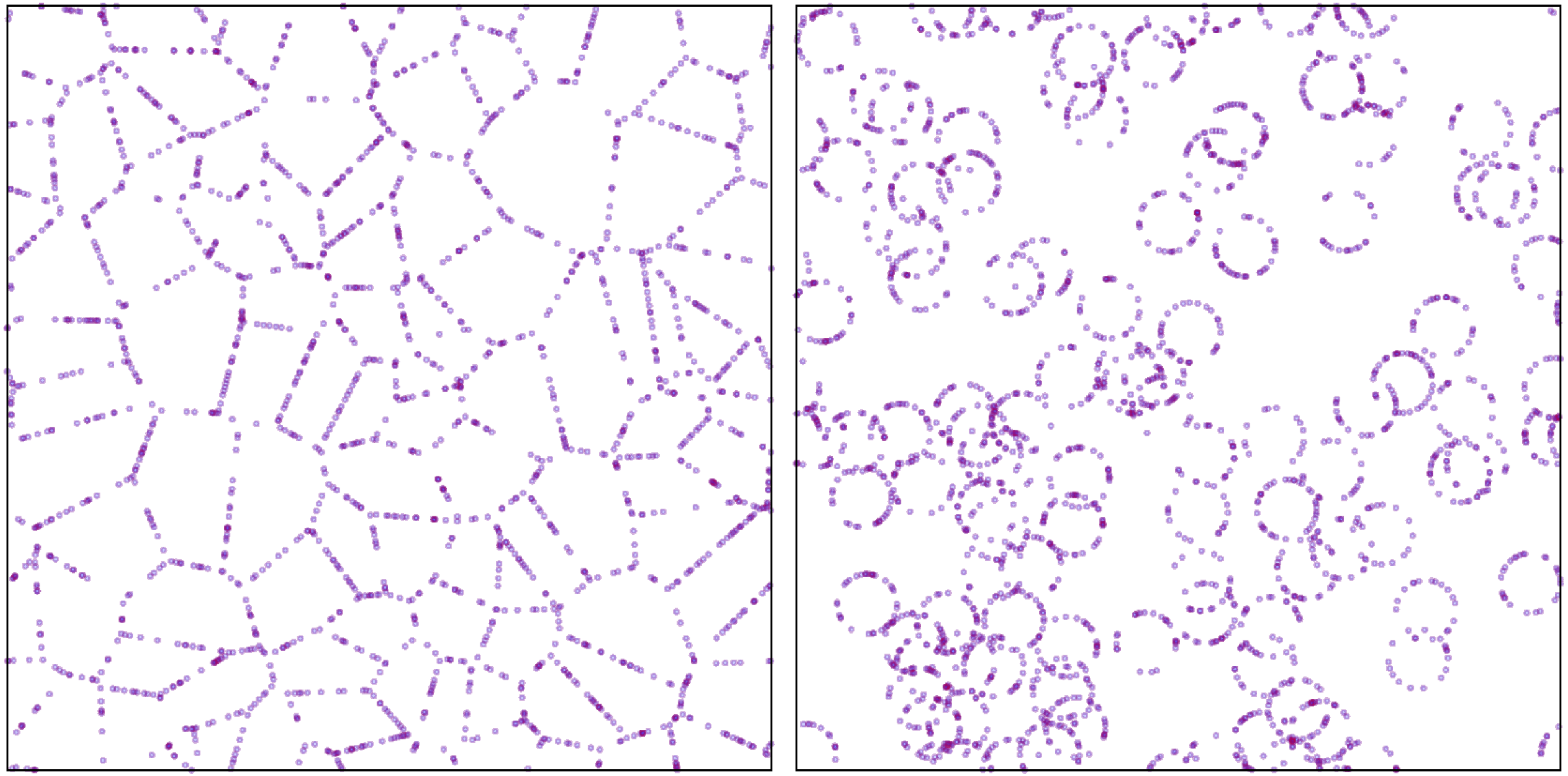
Ergodic learning of spatial geometric structures

B. Błaszczyszyn (Inria/ENS)

Anzère, École d'été CUSO, September 1-4, 2024

INTRODUCTION

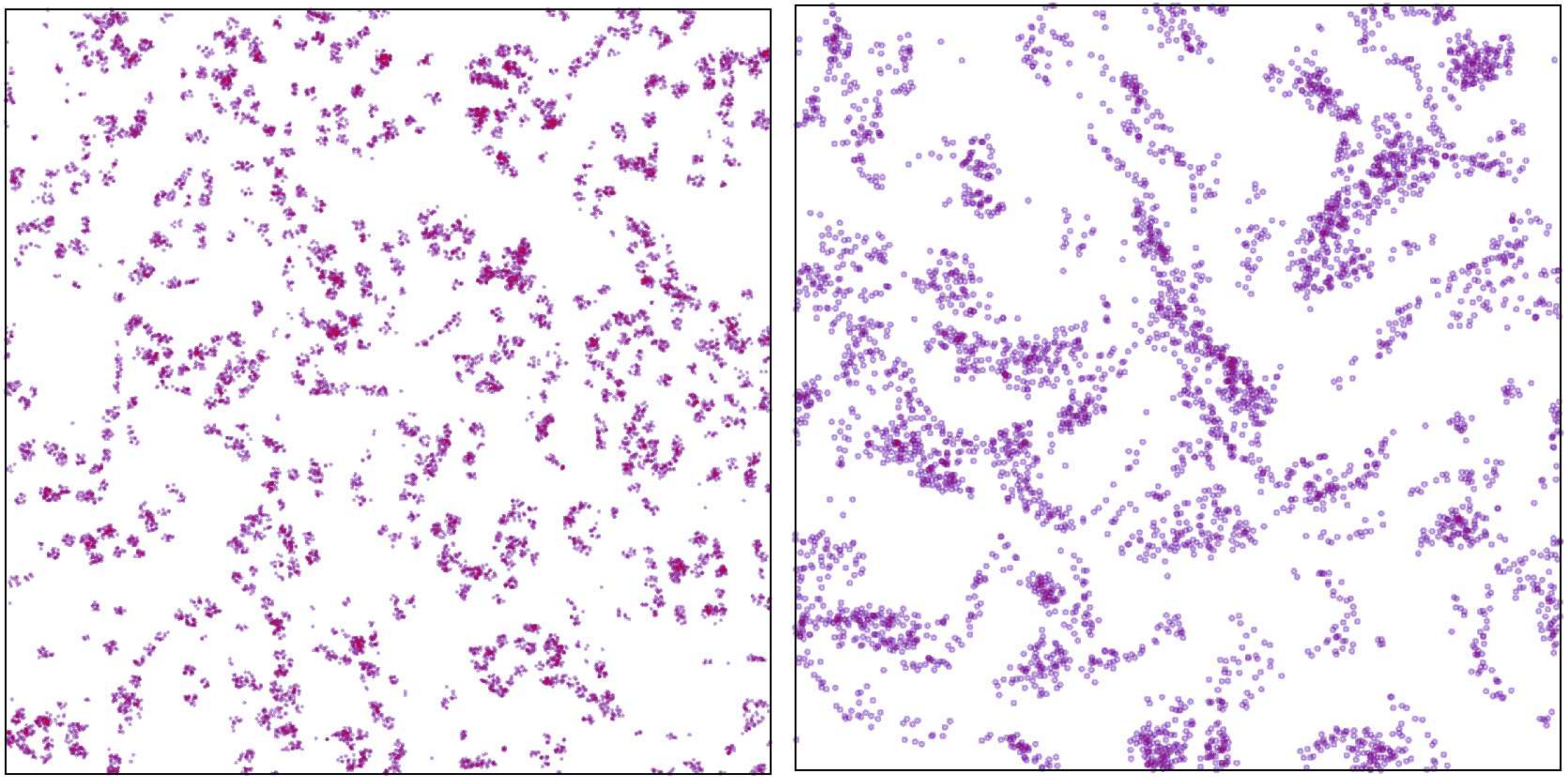
Samples from a point process. Can you recognize the model?



Sure, Cox-Voronoi and Cox-Boolean.

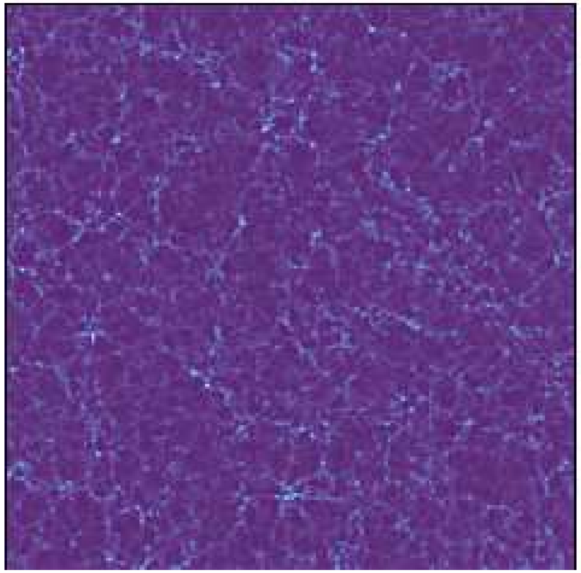
Recall: Cox = doubly stochastic Poisson process.

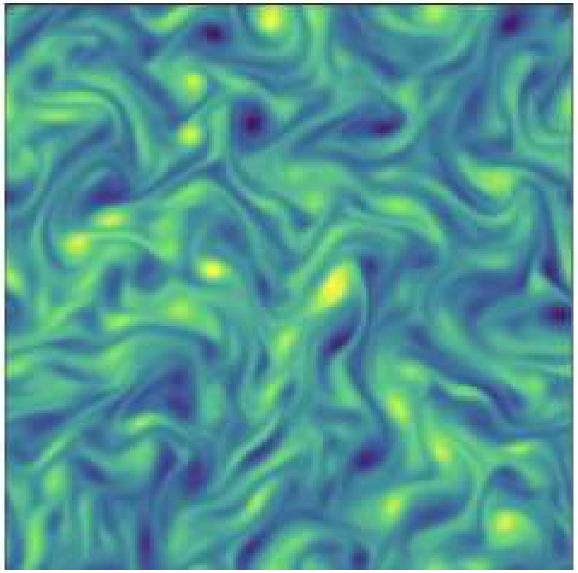
And here?

