. At time $\mathcal{L}$, the surviving lineage has remaining time $n^{3/2}R_n$. Hence the choice of which lineage survives is immaterial asymptotically, and Lemma 6.4 applies despite the dependence between the random time and the lineages. We conclude that

$$\begin{aligned} n^{-3/4}(O_1(\theta_n) + O_2(\theta_n)) &\xrightarrow[n \rightarrow \infty]{(d)} \sqrt{2\sqrt{\frac{2}{3}}|\mathcal{N}|} \\ &= \left(\frac{8}{3}\right)^{1/4} \sqrt{|\mathcal{N}|}. \end{aligned}$$

□

6.2.7 Kingman's coalescent: splitting at the root

We finish this first series of examples with a coalescing process. Although the object produced is a random tree, there is no branching in the continuous-time construction below: all clocks can again be put on the table at time 0.

Fix $n \geq 2$ and start from the partition of $\{1, \dots, n\}$ into singletons, called blocks. At every steps choose uniformly a pair of blocks and merge them. After $n - 1$ steps only one block $\{1, 2, \dots, n\}$ remains. The genealogical merging of the blocks gives rise to a rooted binary tree

with leaves $1, \dots, n$, usually called Kingman's⁵ coalescent tree. The question is then to make a sense of the process in the limit $n \rightarrow \infty$.

To answer this question, it is useful to embed the discrete chain in continuous-time which couples the coalescents for all values of n simultaneously. To do this, fix once and for all independent variables

$$(E_{ij} : 1 \leq i < j < \infty), \quad E_{ij} \sim \mathcal{E}(1).$$

For a fixed n , every block is represented by its smallest label. At absolute time E_{ij} , if i and j are representatives of two distinct blocks, merge these two blocks; otherwise do nothing.

Suppose that at some time there are b blocks, with representatives $i_1, \dots, i_b$. For every pair of representatives the corresponding clock has not rung yet. By memorylessness, the residual clocks are independent $\mathcal{E}(1)$ variables. Hence the next merger occurs after a time of law $\mathcal{E}(\binom{b}{2})$ and the pair which merges is uniform among the $\binom{b}{2}$ pairs. The jump chain is therefore exactly the discrete chain above (in particular there is not simultaneous merging with probability 1).

This construction also makes the consistency in n transparent: Construct the process with $n+1$ labels using the same clocks. Since $n+1$ is larger than every old label, a merger involving $n+1$ can add it to an old block but can never change the representative of that block. Consequently, after deleting $n+1$, one recovers pathwise the process constructed with n labels. Thus all the finite coalescents are coupled coherently.

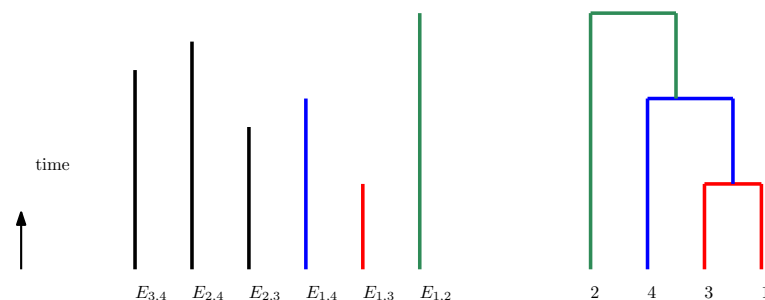


Figure 6.2: Illustration of the construction of Kingman's coalescent from independent exponential clocks.

Furthermore, this construction enables us to start from the "limiting partition of $\mathbb{N}$ into infinitely many singletons":

Proposition 6.16 (Coming down from infinity). *The coherent Kingman coalescent started from the partition of $\mathbb{N}$ into singletons has, almost surely, only finitely many blocks at every time $t > 0$.*



⁵ John Kingman (1939–) British. Also author of [?].

Proof. When there are k blocks, the waiting time before the next merger has law $\mathcal{E}(\binom{k}{2})$, independently of the preceding waiting times. Hence the time θ_n needed to reach complete coalescence when starting from $\{1\}, \{2\}, \dots, \{n\}$ has law

$$\theta_n \stackrel{(d)}{=} \sum_{k=2}^n Z_k, \quad Z_k \sim \mathcal{E}\left(\binom{k}{2}\right)$$

with the variables Z_k independent. In particular, since $\mathbb{E}[\theta_n] = \sum_{k=2}^n 1/\binom{k}{2} \xrightarrow{n \rightarrow \infty} 2$, we deduce that θ_n increases monotonically towards a almost surely finite value θ_∞ . This is the random time it takes to merge into a singleton the infinitely many singletons $\{i\}$ for $i \geq 1$. $\square$

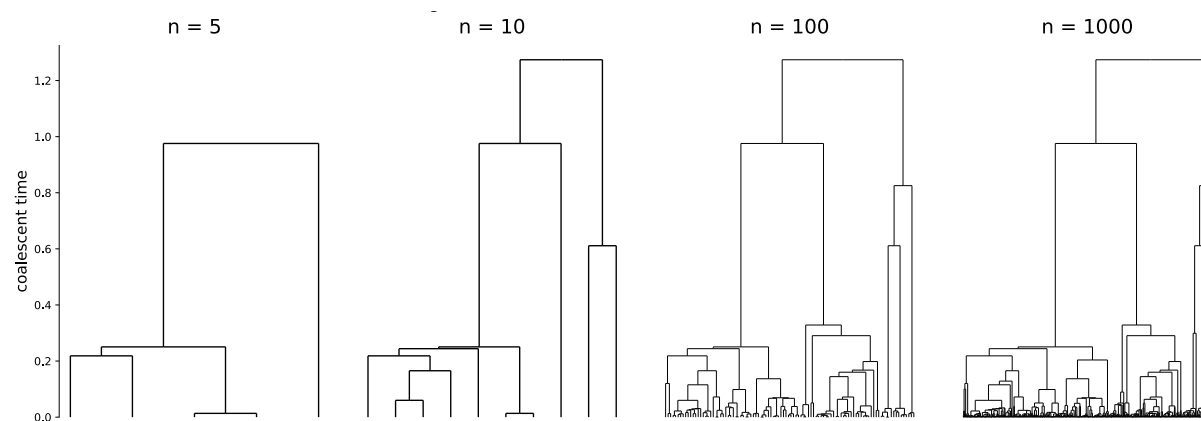


Figure 6.3: A simulation of a coherent Kingman's coalescent for $n = 5, 10, 100$ and 1000, so that the limiting picture is emerging.

We shall now look at Kingman's coalescent tree backwards in time and introduce a related model known as Yule's random tree.

6.3 Branching: Yule and binary search trees

We now turn to a dynamic random tree that underlies many processes with reinforcements such as urns.