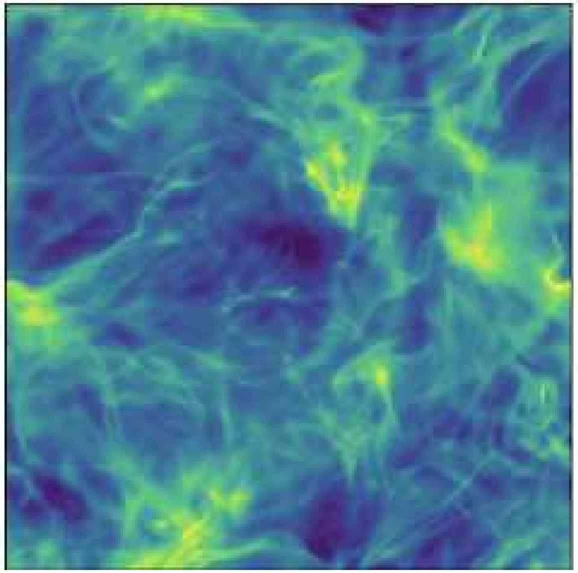


Well... ? Left: Matern cluster process driven by some turbulent field (driven by 2d Navier-Stokes equations). Right: Matern II hard core model applied to a Cox driven by the same turbulent field.

Some more patterns







from astronomy, physics, ...

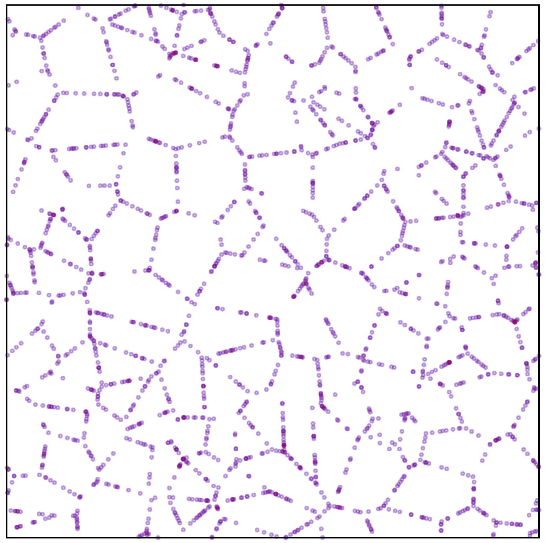
These patterns:

- Exhibit multi-scale properties (e.g. small repulsion, large cluster)
- We want model them with point process with a (very) large number of points (partcles), say $\sim \mathbf{10.000}$, in the window.
- Typically, we have only one original pattern (or, say, very few ones).
- $\Rightarrow$ Ergodic learning of point processes?

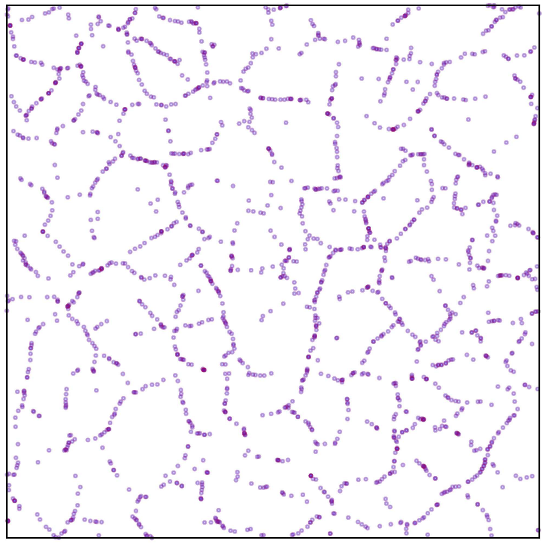
Ergodic learning of point processes

- Recall: Almost surely, any infinite realization of an ergodic point process alows one to fully characterize its distribution and thus (in principle) to sample from this distribution new realizations. $\Rightarrow$ Spatial averaging!

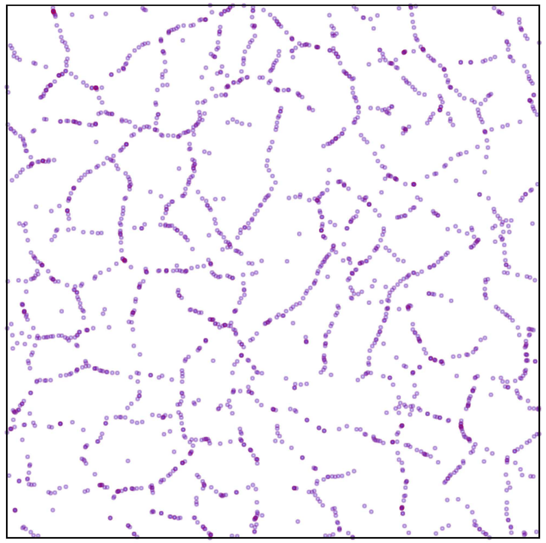
- But in practice, we have only a finite learning window. Can we get approximations of the unknown distribution?



Original image



Synthesis 1



Synthesis 2

samples from "ergodic learning model"

So how it works?

1. Choose statistics (descriptors, moments) that will "summarize" the distribution and not fully "memorize" given patterns.

2. Specify a model deriving from these statistics. Typically a type of "maximum entropy model".

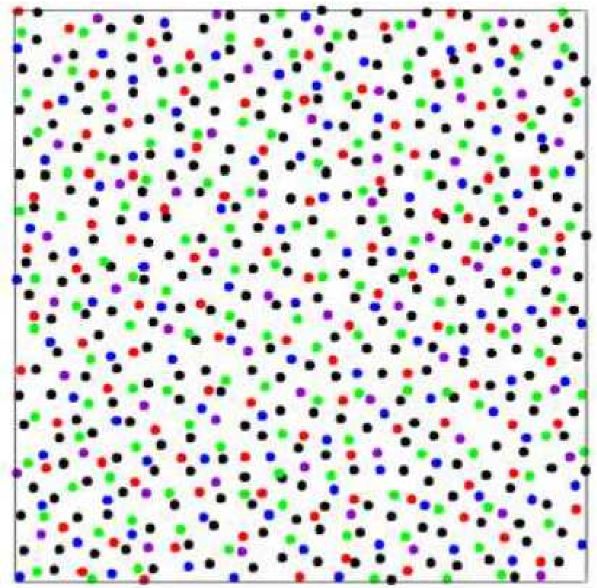
3. Find a way of generating samples from this model. Not always evident!

This is (ergodic) learning of a generative model.

Validation of these models?

- These models involve non-parametric estimation of the entire distribution of the point process and lack mathematical limiting results, making it difficult to perform a confidence analysis.
- The validation of these generative models often relies on visual perception, comparisons of (second) order statistics, or methods from topological data analysis.
- To achieve stronger, more "provable" results, it is necessary to:
  - focus on learning specific features of the given realization,
  - impose mixing conditions, which are stronger than ergodicity.
- For example, consider learning the following feature of the realization...

Hyperuniformity, a striking feature of nature!

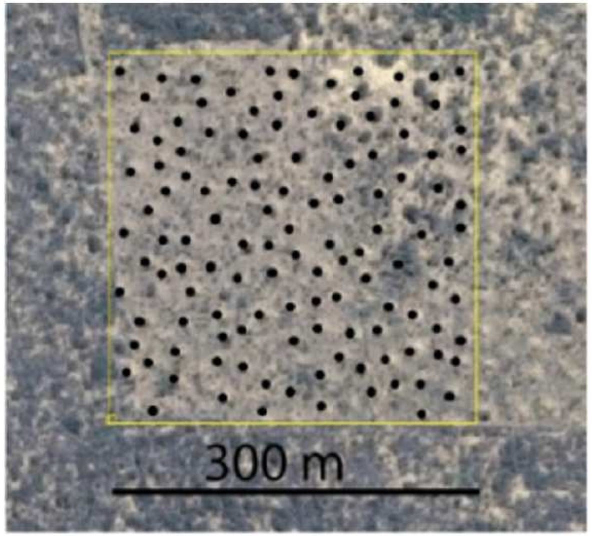


(Jiao et al. 2014)

Avian photoreceptors

(Huang et al. 2021)

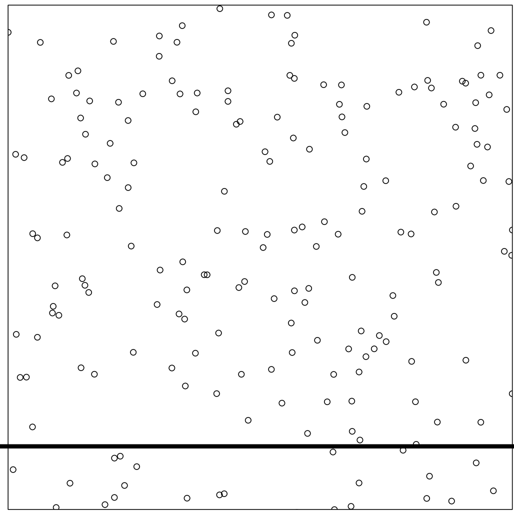
Swimming algae



...

Termite mounds

Patterns more “regular” than complete independence (Poisson model)



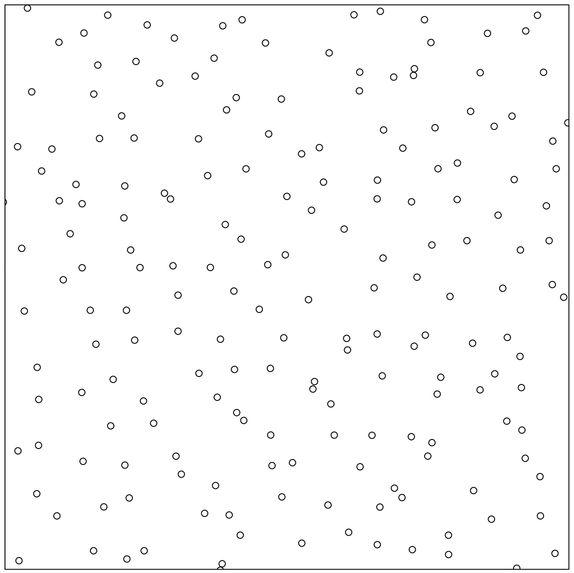
Course Plan

- **Lesson 1:** Crash Course on Point Processes
  - Poisson point process,
  - Stationarity and Palm probabilities,
  - Ergodicity for point processes

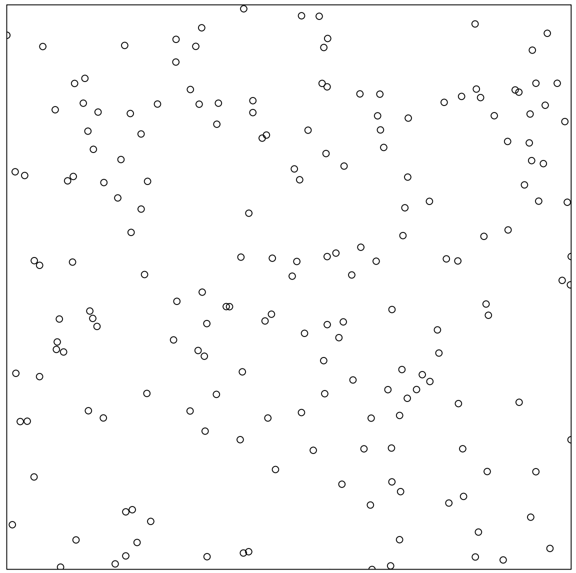
  Based on Chapters 7, 10, and 11 of Lecture Notes on Random Geometric Models by BB; see hal:cel-01654766.
- **Lesson 2:** Ergodic Learning of Point Processes, based on Brochard, A., BB, Mallat, S., and Zhang, S. (2022). Particle Gradient Descent Model for Point Process Generation. *Statistics and Computing*; arXiv:2010.14928.
- **Lesson 3:** Estimating Hyperuniformity, based on Mastrilli, G., BB, Lavancier, F. (2024). Estimating the Hyperuniformity Exponent of Point Processes. arXiv:2407.16797.

POISSON POINT PROCESS

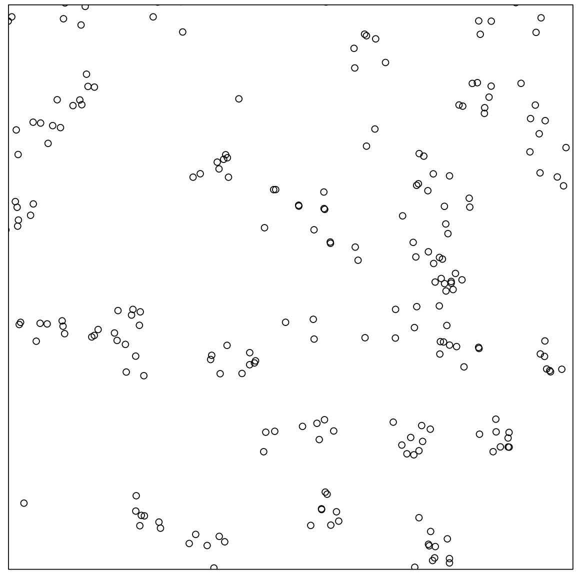
Point process — informally



more regular pattern of repelling points



Poisson point process; independent points



more clustering pattern of attracting points

Point process represents locations of a countable family of particles in some space.

Framework

- Space of points $\mathbb{E}$:
  - A topological space LCSCH (locally compact, second countable, Hausdorff).
  - In particular, $\mathbb{E}$ is Polish space, i.e., separable (there exits a countable, dense subset) and it admits a complete metric. Metric is not unique, in general, we will not use it.
  - $\mathbb{E}$ is σ-compact (i.e., it can be covered by countably many compact sets).
  - We consider Borel σ-algebra $\mathcal{B}$ on $\mathbb{E}$ (generated by the open sets of the topology).
  - A set $B \in \mathcal{B}$ is called (topologically) bounded if it is relatively compact (its closure is compact).
  - Denote by $\mathcal{B}_c$ all bounded Borel subsets of $\mathbb{E}$.
- Standard example of $\mathbb{E}$: d-dimensional Euclidean space $\mathbb{R}^d$, with $1 \leq d < \infty$.

- Space of configurations of points $\mathbb{M}$:
  - A point $x \in \mathbb{E}$ is identified with Dirac measure δ_x; $\delta_x(B) = 1$ if $x \in B$ and 0 otherwise.
  - At most countable subset (configuration) of points $\{x_1, \ldots, x_J\} \subset \mathbb{E}$ is identified with the counting measure on $(\mathbb{E}, \mathcal{B})$

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

  - We consider locally finite configuration of points; $\mu(B) < \infty$ for all $B \in \mathcal{B}_c$ (bounded Borel sets).
  - $\mathbb{M}$ set of all locally finite counting measures on $(\mathbb{E}, \mathcal{B})$; expressed as in (1), where the $(x_i)_{i=1,\ldots,J}$ is a sequence of points of $\mathbb{E}$ without accumulation points.
  - Sometimes, less formally, we write $x \in \mu$ to say $\mu(\{x\}) \geq 1$.

- A counting measure $\boldsymbol{\mu}$ as in (1) is called simple if its atoms (points) $\boldsymbol{x_i}$ are distinct.
- A non-simple measure $\boldsymbol{\mu}$ (corresponding to a configuration with multiple points) can be represented as

$$\boldsymbol{\mu = \sum_{i=1}^{J'} k_i \delta_{x'_i}} \qquad (*)$$

with $\boldsymbol{k_i \in \{1, 2, \ldots\}}$, where atoms $\boldsymbol{x'_i}$ are distinct.

- Measurable configurations in $\mathcal{M}$:
  - We define σ-field $\mathcal{M}$ on $\mathbb{M}$ generated by the mappings $\mu \mapsto \mu(B)\,, B \in \mathcal{B}$ (equivalently for all $B \in \mathcal{B}_c$), i.e.; the smallest σ-algebra making these mappings measurable.
  - Moreover, we can choose the measurable enumeration of the atoms x'_k in (*) such a way that the mappings $\mu \mapsto x'_i$ and $\mu \mapsto k_i$ are measurable.
  - Canonical measurable enumeration of points on $\mathbb{R}$
  $$\ldots < x_{-2} < x_{-1} < x_0 \leq 0 < x_1 < x_2 < \ldots.$$
  - For μ as in (1) and a real function f the mapping
  $$\mu \mapsto \int f \mathrm{d}\mu := \int_{\mathbb{E}} f(x)\,\mu(\mathrm{d}x) = \sum_{i=1}^{J} f(x_i)$$
  is measurable (provided the integrals are defined).

Point process

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

$$(\Phi(B_1), \ldots, \Phi(B_k)),$$

where $k \geq 1$ and $B_1, \ldots, B_k$ run over $\mathcal{B}_c$. (This follows from Kolmogorov's extension theorem as Φ can be seen as a stochastic process $\Phi = \{\Phi(B)\}_{B \in \mathcal{B}_c}$ with the state space $\{0, 1, \ldots\} \ni \Phi(B)$ and where the index B runs over bounded Borel subsets $\mathcal{B}_c$ of $\mathbb{E}$.)
- Point process Φ is simple if $\mathbf{P}\{\, \forall x \in \mathbb{E}, \Phi(\{x\}) \leq 1 \,\} = 1$ (measurability! - Exercise).
- We say Φ has a fixed atom at x_0 if $\mathbf{P}\{\, \Phi(\{x_0\}) > 0 \,\} > 0$.

A few characteristics of point process

- Mean measure $M = M_{\Phi}$ is a measure defined on $(\mathbb{E}, \mathcal{B})$ as $M(B) := \mathbf{E}\left[\Phi(B)\right]$. Note $M(B)$ is well defined for all $B \in \mathcal{B}$ but can be infinite even for $B \in \mathcal{B}_c$.
- Void probability $\nu = \nu_{\Phi}$ is a set function defined on $(\mathbb{E}, \mathcal{B})$ as $\nu(B) := \mathbf{P}\{\,\Phi(B) = 0\,\}$.
- FACT 1: **[Rényi's theorem]** The probability distribution of a simple point process Φ is characterized by the family of its void probabilities $\nu_{\Phi}(B)$ for all $B \in \mathcal{B}_c$.
- REMARK 2: Comparison of void probabilities and (higher order) moment measures allows one to compare clustering properties of point processes. Smaller void probabilities and smaller moment measures indicate more regular point processes, cf figures on page 12.

□ Proof: (of Rényi's theorem; Th. 1)

- Express probabilities of the form $\pi_k(B_1,\ldots,B_k;B) := \mathbf{P}\{\,\Phi(B_1)>0,\ldots,\Phi(B_k)>0,\Phi(B)=0\,\}$ for all $k\ge 1$ and $B_i\in\mathcal{B}_c$.
- Finite dimensional distributions $\mathbf{P}\{\,\Phi(A_1)=n_1,\ldots,\Phi(A_l)=n_l\,\}$ are limits of some expressions involving $\pi_k(\cdot;\cdot)$ with $k\to\infty$ and the sets B_i's dissecting A_s's more and more precisely in a nested way.
- Since a realization of the point process is a locally finite measure at some level of precision in each set B_i there is at most one point of the point process.

- Laplace functional (transform) $\mathcal{L} = \mathcal{L}_\Phi$ is a functional on the space of non-negative, measurable functions $f : \mathbb{E} \mapsto \mathbb{R}^+$ as $\mathcal{L}(f) := \mathbf{E}\left[e^{-\int f \mathrm{d}\Phi}\right]$. Its domain can be extended to functions f for which the expectation is well defined.
- FACT 3: The Laplace functional completely characterizes the distribution of the point process.
- Proof: (Sketch)
  - For $f(x) = \sum_{i=1}^{k} t_i 1(x \in B_i)$, $\mathcal{L}_\Phi(f) = \mathbf{E}\left[e^{-\sum_i t_i \Phi(B_i)}\right]$, seen as a function of the vector $(t_1, \ldots, t_k)$, is the joint Laplace transform of the random vector $(\Phi(B_1), \ldots, \Phi(B_k))$, whose distribution is characterized by this transform.
  - When $B_1, \ldots, B_k$ run over all bounded subsets of the space, one obtains a characterization of all finite-dimensional distributions of the point process.

Campbell's averaging formula

- Basic formula allowing one to evaluate expected values of (random) integrals of deterministic functions with respect to point process.

- THEOREM 4. **[Campbell's averaging formula]** Let $\mathbf{\Phi}$ be a point process on $\mathbb{E}$ with intensity measure $\boldsymbol{M}$. Then for any measurable function $\boldsymbol{f} : \mathbb{E} \to \mathbb{R}$ which is either non-negative or integrable with respect to $\boldsymbol{M}$, the integral $\int_{\mathbb{E}} \boldsymbol{f} \mathbf{d} \mathbf{\Phi}$ is almost surely well defined and

$$\mathbf{E}\left[\int_{\mathbb{E}} \boldsymbol{f}(\boldsymbol{x})\, \mathbf{\Phi}(\mathbf{d}\boldsymbol{x})\right] = \int_{\mathbb{E}} \boldsymbol{f}(\boldsymbol{x})\, \boldsymbol{M}(\mathbf{d}\boldsymbol{x})\,. \tag{2}$$

- Later, Palm theory will offer us an extension of (2) allowing one to consider expectations of the integrals of stochastic processes, i.e.; expressions $\mathbf{E}\left[\int \boldsymbol{f}(\boldsymbol{x}, \mathbf{\Phi})\, \mathbf{\Phi}(\mathbf{d}\boldsymbol{x})\right]$.