6.3.1 From the discrete binary search tree to the Yule tree

The binary search tree is a random process of binary plane trees $(B_n)_{n \geq 0}$. The tree B_0 consists of a single leaf, and, conditionally on B_n , the tree B_{n+1} is obtained by choosing one of the $n+1$ leaves uniformly and grafting a cherry on it.

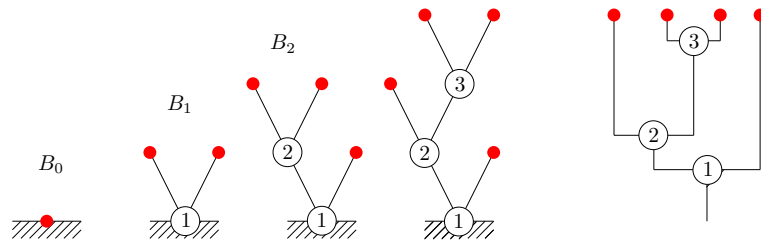


Figure 6.4: Left: the binary search tree growth process. At each step, a leaf is chosen uniformly and replaced by a cherry. Right: its continuous-time evolution, the Yule tree.

To embed this chain into continuous time, start from the infinite binary tree

$$\bigcup_{k \geq 0} \{0,1\}^k$$

and attach to every vertex an independent $\mathcal{E}(1)$ lifetime. The root is alive at time 0; when a leaf dies, it is replaced by its two children, which start their own independent clocks. The resulting random real tree is the standard Yule⁶ tree. We denote by $[\mathbb{T}]_t$ the plane tree alive at time t and by

$$\mathcal{Y}_t := \#\partial[\mathbb{T}]_t$$

its number of leaves.

If there are k leaves, their residual clocks are independent $\mathcal{E}(1)$ variables. Proposition ?? says that the next splitting time is $\mathcal{E}(k)$ and, independently of this time, the leaf which splits is uniform among the k leaves. Let

$$0 = \tau_0 < \tau_1 < \tau_2 < \dots$$

be the successive splitting times. As in the non-explosion argument for the Hawkes process in Section 5.2, but here even more simply, if $(E_k)_{k \geq 1}$ are independent $\mathcal{E}(1)$ variables, then

$$\tau_n = \sum_{k=1}^n \frac{E_k}{k} \xrightarrow[n \rightarrow \infty]{a.s.} \infty$$

by the three-series theorem. Thus $[\mathbb{T}]_t$ is finite for every finite t . Consequently,

$$([\mathbb{T}]_{\tau_n} : n \geq 0) \stackrel{(d)}{=} (B_n : n \geq 0).$$



⁶ George Udny Yule (1871–1951) British

This is the continuous-time embedding of the binary search tree. We shall be interested in the following to the counting process $\mathcal{Y}$. Since Y jumps from k to $k+1$ at rate k , its expectation $m(t) := \mathbb{E}[\mathcal{Y}_t]$ satisfies $m'(t) = m(t)$ and $m(0) = 1$, so

$$\mathbb{E}[\mathcal{Y}_t] = e^t.$$

Combined with the branching Markov property, this shows that

$$(e^{-t}\mathcal{Y}_t : t \geq 0) \text{ is a càdlàg martingale.} \quad (6.1)$$

We now identify its limit.

Proposition 6.17. *We have*

$$e^{-t}\mathcal{Y}_t \xrightarrow[t \rightarrow \infty]{a.s.} W, \quad W \sim \mathcal{E}(1).$$

Proof. We keep the proof based on the jump times, since it makes the relation with the coupon collector particularly transparent. By the discussion above,

$$\tau_i - \tau_{i-1} \sim \mathcal{E}(i), \quad i \geq 1,$$

and these variables are independent. If $h_n = \sum_{i=1}^n 1/i$, then

$$\tau_n - h_n = \sum_{i=1}^n \left((\tau_i - \tau_{i-1}) - \frac{1}{i} \right)$$

is a martingale bounded in L^2 , because

$$\sum_{i \geq 1} \text{Var}(\tau_i - \tau_{i-1}) = \sum_{i \geq 1} \frac{1}{i^2} < \infty.$$

It follows that $\tau_n - h_n$ converges almost surely and in L^2 . Since $h_n - \log n$ converges, there exists a finite random variable X_∞ such that

$$\tau_n - \log n \xrightarrow[n \rightarrow \infty]{a.s.} X_\infty.$$

On the other hand, the vector of spacings

$$(\mathcal{E}(1), \mathcal{E}(2), \dots, \mathcal{E}(n))$$

appears, in reverse order, in the order statistics of n independent $\mathcal{E}(1)$ variables. Hence τ_n has the law of their maximum. By the Gumbel computation used in Proposition 6.6,

$$\tau_n - \log n \xrightarrow[n \rightarrow \infty]{(d)} G.$$

Therefore $X_\infty \stackrel{(d)}{=} G$. Exponentiating the almost sure convergence gives

$$(n+1)e^{-\tau_n} \longrightarrow e^{-X_\infty} \stackrel{(d)}{=} e^{-G} \stackrel{(d)}{=} \mathcal{E}(1).$$

Since $\mathcal{Y}_{\tau_n} = n+1$, this proves the claimed limit along the jump times.

It only remains to pass from jump times to arbitrary times. We have

$$\mathbb{P}\left(\tau_{n+1} - \tau_n \geq n^{-1/2}\right) = \exp\left(-\frac{n+1}{\sqrt{n}}\right),$$

and the right-hand side is summable. By Borel–Cantelli,

$$\tau_{n+1} - \tau_n \longrightarrow 0 \quad \text{a.s.}$$

If $\tau_n \leq t < \tau_{n+1}$, then $\mathcal{Y}_t = n+1$ and

$$e^{-\tau_{n+1}}(n+1) \leq e^{-t}\mathcal{Y}_t \leq e^{-\tau_n}(n+1).$$

The two bounds have the same almost sure limit, which concludes the proof. $\square$

Remark 6.1. In the case of Yule tree of order 2, the distribution of $\mathcal{Y}_t$ is actually explicit:

$$\mathbb{P}(\mathcal{Y}_t = k) = e^{-t}(1 - e^{-t})^{k-1}, \quad k \geq 1. \quad (6.2)$$

Indeed, if $p_k(t) = \mathbb{P}(\mathcal{Y}_t = k)$, the forward equations are

$$p'_k(t) = -kp_k(t) + (k-1)p_{k-1}(t), \quad p_1(t) = e^{-t},$$

and the geometric probabilities in (6.2) solve this system. Taking $t \rightarrow \infty$ in (6.2) also recovers the exponential law of the martingale limit.