- Proof: (of Campbell's averaging formula; Th. 4)
  - Consider first a simple function $f = \sum_{j=1}^{k} a_j \mathbf{1}_{B_j}$, where $a_j \geq 0$ and $B_j \in \mathcal{B}$.
  - Then

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

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

Poisson point process

- DEFINITION: Let $\boldsymbol{\Lambda}$ be a deterministic, locally finite, measure on $(\mathbb{E}, \boldsymbol{\mathcal{B}})$. A point process $\boldsymbol{\Phi}$ on $\mathbb{E}$ is a Poisson point process of intensity (measure) $\boldsymbol{\Lambda}$ if the following two conditions are satisfied:
  1. For any $B \in \mathcal{B}_c$, $\Phi(B)$ is a Poisson random variable of intensity $\Lambda(B)$, i.e.; $\mathbf{P}\{\,\Phi(B) = n\,\} = e^{-\Lambda(B)}(\Lambda(B))^n/n!$.
  2. For every $k = 1, 2, \ldots$ and all sets $B_i \in \mathcal{B}_c$, $i = 1, \ldots, k$, pairwise disjoint, random variables $(\Phi(B_1), \ldots, \Phi(B_k))$ are independent.

- Clearly the above two conditions characterize finite dimensional distributions of a point process, provided it exists(!). We shall construct $\boldsymbol{\Phi}$ later.

Simple characteristics of Poisson point process

- Mean measure is equal to its intensity measure $M(B) = \mathbf{E}[\Phi(B)] = \Lambda(B)$.

- Void probability $\nu(B) = \mathbf{P}\{\, \Phi(B) = 0 \,\} = e^{-\Lambda(B)}$.

- FACT 5: Laplace functional

$$\mathcal{L}(f) = \mathbf{E}\left[e^{-\int_{\mathbb{E}} f \mathrm{d}\Phi}\right] = e^{-\int_{\mathbb{E}}(1-e^{-f(x)})\,\Lambda(\mathrm{d}x)}\,. \tag{3}$$

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

–

$$\mathcal{L}(f) = \mathbf{E}\left[e^{-\int_{\mathbb{E}} f \mathrm{d}\Phi}\right]$$

$$= \mathbf{E}\left[\prod_{j=1}^{k} e^{-a_j \Phi(B_j)}\right]$$

by the independence of $\Phi(B_j), j = 1, \dots, k$ $$= \prod_{j=1}^{k} \mathbf{E}\left[e^{-a_j \Phi(B_j)}\right]$$

by Poisson distribution of $\Phi(B_j)$ $$= \prod_{j=1}^{k} e^{-\Lambda(B_j)(1-e^{-a_j})}$$

$$= e^{-\sum_{j=1}^{k} \Lambda(B_j)(1-e^{-a_j})}$$

$$e^{-\int (1-e^{-f}) \mathrm{d}\Lambda} .$$

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

Conditional distribution of Poisson points given the number

- FACT 6: Consider $B_1, \ldots, B_k \in \mathcal{B}_c$ pariwise disjoint and denote $W := \sum_{i=1}^n B_k$. For all $n, n_1, \ldots, n_k \in \{0, 1, \ldots\}$ with $\sum_i n_i = n$,

$$\mathbf{P}\{\, \Phi(B_1) = n_1, \ldots, \Phi(B_k) = n_k \mid \Phi(W) = n \,\} \tag{4}$$
$$= \frac{n!}{n_1! \ldots n_k!} \frac{1}{\Lambda(W)^n} \prod_{i=1}^{k} \Lambda(B_i)^{n_i} \,.$$

- Proof: (Exercise)
- REMARK 7:
  - We recognize in the above conditional distribution is a multinomial distribution.
  - We can conclude form Fact 6 that given there are n points of the Poisson process in the window W, these points are i.i.d. in W according to the law $\frac{\Lambda(\cdot)}{\Lambda(W)}$.

Construction of Poisson point process

- Given a Radon measure $\boldsymbol{\Lambda}$ on $\mathbb{E}$ and bounded $\boldsymbol{W} \in \boldsymbol{\mathcal{B}_c}$. Consider the following independent random objects $\{\boldsymbol{N}, \boldsymbol{X_1}, \boldsymbol{X_2}, \ldots\}$, where
  - $\boldsymbol{N}$ is a Poisson random variable with parameter $\boldsymbol{\Lambda(W)}$,
  - $\boldsymbol{X_1, X_2, \ldots}$ are identically distributed random vectors (points) taking values in $\boldsymbol{W}$ with

$$\mathbf{P}\{\, \boldsymbol{X_1} \in \cdot \,\} = \boldsymbol{\Lambda}(\cdot)/\boldsymbol{\Lambda}(\boldsymbol{W}). \tag{5}$$

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

Simple Poisson point process

- For a general point processes, simple property (not have multiple points) and not having fixed atoms are two different properties, except in case of a Poisson process.

- FACT 8: Let $\mathbf{\Phi}$ be a Poisson process on $\mathbb{E}$ with intensity measure $\mathbf{\Lambda}$.

  1. $\mathbf{\Phi}$ has a fixed atom at $x_0 \in \mathbb{E}$ iff $\mathbf{\Lambda}$ has an atom at x_0 (i.e., $\mathbf{\Lambda}(\{x_0\}) > 0$).

  2. $\mathbf{\Phi}$ is simple iff $\mathbf{\Lambda}$ is non-atomic, i.e.; $\mathbf{\Lambda}(\{x\}) = 0$ for all $x \in \mathbb{E}$.

- Proof:

  – The first statement is straightforward from the definition.

– For the second statement, use conditional distribution of Poisson points in a bounded subset $B \subset \mathcal{B}_c$ (see Remark 7)

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

– Now

$$\int_{B^n} \mathbf{1}(\exists_{i\neq j} x_i = x_j)\, \Lambda(\mathrm{d}x_1)\ldots\Lambda(\mathrm{d}x_n)$$

$$\leq \sum_{i<j=1}^{n} \int_{B^n} \mathbf{1}(x_i = x_j)\, \Lambda(\mathrm{d}x_1)\ldots\Lambda(\mathrm{d}x_n)$$

$$= \frac{n(n-1)}{2}(\Lambda(B))^{n-2}\int_B \Lambda(\{x\})\Lambda(\mathrm{d}x)$$

$\Lambda(\{x\}) = 0$ since Λ is non-atomic

$$= 0$$

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

Homogeneous Poisson process on $\mathbb{R}^d$

- DEFINITION: Poisson process of intensity $\boldsymbol{\Lambda(\mathrm{d}x) = \lambda \mathrm{d}x}$ on $\mathbb{R}^{\boldsymbol{d}}$, where $\boldsymbol{\lambda}$ ($\boldsymbol{0 < \lambda < \infty}$) is a constant, is called homogeneous Poisson process of intensity $\boldsymbol{\lambda}$.
- Homogeneous Poisson process is simple (by Fact 8).

Markov property of Poisson process

- Let $\mathbf{\Phi}$ is a Poisson point process and consider a real measurable function f on $(\mathbb{M}, \mathcal{M})$. For any $B \in \mathcal{B}_c$

$$\mathbf{E}\left[f(\mathbf{\Phi})\right] = \mathbf{E}\left[f\left(\mathbf{\Phi}|_B + \mathbf{\Phi}'|_{\mathbb{E}\setminus B}\right)\right], \quad (6)$$

  where $\mathbf{\Phi}'$ is an independent copy of $\mathbf{\Phi}$ and $\mu|_B$ denote the truncation of the measure $\mu|_B(\cdot) = \mu(\cdot \cap B)$.

- (6) follows directly from the definition of Poisson process.

- The strong Markov property of Poisson process says that the above statement hold when B is not necessarily constant but a random stopping set with respect to $\mathbf{\Phi}$.

Random stopping set

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

- EXAMPLE: **[k th smallest random ball]**
  - For a point process Φ on $\mathbb{R}^d$, consider the random (closed) ball $B_0(R_k^*)$ centered at the origin, with the random radius equal to the k th smallest norm of $x_i \in \Phi$; i.e.,

$$R_k^* = R_k^*(\Phi) = \min\{r \geq 0 : \Phi(B_0(r)) = k\}.$$

  - Think of start 'growing' a ball $B_0(r)$ centered at the origin, increasing its radius r from 0 until the moment when either (1) it accumulates k or more points or (2) it hits the complement K^c of K.
  - If (1) happens, then $B_0(R_k^*) \subset K$. If (2) happens, then $B_0(R_k^*) \not\subset K$.
  - In either case, we have not used any information about points of Φ in K^c; so $B_0(R_k^*) = B_0(R_k^*(\Phi))$ is a stopping set with respect to Φ.

Strong Markov property of Poisson point process

- The following result extends (6) to the case when B is a stopping set.

- FACT 9: **[Strong Markov property of Poisson point process]** Let Φ be a Poisson point process and $S = S(\Phi)$ a random stopping set relative to Φ (one can know if $\{ S(\Phi) \subset K \}$ is true or not knowing only the points of Φ in K). Then the following holds

$$\mathbf{E}\left[f(\Phi)\right] = \mathbf{E}\left[f\left(\Phi|_{S(\Phi)} + \Phi'|_{\mathbb{E}\setminus S(\Phi)}\right)\right] . \qquad (7)$$

- Proof: The Poisson process is a Markov field indexed by $B \in \mathcal{B}_c$. The result follows by a general result for Markov fields (see e.g. Rozanov 1982).

Exponential construction of Poisson process on $\mathbb{R}$

- Consider homogeneous Poisson point process Φ of intensity λ on the real line $\mathbb{R}$ ($0 < \lambda < \infty$) and enumerate its points $\Phi = \sum_k \delta_{X_k}$ in the canonical way (see page 16).
- In particular, $X_1 = \sup\{x > 0 : \Phi((0,x)) = 0\}$ is the first atom of Φ in the open positive half-line $(0,\infty)$.
- $\{X_k\}$ can be constructed as a renewal process with exponential holding times, i.e., $X_k = \sum_{i=1}^k F_i$ for $k \geq 1$ and $X_k = -\sum_{i=k}^0 F_i$ for $k \leq 0$, where $\{F_k : k = \ldots, -1, 0, 1 \ldots\}$ is a sequence of independent, identically distributed exponential random variables.
- Indeed,

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

so $X_1 = F_1$ is exponential random variable with parameter λ.