6.3.2 Pólya urn and splitting at the root

Let L_n and R_n be the numbers of internal vertices in the left and right subtrees of B_n . Thus

$$L_n + R_n = n - 1, \quad n \geq 1.$$

The left and right subtrees have respectively $L_n + 1$ and $R_n + 1$ leaves. Since the next leaf of B_n is chosen uniformly, conditionally on $(L_n, R_n) = (\ell, r)$ the next split occurs on the left with probability $(\ell + 1)/(n + 1)$ and on the right with probability $(r + 1)/(n + 1)$. Hence the discrete Markov chain

$$(L_n + 1, R_n + 1)_{n \geq 1}$$

is exactly the standard Pólya urn started from (1,1): at each step a ball is chosen uniformly and a new ball of the same color is added.

Proposition 6.18 (Pólya urn and root splitting). *There exists a random variable U , uniform on $[0,1]$, such that*

$$\left(\frac{L_n + 1}{n + 1}, \frac{R_n + 1}{n + 1} \right) \xrightarrow[n \rightarrow \infty]{a.s.} (U, 1 - U).$$

In particular, $L_n/n \rightarrow U$ and $R_n/n \rightarrow 1 - U$ almost surely.

Proof. Use the Yule embedding of the binary search tree. After the root splits, its left and right descendants evolve as two independent Yule trees. If $\mathcal{Y}_t^L$ and $\mathcal{Y}_t^R$ denote their respective numbers of leaves at age t , Proposition 6.17 gives

$$e^{-t}(\mathcal{Y}_t^L, \mathcal{Y}_t^R) \xrightarrow[t \rightarrow \infty]{a.s.} (E, E'),$$

where E and E' are independent $\mathcal{E}(1)$ variables. Hence

$$\frac{\mathcal{Y}_t^L}{\mathcal{Y}_t^L + \mathcal{Y}_t^R} \xrightarrow[t \rightarrow \infty]{a.s.} \frac{E}{E + E'}.$$

The variable $E/(E + E')$ is uniform on $[0,1]$; indeed, for $0 \leq x < 1$,

$$\mathbb{P}\left(\frac{E}{E + E'} \leq x\right) = \int_0^\infty \left(1 - e^{-xy/(1-x)}\right) e^{-y} dy = x.$$

At the n th splitting time τ_n , the Yule tree is B_n , and its left and right subtrees have $L_n + 1$ and $R_n + 1$ leaves. Since $\tau_n - \tau_1 \rightarrow \infty$ almost surely, evaluating the preceding almost sure limit at time $\tau_n - \tau_1$ yields

$$\frac{L_n + 1}{n + 1} \xrightarrow[n \rightarrow \infty]{a.s.} U.$$

Since $L_n + R_n = n - 1$, this proves the result. $\square$

Exercise 6.1. Start instead with $r \geq 1$ red balls and $b \geq 1$ blue balls. Using $r + b$ independent Yule trees, show that the proportion of blue balls converges almost surely to a beta random variable with parameters (b, r) , whose density is

$$\frac{\Gamma(b + r)}{\Gamma(b)\Gamma(r)} x^{b-1}(1-x)^{r-1} \mathbf{1}_{(0,1)}(x) dx.$$

Bibliographical notes. Passing from discrete processes to continuous time in order to gain independence is usually called randomization, Poissonization, or continuous-time embedding; see [?]. Background on the Yule process can be found in [?]. Proposition 6.17 is stated there with an incorrect limiting law; a correction appears in [?, Lemma 3]. The continuous-time embedding of the O.K. Corral model is taken from [?]. The pill problem was studied in [?] using analytic combinatorics. The “earworm” variant of the coupon collector is discussed in MathOverflow Question 215664.

Part III:

Stochastic Geometry

We now use Poisson point measures to generate random geometric constructions on various spaces. We shall mainly use Poisson point measures with constant intensity on $\mathbb{R}^d$ and Poisson point measures on lines in the space of affine lines of $\mathbb{R}^d$. Our goal is not to dive into the detailed properties of those constructions but rather to illustrate the beauty and usefulness of Poisson point measures in stochastic geometry.

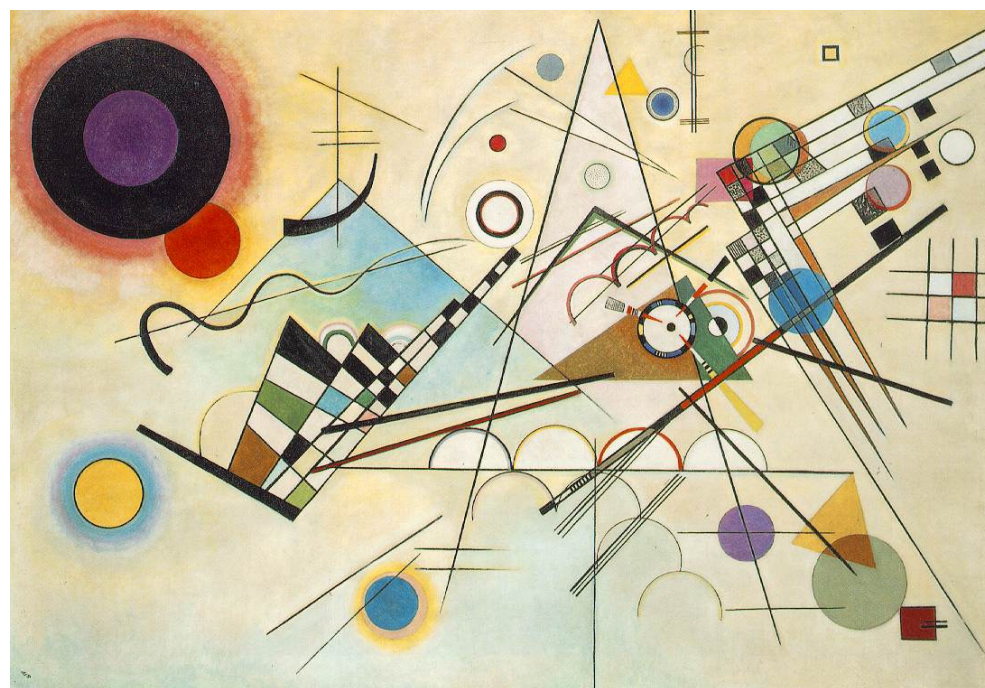


Figure 6.5: Composition VIII by Kandinsky (1923)

VII: Models based on points

This chapter develops several central models of stochastic geometry from a homogeneous Poisson point process in Euclidean space. The underlying random measure is

$$\Pi = \sum_{i \geq 1} \delta_{X_i} \sim \text{PPM}(\lambda \, dx)$$

on $\mathbb{R}^d$, where $\lambda > 0$ and dx denotes Lebesgue measure. The random points X_i are often called *germs* or *nuclei* and the parameter λ is called the **intensity**. We shall survey a few geometric constructions based on Π . Since Lebesgue measure is diffuse, Π is simple almost surely: no point occurs with multiplicity and Π is locally finite.

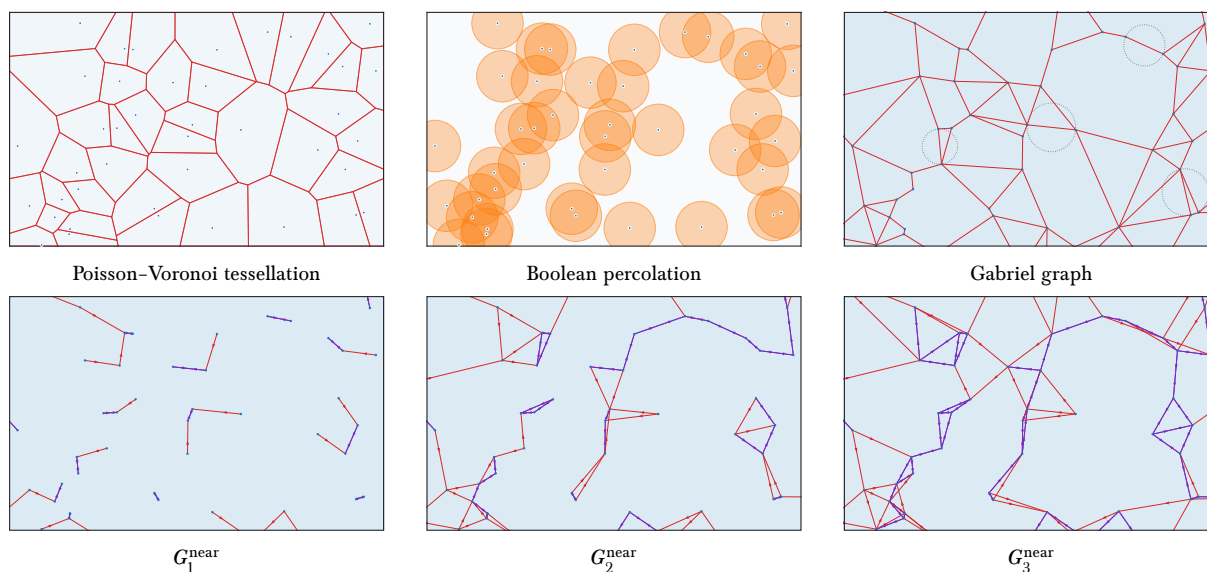


Figure 7.1: Overview of the geometric graphs based on PPM studied in this chapter. All graphs share the same underlying PPM (in blue).

7.1 Basic properties and intensities

We start with basic distributional properties of Π that follow from the mapping theorem and invariance of the Lebesgue measure.

Proposition 7.1 (Stationarity, isotropy and scaling). *If $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a rigid motion¹ then*

$$g_{\#}\Pi \stackrel{d}{=} \Pi.$$

For $a > 0$

$$a \cdot \Pi := (x \mapsto a \cdot x)_{\#}\Pi \sim \text{PPM}\left(\frac{\lambda}{a^d} dx\right).$$

Proof. Recall that Lebesgue measure is invariant under rigid motions and if $S_a(x) = ax$ is a dilation, then for every Borel set A ,

$$(S_a)_{\#}(\lambda dx)(A) = \lambda \text{Vol}_d(a^{-1}A) = \lambda a^{-d} \text{Vol}_d(A).$$

The mapping theorem (Proposition 1.8) proves both assertions. $\square$

We write $B(x, r) = \{y \in \mathbb{R}^d : \|y - x\| \leq r\}$ for the ball of radius r around x and

$$\kappa_d = \text{Vol}_d(B(0, 1)) = \frac{\pi^{d/2}}{\Gamma(1 + d/2)}.$$

In the following we write

$$\Pi = \sum_{i \geq 1} \delta_{X_i},$$

where the atoms X_i are ranked according to their strictly increasing distances from the origin.

Proposition 7.2 (Poisson process of radial volumes). *The point process*

$$\sum_{i \geq 1} \delta_{\lambda \kappa_d \|X_i\|^d}$$

on $\mathbb{R}_+$ is a homogeneous Poisson point process of rate 1. Consequently,

$$\lambda \kappa_d \|X_k\|^d - \lambda \kappa_d \|X_{k-1}\|^d, \quad k \geq 1,$$

with $X_0 = 0$, are independent $\text{Exp}(1)$ variables. Moreover, the directions

$$U_k = \frac{X_k}{\|X_{(k)}\|}$$

are independent, uniform on $\mathbb{S}^{d-1}$, and independent of $(\|X_k\|)_{k \geq 1}$.

¹Every rigid motion of $\mathbb{R}^d$ is the composition of a translation, an orthogonal transformation and possibly a reflection.

Proof. The function $x \in \mathbb{R}^d \mapsto (\|x\|^d, x/\|x\|) \in \mathbb{R}_+ \times \mathbb{S}^{d-1}$ maps $\text{Leb}_{\mathbb{R}^d}$ onto $\kappa_d \cdot \text{Leb}_{\mathbb{R}_+} \times \text{Unif}_{\mathbb{S}^{d-1}}$. The statement then follows from the mapping theorem (Proposition 1.8) together with the temporal representation from the preceding chapter (in particular $i \mapsto \|X_i\|$ is strictly increasing by Proposition 1.10). $\square$

7.1.1 Intensities

Using Π , we shall construct below several random graphs G embedded in $\mathbb{R}^d$. Since these graphs are typically infinite, their total numbers of vertices, edges, or faces are usually $+\infty$; the appropriate quantities are their mean numbers per unit volume. We now make this notion precise.

Let N be a stationary random point measure on $\mathbb{R}^d$, meaning that $\theta_a N$ defined via $\theta_a N(A) := N(A + a)$ has the same law as N for every $a \in \mathbb{R}^d$. Its mean measure is

$$\mu_N(A) := \mathbb{E}[N(A)], \quad A \in \mathcal{B}(\mathbb{R}^d).$$

It is easy to check that μ_N is a measure which is translation invariant. By the standard characterization of translation-invariant Borel measures on $\mathbb{R}^d$, there is a unique constant $i_N \in [0, \infty]$ such that

$$\mu_N = i_N \text{Leb}_{\mathbb{R}^d}.$$

The constant i_N is called the **intensity** of N ; equivalently, for every bounded Borel set A of positive volume,

$$i_N = \frac{\mathbb{E}[N(A)]}{\text{Vol}_d(A)}.$$

For example, in the case of a homogeneous Poisson point process, the intensity agrees with the parameter λ . For a homogeneous Poisson point process, the strong law of large numbers applied to the numbers of points in unit cubes also gives

$$\frac{\Pi([0, R]^d)}{R^d} \xrightarrow[R \rightarrow \infty]{\text{a.s.}} \lambda.$$

Similar empirical limits can often be proved for constructions obtained from Π , but they are not part of the definition of intensity and will be checked separately when needed.