- By the strong Markov property for $k \geq 2$,

$$\begin{aligned} \mathrm{P}\{\, F_k > t \mid F_1, \ldots, F_{k-1} \,\} &= \mathrm{P}\{\, X_k - X_{k-1} > t \mid X_1, \ldots, X_{k-1} \,\} \\ \boxed{\text{By (7) with } S(\Phi) = [0, X_{k-1}]} \quad &= \mathrm{P}\{\, \Phi((X_{k-1}, X_{k-1} + t]) = 0 \mid X_{k-1} \,\} \\ &= e^{-\lambda t} \end{aligned}$$

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

Ordering of Poisson points according to their distance

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

$$\mathrm{P}\Big\{\, R_k^* > t \mid R_{k-1}^* = s \,\Big\} = \begin{cases} e^{-\Lambda(B_0(t)) - \Lambda(B_0(s))} & \text{if } t > s \\ 1 & \text{if } t \leq s\,. \end{cases} \tag{8}$$

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

Equivalent characterizations of Poisson process

- Under mild assumptions Poisson distribution or independence alone characterizes the Poisson process.
- PROPOSITION 10. **[Characterization by the form of the distribution]** Suppose that Φ is a simple. Then Φ is a Poisson point process iff there exists a locally finite, non-atomic measure Λ such that for any subset $B \in \mathcal{B}_c$, $\mathbf{P}\{\,\Phi(B) = 0\,\} = e^{-\Lambda(B)}$.
- Proof: This is a consequence of the Rényi's theorem (cf Fact 1).
- COROLLARY 11. Φ is a Poisson process provided it is simple and all marginal distributions $\Phi(B)$ for $B \in \mathcal{B}_c$ are Poisson.
- The assumption that Φ is simple cannot be relaxed since one can construct two Poisson random variables N_1 and N_2, of parameters μ_1, μ_2, respectively, and such that $N_1 + N_2$ is Poisson of parameter $\mu_1 + \mu_2$, with N_1 and N_2 not being independent.

Complete independence

- One says that the point process Φ has the property of complete independence if for any finite family of subsets $B_1, \ldots, B_k \in \mathcal{B}_c$ that are mutually disjoint, the random variables $\Phi(B_1), \ldots, \Phi(B_k)$ are independent.
- PROPOSITION 12. **[Characterization by complete independence]** Suppose that Φ is a point process without fixed atoms. Then Φ is a Poisson point process iff Φ is simple and has the property of complete independence.
- Proof:
  - The necessity follows from Fact 8 (Poisson without fixed atoms is simple).
  - For sufficiency, one shows that the measure $\Lambda(A) = -\log(\mathrm{P}\{\Phi(A) = 0\})$ satisfies the assumptions of Proposition 10 (Characterization by the form of the distribution).
- The assumption that Φ has no fixed atoms cannot be relaxed to a simple Poisson process; it destroys Poisson distribution of $\Phi(\{x_0\})$.

Operations preserving the Poisson law

- Superposition: sum of point processes $\Phi = \sum_k \Phi_k$.
- If infinite sum, one has to ensure that Φ is a locally finite measure. A crude, sufficient condition $\sum_k \mathbf{E}[\Phi_k(B)] < \infty$ for bouned $B \in \mathcal{B}_c$. A refined sufficient condition may be found by the Borel–Cantelli lemma.
- FACT 13: The superposition of independent Poisson point processes with intensities Λ_k is a Poisson point process with intensity measure $\sum_k \Lambda_k$ iff the latter is a locally finite measure.

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

$$\mathbf{\Phi}^p = \sum_k I_k \delta_{X_k} , \tag{9}$$

where the random variables $\{I_k\}_k$ are independent given $\mathbf{\Phi}$, and $\mathbf{P}\{ I_k = 1 \mid \mathbf{\Phi} \} = 1 - \mathbf{P}\{ I_k = 0 \mid \mathbf{\Phi} \} = p(X_k)$.
- Fact 14: The thinning of the Poisson point process of intensity measure $\mathbf{\Lambda}$ with the retention probability p yields a Poisson point process of intensity measure $p\mathbf{\Lambda}$ with $(p\mathbf{\Lambda})(B) = \int_B p(x)\, \mathbf{\Lambda}(\mathrm{d}x)$.
- Proof: Exercise; use Laplace functional characterization.

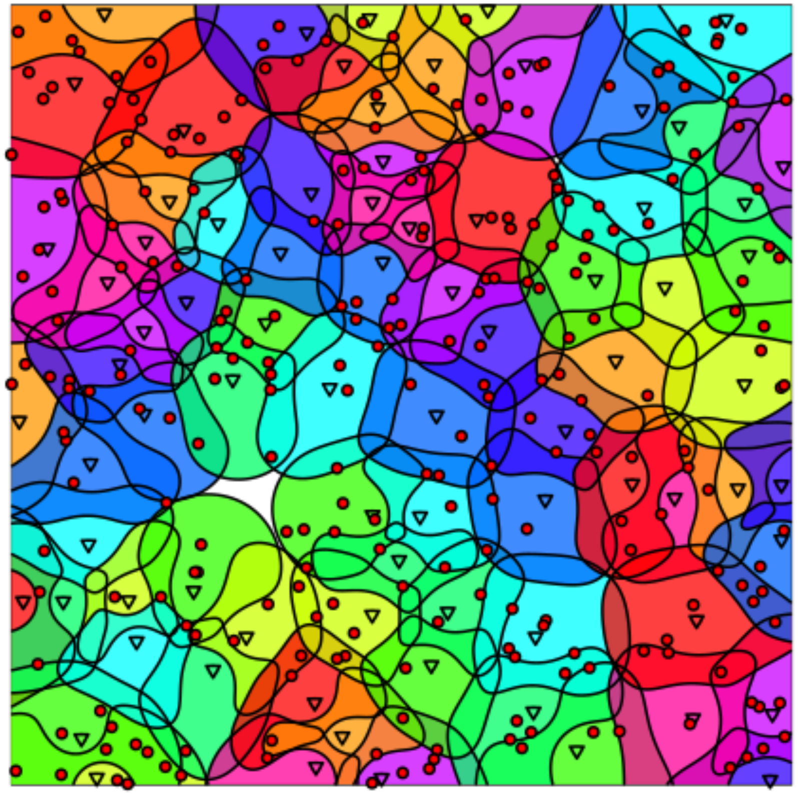
- Random Transformation of points: independent displacing each point according to some probability kernel p.
- Formally, consider a probability kernel $p(x, B)$ from $(\mathbb{E}, \mathcal{B})$ to some LCSCH space $(\mathbb{E}', \mathcal{B}')$, The transformation Φ^p of a point process $\Phi = \sum_k \delta_{X_k}$ by a probability kernel $p(\cdot, \cdot)$ is a point process on $\mathbb{E}'$ given by

$$\Phi^p = \sum_k \delta_{Y_k} , \tag{10}$$

  where the $\mathbb{E}'$-valued random vectors $\{Y_k\}_k$ are independent given Φ, with $\mathbf{P}\{ Y_k \in B' \mid \Phi \} = p(X_k, B')$. W tacitly assume that Φ^p is locally finite measure.
- Fact 15: **[Displacement Theorem]** The transformation of the Poisson point process of intensity measure Λ by a probability kernel p is the Poisson point process with intensity measure $\Lambda'(B') = \int_{\mathbb{E}} p(x, B) \, \Lambda(\mathrm{d}x)$, $B' \subset \mathbb{E}'$, provided Λ' is a locally finite measure.
- Proof: Exercise; use Laplace functional characterization.

STATIONARITY

An example of conservation law



$\mathbf{E}^{\circ}$[Number of cells covering a typical user]
$=$ Intensity of cell-centers
$\times\ \mathbf{E}^{\triangledown}$[Area of the typical cell]

Realization of a random network with cells "centered" at points "$\triangledown$" and locations at users "$\circ$". If the random model is stationary, independent of stationary users, then the *mean number of cells covering a typical user is equal to the mean area of the typical cell multiplied by the intensity of cell centers.*

Stationary of point and random processes

- Consider point processes on d-dimensional Euclidean space $\mathbb{E} = \mathbb{R}^d$, with Borel σ-algebra $\mathcal{B}$ and the corresponding space of counting measures $(\mathbb{M}, \mathcal{M})$ on $(\mathbb{R}^d, \mathcal{B})$.
- Point process $\Phi : (\Omega, \mathcal{A}, \mathrm{P}) \longrightarrow (\mathbb{M}, \mathcal{M})$ is called stationary if its distribution $\mathbf{P}_\Phi$ on $(\mathbb{M}, \mathcal{M})$ invariant with respect to the translation by any vector $t \in \mathbb{R}^d$

$$\mathbf{P}_\Phi = \mathbf{P}_{S_t \Phi} \qquad \text{for all } t \in \mathbb{R}^d, \tag{11}$$

where t-shit $S_t\Phi$ is the translation of all atoms of Φ by $-t$.

- Stochastic process $X = \{X(x)\}_{x \in \mathbb{R}}$ on $(\Omega, \mathcal{A}, \mathrm{P})$ with values $X(x)$ in some measurable space $(\mathbb{K}, \mathcal{K})$ is called stationary if its distribution $\mathbf{P}_{\{X(x)\}_{x \in \mathbb{R}^d}}$ is invariant with respect to the translation of its argument x by any vector $t \in \mathbb{R}^d$

$$\mathbf{P}_{\{X(x)\}_{x \in \mathbb{R}^d}} = \mathbf{P}_{\{X(x+t)\}_{x \in \mathbb{R}^d}} \qquad \text{for all } t \in \mathbb{R}^d. \tag{12}$$

Need a framework for joint stationarity

- A family of point processes Φ_i and stochastic processes X_i, $i = 1, 2, \ldots$ defined on the same probability space is called jointly stationary if the joint distribution of all these random objects is invariant with respect to the respective translation by any vector $t \in \mathbb{R}^d$.
- In order to facilitate the analysis of such jointly stationary objects (e.g. their conservation laws) we shall introduce some stationary framework assuming a flow (ω-shift) directly on the probability space $(\Omega, \mathcal{A}, \mathbf{P})$.

Shift operator on measures and functions

- For any $t \in \mathbb{R}^d$, let S_t be the shift operator on the space of measures: for any measure μ on $(\mathbb{R}^d, \mathcal{B})$ $S_t\mu$ is a measure on $(\mathbb{R}^d, \mathcal{B})$ such that

$$S_t\mu(B) = \mu(B + t),$$

where $B + t = \{x + t \in \mathbb{R}^d : x \in B\}$.