The intensities can be defined for various geometric quantities. For a stationary embedded graph $G = (V, E)$, we can for example define the vertex and edge intensities. To do that, define the following stationary point measures

$$N_V = \sum_{x \in V} \delta_x, \quad N_{V_k} = \sum_{x \in V} \mathbf{1}_{\{\deg_G(x)=k\}} \delta_x, \quad k \geq 0,$$

and denote their intensities by i_V and i_{V_k} . When $0 < i_V < \infty$, then

$$p_k := \frac{i_{V_k}}{i_V}, \quad k \geq 0,$$

defines the **mean degree distribution**: p_k is the mean proportion of vertices having degree k , and $\sum_{k \geq 0} p_k = 1$.

To define the edge intensity, we need similarly to define a point measure. To do that, we attach to every unoriented edge $\{x, y\}$ a translation-covariant representative, for instance its midpoint, and consider the point measure

$$N_E = \sum_{\{x, y\} \in E} \delta_{(x+y)/2}.$$

Thus every unoriented edge is counted exactly once. Counting the two endpoints of every edge gives

$$\text{mean degree} = \sum_{k \geq 0} k i_{V_k} = 2i_E.$$

Face intensities of a stationary planar tessellation are defined similarly, using a translation-covariant center for each bounded face (e.g. its center of gravity).

7.2 Neighbor models

Given Π , we shall build various random graphs whose vertices are the atoms of Π and whose edges are determined by interpoint distances. The simplest is the Gilbert graph (or Boolean model): for a parameter $r > 0$, two distinct points $x, y \in \Pi$ are adjacent if and only if $\|x - y\| \leq r$.

7.2.1 Boolean percolation

We now state (without proof) the classic Boolean percolation result. For $r > 0$, let $G_r^{\text{Bool}}(\Pi)$ join two Poisson points whenever their Euclidean distance is at most r . Equivalently, its connected components are those of the Boolean set

$$\mathcal{B}_r = \bigcup_{i \geq 1} B(X_i, r/2).$$

Theorem 7.3 (Gilbert–Boolean phase transition)

If $d \geq 2$, there is a constant $0 < r_c(\lambda, d) < \infty$ such that, almost surely,

- (i) *if $r < r_c(\lambda, d)$, every component of $G_r^{\text{Bool}}(\Pi)$ is bounded;*

(ii) if $r > r_c(\lambda, d)$, there is a unique unbounded component.

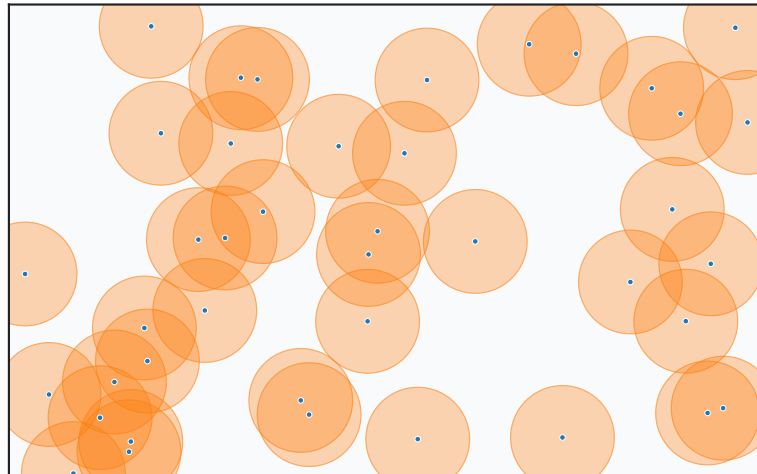


Figure 7.2: The Boolean set $\mathcal{B}_r$.

Exercise 7.1. Show that $r_c(\lambda, d) = \lambda^{-1/d} r_c(1, d)$.

This is a typical percolation result: A zero-one law² for homogeneous Poisson processes shows that the probability of percolation is either zero or one. Monotonicity in r gives a critical value, and Peirls or first-moment type argument shows that it is nontrivial when $d \geq 2$. The exact value is not known, even in the plane. The behavior exactly at criticality is a subtler question which is not yet understood in low dimensions.

Proposition 7.4 (Geometric intensities of the Boolean model). *Put $m = \lambda \kappa_d r^d$. The vertex intensity of $G_r^{\text{Bool}}(\Pi)$ is λ and the mean degree distribution is*

$$p_k = i_{V_k} / i_V = e^{-m} \frac{m^k}{k!}, \quad \text{for } k \geq 0.$$

The intensity of unoriented edges (represented by their midpoints) is

$$i_E = \frac{\lambda^2}{2} \kappa_d r^d = \frac{\lambda m}{2}.$$

Proof. The vertex process is Π , so its intensity is λ . For any bounded Borel set $A \subset \mathbb{R}^d$, the

²after all, the PPM Π can be written as a function of iid random variables. We can even code Π from valid $(U_i)_{i \geq 1}$ so that if we change finitely many U_i 's it only perturb Π on finitely many points. Represent Π by $(U_{i,j})_{i,j \geq 1}$ and use the variables $U_{i_0,j}$ to code for Π on a given region.

Slivnyak–Mecke formula, Theorem 2.3, gives

$$\begin{aligned} i_{V_k} \text{Vol}_d(A) &= \mathbb{E} \left[\int_A d\Pi(x) \mathbf{1}_{\{\deg_{G_r^{\text{Bool}}}(x)=k\}} \right] \\ &\stackrel{\text{Mecke}}{=} \lambda \int_A \mathbb{P}(\Pi(B(x,r)) = k) \, dx \\ &= \lambda \text{Vol}_d(A) e^{-m} \frac{m^k}{k!}. \end{aligned}$$

Indeed, after the distinguished point x is inserted, its neighbors are exactly the points of the independent process Π lying in $B(x,r)$. That number has the $\text{Pois}(m)$ distribution. Dividing by $\text{Vol}_d(A)$ proves the formula for i_{V_k} .