- Equivalently, for atomic measure $\mu = \sum_i \delta_{x_i}$ we have

$$S_t\mu = \sum_i \delta_{x_i - t} \,. \tag{13}$$

Shift operator on functions

- We extend the shift operator to functions $\boldsymbol{X}(\cdot)$ defined on $\mathbb{R}^{\boldsymbol{d}}$ with values is some arbitrary space, by putting

$$\boldsymbol{S_t X}(\boldsymbol{x}) = \boldsymbol{X}(\boldsymbol{x} + \boldsymbol{t})\,.$$

- The following immediate relation will be often used, for $\boldsymbol{B} \in \boldsymbol{\mathcal{B}}$, $\boldsymbol{t} \in \mathbb{R}^{\boldsymbol{d}}$

$$\int_{\boldsymbol{B}} \boldsymbol{X}(\boldsymbol{x})\, \boldsymbol{S_t}\boldsymbol{\mu}(\mathrm{d}\boldsymbol{x}) = \int_{\boldsymbol{B}+\boldsymbol{t}} \boldsymbol{X}(\boldsymbol{x}-\boldsymbol{t})\, \boldsymbol{\mu}(\mathrm{d}\boldsymbol{x}) = \int_{\boldsymbol{B}+\boldsymbol{t}} \boldsymbol{S_{-t}X}(\boldsymbol{x})\, \boldsymbol{\mu}(\mathrm{d}\boldsymbol{x})\,. \tag{14}$$

Flow on the probability space

- Consider a measurable space $(\Omega, \mathcal{A})$ that will serve as the probability space. We assume there exists a family of measurable mappings $\theta_t : \Omega \to \Omega$, $t \in \mathbb{R}^d$ satisfying the following conditions:

  1. For each $t \in \mathbb{R}^d$, the mapping θ_t is a bijection from Ω to Ω.
  2. For all $t, s \in \mathbb{R}^d$, $\theta_t \circ \theta_s = \theta_{s+t}$, with $\circ$ denoting the composition of mappings on Ω.
  3. The mapping $(\mathbb{R}^d, \Omega) \ni (t, \omega) \longmapsto \theta_t(\omega)$ is $\mathcal{B} \otimes \mathcal{A}$ measurable.

- Observe that for any given $t \in \mathbb{R}^d$, the inverse of θ_t is equal to

$$\theta_t^{-1} = \theta_{-t}.$$

- The family $\{\theta_t\}_t$ of mappings satisfying conditions $1, 2, 3$ above will be called (measurable) flow on $(\Omega, \mathcal{A})$. We shall denote the space equipped with such flow by $(\Omega, \mathcal{A}, \{\theta_t\})$.
- More abstractly: θ is a measurable action of the group $(\mathbb{R}^d, +)$ on Ω.

Point processes compatible with the flow

- We shall say that a point process $\Phi : (\Omega, \mathcal{A}, \{\theta_t\}) \to (\mathbb{M}, \mathcal{M})$ is compatible with the flow if for all $t \in \mathbb{R}^d$

$$\Phi \circ \theta_t = S_t \Phi \,,$$

  where $S_t\Phi$ is the shift of the counting measure Φ (here, S is a measurable action of the group $(\mathbb{R}^d, +)$ on $\mathbb{M}$.

- In other words:

$$\Phi(\theta_t(\omega))(B) = S_t\Phi(\omega)(B) = \Phi(\omega)(B + t)$$

  for all $\omega \in \Omega$, $B \in \mathcal{B}$.

- EXAMPLE: [Canonical probability space with the flow] The space $(\Omega, \mathcal{A}, \{\theta_t\}) = (\mathbb{M}, \mathcal{M}, \{S_t\})$ is the canonical space supporting point process $\Phi(\mu) = \mu$, $\mu \in \mathbb{M}$, compatible with the flow.

Stochastic processes compatible with the flow

- Similarly, stochastic process $X = \{X(x)\}_{x\in\mathbb{R}^d}$ defined on $(\Omega, \mathcal{A}, \{\theta_t\})$, with values some measurable space, will be said compatible with the flow if for all $t \in \mathbb{R}^d$

$$X \circ \theta_t = S_t X\,;$$

 that is

$$\{X(\theta_t(\omega))(x)\}_{x\in\mathbb{R}^d} = \{S_t X(\omega)(x)\}_{x\in\mathbb{R}^d} = \{X(\omega)(x+t)\}_{x\in\mathbb{R}^d}\,.$$

 Here S is a measurable action of the group $(\mathbb{R}^d, +)$ on the space of functions $X : \mathbb{R}^d \mapsto \mathbb{K}$.

Stochastic processes related to point processes

- Most of our stochastic processes compatible with the flow will be fonctionals of some point processes compatible with the flow, as in the following example.

- EXAMPLE: Let Φ be a point process compatible with the flow in $(\Omega, \mathcal{A}, \{\theta_t\})$. Consider stochastic process

$$X(x) := \min_{y \in \Phi} |y - x|$$

describing the distance from the argument x to the nearest point of Φ. X is generated by random variable $R^* := X(0) = \min_{y \in \Phi} |y|$. Indeed,

$$\begin{aligned} R^* \circ \theta_x &= \min_{y \in \Phi \circ \theta_x} |y| \\ \boxed{\text{compatibility of } \Phi} \quad &= \min_{y \in S_x \Phi} |y| \\ \boxed{\text{by (13)}} \quad &= \min_{y' \in \Phi} |y - x| = X(x)\,. \end{aligned}$$

Stationary probability

- Let $(\Omega, \mathcal{A}, \{\theta_t\})$ be a measurable space with the flow. Let $\mathbf{P}$ be a probability measure on $(\Omega, \mathcal{A})$ invariant with respect to all elements of the flow

$$\mathbf{P}\theta_t^{-t} = \mathbf{P} \quad \text{for all } t \in \mathbb{R}^d; \tag{15}$$

  that is $\mathbf{P}\{\, \omega : \theta_t(\omega) \in A \,\} = \mathbf{P}\{\, A \,\}$ for all $A \in \mathcal{A}$.

- We call $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$ a stationary framework.
- The following result follows immediate from the above definition...

- FACT 16: Let $\boldsymbol{\Phi_i}$ and $\boldsymbol{X_i}$, $\boldsymbol{i = 1, 2, \ldots}$ be a family of point processes and stochastic processes,respectively, defined on a stationary framework $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \{\boldsymbol{\theta_t}\}, \mathbf{P})$ compatible with the flow. Then $\boldsymbol{\Phi_i}$ and $\boldsymbol{X_i}$ are jointly stationary

$$\mathbf{P}_{(S_t\Phi_1, S_t\Phi_2, \ldots, S_t X_1, S_t X_2, \ldots)} = \mathbf{P}_{(\Phi_1, \Phi_2, \ldots, X_1, X_2, \ldots)} \quad \text{for all } t \in \mathbb{R}^d .$$

(Exercise)

- We will call $\mathbf{P}$ the stationary probability on $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \{\boldsymbol{\theta_t}\})$ to distinguish it form the Palm probabilities of different point processes to be defined on the same space...

- EXAMPLE: The distribution of a homogeneous Poisson process (having intensity $\boldsymbol{\Lambda}(\mathbf{d}\boldsymbol{x}) = \boldsymbol{\lambda}\mathbf{d}\boldsymbol{x}$) is invariant with respect to any shift $\boldsymbol{S_t}$, $\boldsymbol{t} \in \mathbb{R}^{\boldsymbol{d}}$. The canonical probability space can serve as a stationary framework for it.

Intensity of stationary point process

- From now on $\boldsymbol{\Phi}$ will be a point process defined on the stationary framework $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \{\boldsymbol{\theta_t}\}, \mathbf{P})$ and compatible with the flow.

- FACT 17: The mean measure $\boldsymbol{M(\mathrm{d}x) = M_{\Phi}(\mathrm{d}x)}$ of $\boldsymbol{\Phi}$ is equal to the Lebesgue measure multiplied by a constant

$$\boldsymbol{M(\mathrm{d}x) = \lambda\,\mathrm{d}x,}$$

with $\boldsymbol{0 \le \lambda \le \infty}$. We call the constant $\boldsymbol{\lambda = \lambda_{\Phi}}$ the intensity of the point process $\boldsymbol{\Phi}$.

□ Proof:

– M is invariant with respect to any shift S_t, $t \in \mathbb{R}^d$.
Indeed, for $B \in \mathcal{B}$

$$\begin{aligned}
S_t M(B) &= M(B+t) \\
&= \mathbf{E}\left[\Phi(B+t)\right] \\
&= \mathbf{E}\left[S_t \Phi(B)\right] \\
\boxed{\text{compatibility of } \Phi} \quad &= \mathbf{E}\left[\Phi \circ \theta_t(B)\right] \\
\boxed{\text{invariance of } \mathbf{P}} \quad &= \mathbf{E}\left[\Phi(B)\right] \\
&= M(B)\,.
\end{aligned}$$

– The only measure on $(\mathbb{R}^d, \mathcal{B})$ that is invariant with respect to all shifts is a constant-multiple of the Lebesque measure.

Palm probability w.r.t. a point process

- Let Φ be a point process compatible with the flow on the probability space $(\Omega, \mathcal{A}, \{\theta_t\})$, with finite, non-null intensity $0 < \lambda < \infty$. The Palm probability of Φ (or related to Φ) is the unique probability measure $\mathbf{P}^0$ on $(\Omega, \mathcal{A})$ given by

$$\mathbf{P}^0(A) = \frac{1}{\lambda|B|}\mathbf{E}\left[\int_{\mathbb{R}^d} 1(x \in B)1(\theta_x \in A)\,\Phi(\mathrm{d}x)\right] \qquad A \in \mathcal{A}; \quad (16)$$

  with any set $B \in \mathcal{B}$ of finite, non-null Lebesgue measure $|B|$.

- One can verify that $\mathbf{P}^0$ is indeed a probability measure and its value does not depend on the choice of the set B.

- In what follows we shall denote by $\mathbf{E}^0$ the expectation with respect to $\mathbf{P}^0$.

- The following result is the variant of the Campbell-Little-Mecke result (for general point processes).
- THEOREM 18. [CLMM] Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, compatible with the flow, and having finite, non-null intensity $0 < \lambda < \infty$. Denote by $\mathbf{P}^0$ the Palm probability of Φ. For any non-negative measurable functions f on $\mathbb{R}^d \times \Omega$ (but not necessarily compatible with the flow), we have

$$\mathbf{E}\left[\int_{\mathbb{R}^d} f(x, \theta_x)\, \Phi(\mathrm{d}x)\right] = \lambda \int_{\mathbb{R}^d} \mathbf{E}^0\left[f(x, \omega)\right] \mathrm{d}x. \tag{17}$$

  The result extends to all functions f for which either of the two sides of the equality (17) is finite when f is replaced by $|f|$.
- Proof: Directly from the definition of $\mathbf{P}^0$ one easily shows the desired equality for $f(x, \omega) = 1(x \in B, \omega \in A)$ where $B \in \mathcal{B}$, $A \in \mathcal{A}$. The result follows by usual measure theoretic approximation arguments.

First properties of Palm P^0: typical point $X_0 = 0$

- COROLLARY 19. Under the assumptions of Theorem 18, $\mathbf{P}^0$-almost surely $\mathbf{0} \in \mathbf{\Phi}$;

$$\mathbf{P}^0\{\,\mathbf{0} \in \mathbf{\Phi}\,\} = \mathbf{1}.$$

  (Exercise)

- The point $\mathbf{0}$ of $\mathbf{\Phi}$ under $\mathbf{P}^0$ is called the typical point of $\mathbf{\Phi}$. (We shall see arguments for this.) By the convention, we assign to the typical point the index 0; thus $X_0 = \mathbf{0}$ under $\mathbf{P}^0$.