For the edges, count each unordered pair once and select its midpoint. The two-point Mecke formula, Theorem 2.4, yields

$$\mathbb{E}[N_E(A)] = \frac{\lambda^2}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{\|x-y\| \leq r\}} \mathbf{1}_{\{(x+y)/2 \in A\}} \, dx \, dy.$$

With $z = (x+y)/2$ and $h = y-x$, the transformation $(x,y) \mapsto (z,h)$ has Jacobian one. Hence the last expression equals

$$\frac{\lambda^2}{2} \int_A dz \int_{B(0,r)} dh = \frac{\lambda^2 \kappa_d r^d}{2} \text{Vol}_d(A),$$

which identifies the edge intensity. $\square$

In the above case, it is easy to prove the the empirical counts in $[0,L]^d$ converge after normalization towards the mean intensities. Namely, consider $N_{V_k}([0,L]^d)$ the number of vertices of degree k in the box $[0,L]^d$. Mecke formula shows that

$$\mathbb{E}[N_{V_k}([0,L]^d)] = \int_{[0,L]^d} dx \, \mathbb{P}(x \text{ has } k \text{ neighbors in } \Pi) = L^d \lambda e^{-m} \frac{m^k}{k!}.$$

Similarly the multi-point Mecke formula shows that

$$\begin{aligned} \mathbb{E}[N_{V_k}([0,L]^d)^2 - N_{V_k}([0,L]^d)] &= \mathbb{E} \left[\sum_{i \neq j} \mathbf{1}_{X_i, X_j \in [0,L]^d \text{ and have degree } k} \right] \\ &\stackrel{\text{Mecke}}{=} \iint_{[0,L]^d} dx dy \, \mathbb{P}(x,y \text{ have } k \text{ neighbors in } \Pi). \end{aligned}$$

Notice that that as soon as $\|x-y\| \geq 2r$ the events $\{x \text{ has } k \text{ neighbors in } \Pi\}$ and $\{y \text{ has } k \text{ neighbors in } \Pi\}$ are independent since they depend on the balls of radius r around each point. In particular, up to an error of order L^d we have

$$\iint_{[0,L]^d} dx dy \, \mathbb{P}(x,y \text{ have } k \text{ neighbors in } \Pi) = \left(\int_{[0,L]^d} dx \mathbb{P}(x \text{ has } k \text{ neighbors in } \Pi) \right)^2 + O(L^d).$$

This implies that the variance of $N_{V_k}([0, L]^d)$ is of order L^2 and so by Markov's inequality that

$$\mathbb{P}\left(\left|\frac{N_{V_k}([0, L]^d)}{L^d} - \lambda e^{-m} \frac{m^k}{k!}\right| \geq \varepsilon\right) \leq \frac{\text{Var}(N_{V_k}([0, L]^d))}{\varepsilon^2 L^d} \xrightarrow{L \rightarrow \infty} 0.$$

7.2.2 The k -nearest-neighbor graph

We can build another natural graph from Π without the need of an extra length parameter. Fix $k \geq 1$. For each point $x \in \Pi$, draw a directed edge from x to each of its k nearest other points. Ties occur with probability zero, see below. The resulting directed graph is denoted by $G_k^{\text{near}}(\Pi)$. Figure 7.3 shows the first three directed nearest-neighbor graphs on the same point set.

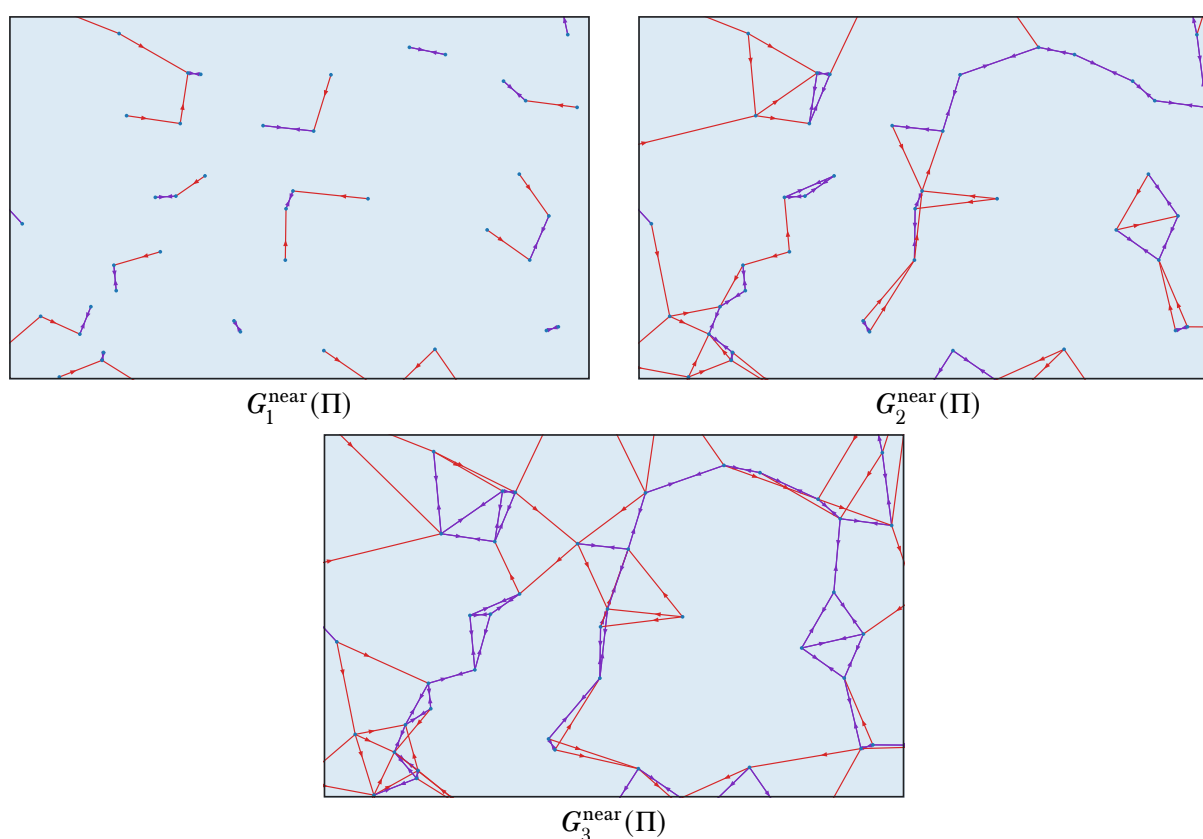


Figure 7.3: The directed graphs $G_1^{\text{near}}(\Pi)$, $G_2^{\text{near}}(\Pi)$, and $G_3^{\text{near}}(\Pi)$, using the same visible points as Figure 7.2. Arrows indicate the nearest neighbors. Reciprocal edges are purple.