□ COROLLARY 20. Under the assumptions of Theorem 18, let $\mathbf{P}^{\mathbf{0}}_{\mathbf{\Phi}}$ be the distribution of $\mathbf{\Phi}$ under its Palm probability $\mathbf{P}^{\mathbf{0}}$ and $\boldsymbol{P_x}$ Palm distributions of $\mathbf{\Phi}$. Then

$$\mathbf{P}^{\mathbf{0}}_{\mathbf{\Phi}} = \boldsymbol{P_x S_x^{-1}} \qquad \text{for Lebesgue almost all } \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{d}}.$$

(Exercise)

Slivnyak-Mecke's characterization for homogeneous Poisson

- COROLLARY 21. A stationary point process with finite intensity is a Poisson point process iff its distribution under the Palm probability (considered in some stationary framework e.g. the canonical one) is equal to the distribution of $\mathbf{\Phi} + \boldsymbol{\delta}_{\mathbf{0}}$ under the original stationary distribution

$$\mathbf{P}^{\mathbf{0}}_{\mathbf{\Phi}} = \mathbf{P}_{\mathbf{\Phi}+\boldsymbol{\delta}_{\mathbf{0}}}.$$

(Cross-) Mass transport formula (between two processes)

- The true benefit form stationary framework is the possibility to study the relations between dependent point processes (living on the same probability)...

- THEOREM 22. [MTP for two point processes] Consider two point processes $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}'$ defined on a common stationary framework $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \{\boldsymbol{\theta_t}\}, \mathbf{P})$ and compatible with the flow, having non-null and finite intensities $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}'$, respectively. We denote by $\mathbf{P}^{\mathbf{0}}$ and $\mathbf{P}^{\mathbf{0}'}$ the respective Palm probabilities with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}'$. For any (say non-negative) measurable functions $\boldsymbol{g}$ on $\mathbb{R}^{\boldsymbol{d}} \times \boldsymbol{\Omega}$ (not necessarily compatible with the flow) we have

$$\boldsymbol{\lambda}\mathbf{E}^{\mathbf{0}}\left[\int_{\mathbb{R}^{\boldsymbol{d}}} \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\omega})\, \boldsymbol{\Phi}'(\mathbf{d}\boldsymbol{y})\right] = \boldsymbol{\lambda}'\mathbf{E}^{\mathbf{0}'}\left[\int_{\mathbb{R}^{\boldsymbol{d}}} \boldsymbol{g}(-\boldsymbol{x}, \boldsymbol{\theta_x})\, \boldsymbol{\Phi}(\mathbf{d}\boldsymbol{x})\right]. \tag{18}$$

- Note, we do not assume any particular dependence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}'$. In particular they might be dependent! (e.g. $\boldsymbol{\Phi}' \subset \boldsymbol{\Phi}$ or other way around, etc).

□ Proof: Let $B \in \mathcal{B}$ of unit Lebesgue measure, $|B| = 1$. We have

$$\lambda \mathbf{E}^0 \left[\int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y) \right] = \lambda \int_{\mathbb{R}^d} 1(x \in B) \mathbf{E}^0 \left[\int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y) \right] \mathrm{d}x$$

CLMM for Φ, $f(x,\omega) = 1(x \in B) \int_{\mathbb{R}^d} g(y,\omega)\, \Phi'(\mathrm{d}y)$

$$= \mathbf{E} \left[\int_{\mathbb{R}^d} 1(x \in B) \int_{\mathbb{R}^d} g(y, \theta_x)\, \Phi' \circ \theta_x(\mathrm{d}y)\, \Phi(\mathrm{d}x) \right]$$

by compatibility of Φ' and (13)

$$= \mathbf{E} \left[\int_{\mathbb{R}^d} 1(x \in B) \int_{\mathbb{R}^d} g(y - x, \theta_x)\, \Phi'(\mathrm{d}y)\, \Phi(\mathrm{d}x) \right]$$

Fubini's theorem

$$= \mathbf{E} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1(x \in B) g(y - x, \theta_x)\, \Phi(\mathrm{d}x)\, \Phi'(\mathrm{d}y) \right]$$

$$= \ldots$$

$$\ldots = \mathbf{E}\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1(x \in B) g(y - x, \theta_x)\, \Phi(\mathrm{d}x)\, \Phi'(\mathrm{d}y)\right]$$

$\theta_{-y} \circ \theta_y = \theta_0$ — indentity function

$$= \mathbf{E}\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1(x \in B) g(y - x, \theta_{x-y} \circ \theta_y)\, \Phi \circ \theta_{-y} \circ \theta_y(\mathrm{d}x)\, \Phi'(\mathrm{d}y)\right]$$

CLMM for Φ', $f(x, \omega) = \int_{\mathbb{R}^d} 1(x \in B) g(y - x, \theta_{x-y})\, \Phi \circ \theta_{-y}(\mathrm{d}x)$

$$= \lambda' \int_{\mathbb{R}^d} \mathbf{E}^{0'}\left[\int_{\mathbb{R}^d} 1(x \in B) g(y - x, \theta_{x-y}) \Phi \circ \theta_{-y}(\mathrm{d}x)\right] \mathrm{d}y$$

$$= \ldots$$

$$\ldots = \boldsymbol{\lambda}' \int_{\mathbb{R}^d} \mathbf{E}^{0'} \left[\int_{\mathbb{R}^d} \mathbf{1}(\boldsymbol{x} \in \boldsymbol{B}) \boldsymbol{g}(\boldsymbol{y} - \boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{x}-\boldsymbol{y}}) \boldsymbol{\Phi} \circ \boldsymbol{\theta}_{-\boldsymbol{y}}(\mathrm{d}\boldsymbol{x}) \right] \mathrm{d}\boldsymbol{y}$$

by compatibility of $\boldsymbol{\Phi}$ and (13)

$$= \boldsymbol{\lambda}' \int_{\mathbb{R}^d} \mathbf{E}^{0'} \left[\int_{\mathbb{R}^d} \mathbf{1}(\boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{B}) \boldsymbol{g}(-\boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{x}}) \boldsymbol{\Phi}(\mathrm{d}\boldsymbol{x}) \right] \mathrm{d}\boldsymbol{y}$$

Foubini's theorem

$$= \boldsymbol{\lambda}' \mathbf{E}^{0'} \left[\int_{\mathbb{R}^d} \boldsymbol{g}(-\boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{x}}) \int_{\mathbb{R}^d} \mathbf{1}(\boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{B}) \, \mathrm{d}\boldsymbol{y} \; \boldsymbol{\Phi}(\mathrm{d}\boldsymbol{x}) \right]$$

$\int_{\mathbb{R}^d} \mathbf{1}(\boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{B}) \, \mathrm{d}\boldsymbol{y} = \mathbf{1}$ since $|\boldsymbol{B} - \boldsymbol{y}| = |\boldsymbol{B}| = \mathbf{1}$

$$= \boldsymbol{\lambda}' \mathbf{E}^{0'} \left[\int_{\mathbb{R}^d} \boldsymbol{g}(-\boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{x}}) \, \boldsymbol{\Phi}(\mathrm{d}\boldsymbol{x}) \right].$$

Equivalent form of the mass transport formula

- Remark 23:
  - Let $m(x, y, \omega)$ be a measurable function on $\mathbb{R}^d \times \mathbb{R}^d \times \Omega$ interpreted as the amount of mass sent from x to y on the event ω.
  - We assume that m is compatible with the flow in the following sense

$$m(x, y, \omega) = m(x - t, y - t, \theta_t); \quad \text{for all } x, y, t \in \mathbb{R}^d.$$

  - Then

$$\lambda \mathbf{E}^0 \left[\int_{\mathbb{R}^d} m(0, y, \omega) \, \Phi'(\mathrm{d}y) \right] = \lambda' \mathbf{E}^{0'} \left[\int_{\mathbb{R}^d} m(x, 0, \omega) \, \Phi(\mathrm{d}x) \right] . \quad (19)$$

  - Interpretation: the proportion between the expected total masses, SENT from the typical point of Φ to all points of Φ' and RECEIVED by the typical point of Φ' from all points of Φ. The proportion involves the respective intensities of processes.
  - Proof using (18) with $g(y, \omega) := m(0, y, \omega)$ and by the compatibility of $m(0, -x, \theta_x) = m(x, 0, \omega)$.

Unimodulartiy of Palm probability

- Remark 24: Assume now $\boldsymbol{\Phi}' = \boldsymbol{\Phi}$. Then equation (18) by $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$ one obtains

$$\mathbf{E}^0\left[\int_{\mathbb{R}^d} g(y,\omega)\,\Phi(\mathrm{d}y)\right] = \mathbf{E}^0\left[\int_{\mathbb{R}^d} g(-x,\theta_x)\,\Phi(\mathrm{d}x)\right] \tag{20}$$

for any measurable functions g on $\mathbb{R}^d \times \Omega$, not necessarily compatible with the flow. Equivalently, (19) becomes

$$\mathbf{E}^0\left[\int_{\mathbb{R}^d} m(0,y,\omega)\,\Phi(\mathrm{d}y)\right] = \mathbf{E}^0\left[\int_{\mathbb{R}^d} m(x,0,\omega)\,\Phi(\mathrm{d}x)\right], \tag{21}$$

for any measurable function on $\mathbb{R}^d \times \mathbb{R}^d \times \Omega$ compatible with the flow.

- Observe complete analogy to the mass transport formula for unimodular graphs (cf Lesson 5 on Unimodular Graphs).

Mass transport formula between point process and Lebesgue

- Proposition 25.

$$\lambda \mathbf{E}^{0}\left[\int_{\mathbb{R}^d} m(0,y,\omega)\,\mathrm{d}y\right] = \mathbf{E}\left[\int_{\mathbb{R}^d} m(x,0,\omega)\,\Phi(\mathrm{d}x)\right], \tag{22}$$

  with m translation invariant.

- Proof (Exercise) using Campbell-Little-Mecke-Matthes' formula.
- Apply to Littles' law.

ERGODICITY

Ergodicity bridges probability theory and real-life application

- Law of Large Numbers (LLN), say on for random variables on $\mathbb{R}$):

Applications		**Probability**
observations $X_1, \ldots, X_n \in \mathbb{R}$		random variable $X = X(\omega)$ on some abstract probability space $(\Omega, \mathcal{A}, \mathbf{P})$
mean: $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$	**Ergodicity** $\xrightarrow[n\to\infty]{}$	$\mathbf{E}[f(X)] = \int_\Omega f(X(\omega))\,\mathbf{P}(\mathrm{d}\omega)$; expectation

- Ergodic theory provides precise conditions for the above converge result, thus bridging the gap between the probability theory and real-life applications. It is particularly important in statistics.