The law of $G_k^{\text{near}}(\Pi)$ does not depend up, to a deterministic dilation, on λ which we fix arbitrarily to 1 in this subsection. Despite its basic definition, many properties of this random graph are still elusive. For $k = 1$, we will see below that almost surely every component is

finite; following directed edges eventually reaches a unique pair of mutually nearest neighbors whose density is explicit. In contrast, it is known that when k is sufficiently large ($k \geq 11$ works but $k \geq 3$ should suffice in dimension $d = 2$) the graph percolates, i.e. it has infinite components (a unique one). Before diving into probability, remark that for any $d \geq 1$ a given vertex x of G_k^{near} cannot have more than C_d neighbors pointing to him for some $C_d > 0$. The solution to this riddle is given in Figure 7.4.

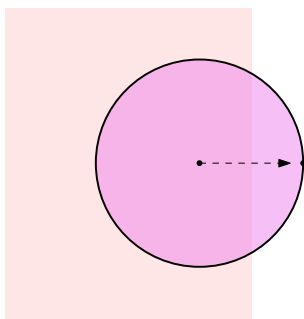


Figure 7.4: Once a point has the origin as nearest neighbor, the purple region must be empty, and any point falling in the red part cannot have the origin as nearest neighbor.

Let us first isolate a useful lemma that rules some configurations in a PPM with constant intensity on $\mathbb{R}^d$:

Lemma 7.5 (No ties, no descending chains). *Almost surely, there are not ties for the pairwise distances, and there is no infinite sequence of atoms $X_{i_0}, X_{i_2}, \dots$ of Π such that*

$$\|X_{i_{k+1}} - X_{i_k}\| \text{ is decreasing for } k \geq 0.$$

Proof. *No ties.* We use multi-point Mecke's formula to show that the expected number of triplet (i, j, k) such that $\|X_i - X_j\| = \|X_j - X_k\|$ is zero. Indeed,

$$\mathbb{E} \left[\sum_{i,j,k \text{ distincts}} \mathbf{1}_{\|X_i - X_j\| = \|X_j - X_k\|} \right] = \iint dx dy dz \mathbf{1}_{\|x-y\| = \|y-z\|} = 0.$$

We obtain the same conclusion for quadruplets (i, j, k, ℓ) such that $\|X_i - X_j\| = \|X_k - X_\ell\|$. We deduce that there are no ties in distances.

No descending chains. For a bounded Borel set $A \subset \mathbb{R}^d$, $R > 0$, and $n \geq 1$, let

$$D_n(A, R) = \sum_{\substack{x_0, \dots, x_n \text{ atoms of } \Pi \\ \text{all distinct}}} \mathbf{1}_{\{x_0 \in A\}} \mathbf{1}_{\{R \geq \|x_1 - x_0\| > \dots > \|x_n - x_{n-1}\|\}}.$$

The multi-point Mecke formula, followed by the change of variables $h_i = x_i - x_{i-1}$, gives

$$\begin{aligned}\mathbb{E}[D_n(A, R)] &= \lambda^{n+1} \text{Vol}_d(A) \int_{R \geq \|h_1\| > \dots > \|h_n\|} dh_1 \cdots dh_n \\ &= \lambda^{n+1} \text{Vol}_d(A) \frac{(\kappa_d R^d)^n}{n!}.\end{aligned}$$

For fixed A and R , this expectation tends to zero as $n \rightarrow \infty$. If an infinite descending chain started in A with first step at most R , then $D_n(A, R) \geq 1$ for every n . Markov's inequality, followed by a countable union over $A = [-m, m]^d$ and $R = m$, $m \geq 1$, proves the claim. $\square$

Proposition 7.6 (Connected components of the one-nearest-neighbor graph). *For every $d \geq 1$, all connected components of $G_1^{\text{near}}(\Pi)$ are finite almost surely. After directions are forgotten, each component is a tree with a distinguished reciprocal edge: its two endpoints are each other's nearest neighbors. For $d = 2$, the intensity of those reciprocal edges is*

$$\lambda p_{\leftrightarrow} \quad \text{where } p_{\leftrightarrow} = \frac{6\pi}{8\pi + 3\sqrt{3}} \approx 0.6215.$$

Proof. Let us first prove that for $d = k = 1$ then all components are finite. Lemma 7.5 already rules out the infinite forward chains. We deduce that all forward path eventually reach a cycle, which must be a reciprocal edge.

It remains to exclude an infinite tree feeding a reciprocal pair. This is done using the mass-transport identity. Imagine that each point of Π has a unit mass of money. Each point then sends and share equally its money among the two endpoint of the reciprocal pair in its component (which exists a.s. by the above discussion). Clearly if there is an infinite backward chain, then a reciprocal edge receives an infinite amount of money. We will see that this is excluded because of the mass-transport principle which says that the average amount of money entering a given region of $\mathbb{R}^d$ is equal to the amount of money exiting this region. Formally if x is not at the extremity of a reciprocal edge, let $H(x, y, \Pi)$ to be 1/2 when y is an endpoint of the reciprocal edge of the cluster of x and 0 otherwise. If x at the extremity of a reciprocal edge then $H(x, y, \Pi) = 1$ only if y is the other extremity. Interpreting H as a quantity of money that x sends to y , we have that a.s. the quantity of money send by any atom x is equal to 1. By Mecke's formula we have

$$\begin{aligned}\mathbb{E}[\text{money received in } A] &= \mathbb{E} \left[\sum_{i \neq j} H(X_i, X_j, \Pi) \mathbf{1}_{X_j \in A} \right] \\ &\stackrel{\text{Mecke}}{=} \int_{\mathbb{R}^d} dx \int_A dy \mathbb{E}[H(x, y, \Pi + \delta_x + \delta_y)].\end{aligned}$$

Notice then that by invariance of Π under rotation and translation $\mathbb{P}(H(x, y, \Pi + \delta_x + \delta_y) = f(\|x - y\|))$. Exchanging the variables and reverse-engineering the above calculations shows that

$$\begin{aligned} \int_{\mathbb{R}^d} dx \int_A dy \mathbb{E}[H(x, y, \Pi + \delta_x + \delta_y)] &= \int_{\mathbb{R}^d} dy \int_A dx \mathbb{E}[H(x, y, \Pi + \delta_x + \delta_y)] \\ &\stackrel{\text{Mecke}}{=} \mathbb{E} \left[\sum_{i \neq j} H(X_i, X_j, \Pi) \mathbf{1}_{X_i \in A} \right] \\ &= \mathbb{E}[\text{money sent from } A]. \end{aligned}$$

This is the mass-transport principle. Notice $\mathbb{E}[\text{money sent from } A] \leq \mathbb{E}[\Pi(A)] = \text{Leb}_{\mathbb{R}^d}(A) < \infty$ so that it is impossible for an infinite backward cluster to start from A .

We finally compute the planar reciprocal-edge intensity. First, we do an area calculation. The union of two disks of radius 1 whose centers are at distance 1 has area

$$\mathfrak{O} := \text{Vol}_2(B(0, 1) \cup B(e_1, 1)) = 2\pi - \text{Overlap},$$

where the lens-shaped intersection (lunule) has area

$$\text{Overlap} = 4 \int_{1/2}^1 \sqrt{1 - r^2} dr = 4 \int_0^{\pi/3} \sin^2(\theta) d\theta = \frac{2\pi}{3} - \frac{\sqrt{3}}{2}.$$

Consequently,

$$\mathfrak{O} = \frac{4\pi}{3} + \frac{\sqrt{3}}{2} = \frac{8\pi + 3\sqrt{3}}{6}.$$

Two points x, y at distance r are mutual nearest neighbors precisely when the union $B(x, r) \cup B(y, r)$ contains no other point of Π . Its area is $r^2 \mathfrak{O}$ by scaling, so the Poisson void probability is

$$\exp(-\lambda r^2 \mathfrak{O}).$$

Applying the two-point Mecke formula to the ordered pair (x, y) gives

$$\mathbb{E}[N_{\leftrightarrow}(A)] = \lambda^2 \iint \mathbf{1}_{\{(x+y)/2 \in A\}} \exp(-\lambda \|x - y\|^2 \mathfrak{O}) dx dy.$$

Performing the change of variables $z = (x + y)/2$ and $h = y - x$, whose Jacobian is one, and then passing to polar coordinates for h , the last integral becomes

$$\lambda^2 \text{Vol}_2(A) 2\pi \int_0^\infty r \exp(-\lambda r^2 \mathfrak{O}) dr.$$

Finally, the substitution $u = \lambda r^2 \mathfrak{O}$ gives

$$\int_0^\infty r \exp(-\lambda r^2 \mathfrak{O}) dr = \frac{1}{2\lambda \mathfrak{O}}.$$

Therefore

$$\mathbb{E}[N_{\leftrightarrow}(A)] = \lambda \frac{\pi}{\mathbb{O}} \text{Vol}_2(A) = \lambda \frac{6\pi}{8\pi + 3\sqrt{3}} \text{Vol}_2(A).$$

□

7.2.3 The Gabriel graph

Closely related to the previous graph, one can also consider the *Gabriel graph* $G^{\text{Gab}}(\Pi)$ whose vertex set are the atoms of Π and where two atoms x, y are joined when the ball of diameter $[x, y]$ has no other point of Π . See Figure 7.5 for a simulation. Every nearest-neighbor edge is a Gabriel edge so that $G_1^{\text{near}}(\Pi) \subset G^{\text{Gab}}(\Pi)$ in terms of unoriented graphs. We leave the following exercise to the reader:

Exercise 7.2. The Gabriel graph of a locally finite set of points in $\mathbb{R}^2$ is a planar graph, but not necessarily connected. If there is no descending chain in the set of points, then its Gabriel graph is connected.

In particular, it follows from Lemma 7.5 that the Gabriel graph of Π is almost surely connected. We compute below its edges intensity:

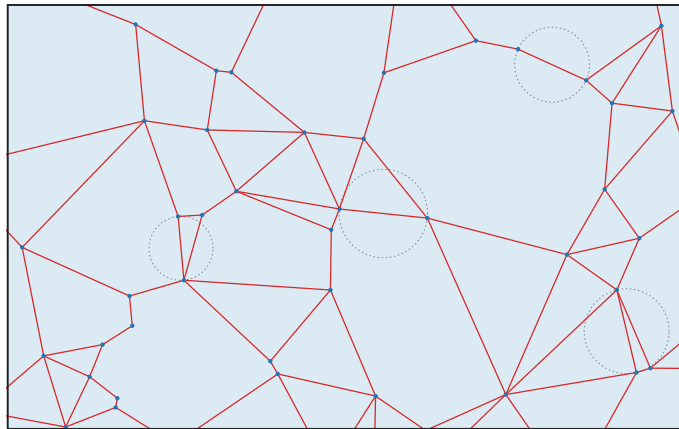


Figure 7.5: Gabriel graph of Figure ??; edges to undisplayed exterior points are clipped at the frame, and dotted circles show four empty diameter balls.

Proposition 7.7 (Degree and edge intensities of the Gabriel graph). *The mean degree of $G^{\text{Gab}}(\Pi)$ is 2^d .*

Proof. The vertex intensity is clearly λ . Let us compute its edge intensity. Let $A \subset \mathbb{R}^d$ be a

bounded Borel set. By the two-point Mecke formula and the avoidance probability,

$$\begin{aligned}\mathbb{E}[N_{\deg}^{\text{Gab}}(A)] &= \lambda^2 \int_A dx \int_{\mathbb{R}^d} dy \mathbb{P}\left(\Pi\left(\dot{B}\left(\frac{x+y}{2}, \frac{\|x-y\|}{2}\right)\right) = 0\right) \\ &= \lambda^2 \int_A dx \int_{\mathbb{R}^d} dy \exp\left(-\frac{\lambda \kappa_d}{2^d} \|x-y\|^d\right).\end{aligned}$$

Writing $h = y - x$ and using polar coordinates gives

$$\begin{aligned}\int_{\mathbb{R}^d} \exp\left(-\frac{\lambda \kappa_d}{2^d} \|h\|^d\right) dh &= d \kappa_d \int_0^\infty r^{d-1} \exp\left(-\frac{\lambda \kappa_d}{2^d} r^d\right) dr \\ &= \frac{2^d}{\lambda}.\end{aligned}$$

Hence

$$\mathbb{E}[N_{\deg}^{\text{Gab}}(A)] = 2^d \lambda \text{Vol}_d(A),$$

which proves the degree-intensity formula. Dividing by the vertex intensity λ gives the mean degree. $\square$

7.3 Poisson–Voronoi tessellations

We now attach to every atom of Π its region of points for which it is a nearest nucleus. These regions form the Poisson–Voronoi tessellation.

7.3.1 Definition and elementary geometry

Let $\varphi \subset \mathbb{R}^d$ be locally finite. For $x \in \varphi$, define its Voronoi cell by

$$V(x, \varphi) = \left\{y \in \mathbb{R}^d : \|y - x\| \leq \|y - z\| \text{ for every } z \in \varphi\right\}.$$

For the Poisson process Π , the collection

$$\{V(x, \Pi) : x \in \Pi\}$$

is the *Poisson–Voronoi tessellation*.

Lemma 7.8 (Half-space representation). *For distinct $x, z \in \mathbb{R}^d$,*

$$\|y - x\| \leq \|y - z\|$$

is equivalent to

$$2y \cdot (z - x) \leq \|z\|^2 - \|x\|^2.$$