Case of spatial data (point processes)

Applications	**Probability**
"homogeneous" pattern $\Phi = \{x_i\}_i$ of points in some observation window $B \subset \mathbb{R}^d$	stationary point process $\Phi(\omega)$ on $(\Omega, \mathcal{A}, \{\theta_t\}, \mathbf{P})$, of intensity λ and Palm probability $\mathbf{P}^0$, modeling the observations;

- Two type of empirical averaging: continuously and discrete (w.r.t. data points), give two types of LLN's...

Continuous LLN in a nutshell

Averaging observations $X(x) = X(x, \Phi) = f(S_x\Phi)$ of Φ from all locations $x \in B$ in the window

Stationary expectation of the observation $X(0) = X(0, \Phi(\omega)) = f(\Phi(\omega))$ of Φ from the origin

$$\frac{1}{|B|}\int_B X(x)\,\mathrm{d}x = \frac{1}{|B|}\int_B f(S_x\Phi)\,\mathrm{d}x \quad \underset{B\nearrow\mathbb{R}^d}{\longrightarrow} \mathbf{E}\left[X(0)\right] = \mathbf{E}\left[f(\Phi)\right]$$

Observe by the invariance of $\mathbf{P}$ that on average

$$\frac{1}{|B|}\mathbf{E}\left[\int_B f(S_x\Phi)\,\mathrm{d}x\right] = \frac{1}{|B|}\int_B \mathbf{E}\left[f(S_x\Phi)\right]\,\mathrm{d}x = \mathbf{E}\left[f(\Phi)\right].$$

Discrete LNN in a nutshell

Averaging observations $X(x_i) = X(x_i, \Phi) = f(S_{x_i}\Phi)$ of Φ from all (discrete) points of Φ $x_i \in \Phi \cap B$ in the window

Palm expectation of the observatio $X(0) = X(0, \Phi(\omega)) = f(\Phi(\omega)$ of Φ from the typical point at the ori gin

$$\frac{1}{\Phi(B)} \sum_{x_i \in \Phi \cap B} X(x_i) = \frac{1}{\Phi(B)} \sum_{x_i \in \Phi \cap B} f(S_{x_i}\Phi) \quad \xrightarrow[B \nearrow \mathbb{R}^d]{} \mathbf{E}^0 [X(0)] = \mathbf{E}^0 [f(\Phi)]$$

Observe by the CLMM theorem that on average

$$\mathbf{E}\left[\sum_{X_i \in \Phi \cap B} f(S_{X_i}\Phi)\right] = \lambda |B| \, \mathbf{E}^0 [f(\Phi)] \quad \text{and} \quad \mathbf{E}[\Phi(B)] = \lambda |B|.$$

Continuum ergodicity

- We will using general ergodic theory approach. They perfectly fit to our stationary framework for point processes $\left(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P}\right)$.
- An event $A \in \mathcal{A}$ is called $(\{\theta_t\}, \mathbf{P})$-invariant (invariant for short) if

$$\mathbf{P}(A \triangle \cap \theta_t A) = 0 \quad \text{for all } t \in \mathbb{R}^d,$$

  where $\triangle$ denotes the symmetric difference: $A \triangle B = (A \cup B) \setminus (A \cap B)$.
- We define invariant σ-algebra:

$$\mathcal{I} := \{A \in \mathcal{A} : A \text{ is invariant}\}\,;$$

  (prove it is indeed a σ-algebra).
- We say $\left(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P}\right)$ is metrically transitive if $\mathcal{I}$ is $\mathbf{P}$-trivial, i.e., if $\forall A \in \mathcal{I}, \mathbf{P}(A) \in \{0, 1\}$.

Invariant events for point process

☐ REMARK 26:

- Consider point process $\mathbf{\Phi}$ on $\left(\mathbf{\Omega}, \boldsymbol{\mathcal{A}}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P}\right)$, compatible with the flow $\{\theta_t\}_{t\in\mathbb{R}^d}$.
- Consider event $A = \{\omega : \mathbf{\Phi}(\omega) \in \mathbf{\Gamma}\}$ for some $\mathbf{\Gamma} \in \boldsymbol{\mathcal{M}}$.
- A is $\mathbf{P}$-invariant iff $\mathbf{1}(S_t\mathbf{\Phi} \in \mathbf{\Gamma}) = \mathbf{1}(\mathbf{\Phi} \in \mathbf{\Gamma}) \quad \mathbf{P}$-a.s. (observing $\mathbf{\Phi} \in \mathbf{\Gamma}$ is $\mathbf{P}$-a.s. invariant with all translations of $\mathbf{\Phi}$).
- Often, we consider events "$\varphi(\mathbf{\Phi}) = a$" for some translation invariant φ and its value a; for example
  - ▹ $\varphi(\mathbf{\Phi}) =$ #infinite components in some translation-invariant graph on $\mathbf{\Phi}$.
- Consequently, if $\mathcal{I}$ (say generated by $\mathbf{\Phi}$) is $\mathbf{P}$-trivial then $\varphi(\mathbf{\Phi})$ is $\mathbf{P}$-a.s. constant.

Averaging observation windows

- A sequence of sets $(B_n)_{n\geq 1}$ in $\mathbb{R}^d$ is said to be a convex averaging sequence if each B_n is bounded Borel and convex set such that

$$B_n \subset B_{n+1}, \quad \forall n$$

and

$$\sup\{r \geq 0 : B_n \text{ contains a ball of radius } r\} \to \infty, \quad n \to \infty.$$

- THEOREM 27. **[Birkhoff's (Individual or Pointwise) Ergodic Theorem]** Let $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ be a stationary framework, $\mathcal{I}$ the invariant σ-algebra, $(B_n)_{n\geq 1}$ a convex averaging sequence in $\mathbb{R}^d$. For a measurable function on $(\Omega, \mathcal{A})$, such that $\mathbf{E}[|f|] < \infty$

$$\lim_{n\to\infty} \frac{1}{|B_n|} \int_{B_n} f\circ\theta_x \, \mathrm{d}x = \mathbf{E}[f \mid \mathcal{I}], \quad \mathbf{P}\text{-a.s.} \tag{23}$$

  where $\mathbf{E}[f \mid \mathcal{I}]$ is the conditional expectation with respect to $\mathcal{I}$.

- COROLLARY 28. Under the assumptions of Theorem 27, if the stationary framework is metrically transitive then (23) holds with $\mathbf{E}[f \mid \mathcal{I}] = \mathbf{E}[f]$.

- Proof: (classical Birkhoff's ergodic theorem) see e.g. Theorem 10 in Kallenberg([2])

[2] *Foundations of modern probability*. Springer, 2002.

- Verifying metrical transitivity is not simple. In what follows we provide an equivalent and a sufficient condition.
- We say stationary framework is ergodic if

$$\lim_{a\to\infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathbf{P}(A_1 \cap \theta_x A_2)\mathrm{d}x = \mathbf{P}(A_1)\mathbf{P}(A_2), \quad \forall A_1, A_2 \in \mathcal{A} \tag{24}$$

- We say it is mixing if

$$\lim_{|x|\to\infty} \mathbf{P}(A_1 \cap \theta_x A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2), \quad \forall A_1, A_2 \in \mathcal{A} \tag{25}$$

□ PROPOSITION 29. For a stationary framework $\left(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \{\boldsymbol{\theta}_t\}_{t\in\mathbb{R}^d}, \mathbf{P}\right)$ the following relations hold true:

$$\text{mixing} \quad \Rightarrow \quad \text{ergodicity} \quad \Leftrightarrow \quad \text{metrical transitivity.} \tag{26}$$

□ Proof:

– Mixing implies ergodicity — Exercise.

– Ergodicity implies metrical transitivity:

▹ Assume that the framework is ergodic. Consider some $\boldsymbol{A} \in \boldsymbol{\mathcal{I}}$.

▹ For any $\boldsymbol{t} \in \mathbb{R}^{\boldsymbol{d}}$, $\mathbf{P}\left(\boldsymbol{A} \triangle \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A}\right) = \mathbf{0}$; and since $\boldsymbol{A} \cap \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A} = \boldsymbol{A} \setminus (\boldsymbol{A} \setminus \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A})$ and $\boldsymbol{A} \setminus \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A} \subset \boldsymbol{A} \triangle \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A}$, then $\mathbf{P}\left(\boldsymbol{A} \cap \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A}\right) = \mathbf{P}\left(\boldsymbol{A}\right) - \mathbf{P}\left(\boldsymbol{A} \setminus \boldsymbol{\theta}_{\boldsymbol{t}} \boldsymbol{A}\right) = \mathbf{P}\left(\boldsymbol{A}\right)$.

▹ On the other hand, we deduce from ergodicity that

$$\lim_{a \to \infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathbf{P}(\boldsymbol{A} \cap \boldsymbol{\theta}_{\boldsymbol{x}} \boldsymbol{A}) \mathrm{d}\boldsymbol{x} = \mathbf{P}(\boldsymbol{A})^2$$

▹ Then $\mathbf{P}\left(\boldsymbol{A}\right) = \mathbf{P}(\boldsymbol{A})^2$, thus $\mathbf{P}\left(\boldsymbol{A}\right) \in \{0, 1\}$. Therefore the invariant $\boldsymbol{\sigma}$-algebra $\boldsymbol{\mathcal{I}}$ is $\mathbf{P}$-tirival, and the framework is consequently metrically transitive.

- Metrical transitivity imples ergodicity:
  - ▹ Assume that the framework is metrically transitive. Let $A_1, A_2 \in \mathcal{A}$.
  - ▹ By Birkhoff's theorem 27 and Corollary 28, we have

$$\lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} 1\left\{\theta_x(\omega) \in A_2\right\} \mathrm{d}x = \mathbf{E}[1\{\omega \in A_2\}] = \mathbf{P}(A_2)$$

▷ Then

$$\mathbf{P}(A_1)\mathbf{P}(A_2) = \mathbf{E}[1\{\omega \in A_1\}] \left(\lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} 1\{\theta_x \omega \in A_2\}\, \mathrm{d}x \right)$$

since $(\lim \dots)$ is $\mathbf{P}$-a.s. constant

$$= \mathbf{E}\left[1\{\omega \in A_1\} \lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} 1\{\theta_x \omega \in A_2\}\, \mathrm{d}x \right]$$

Dominated Convergence Theorem

$$= \lim_{n\to\infty} \mathbf{E}\left[1\{\omega \in A_1\} \frac{1}{(2n)^d} \int_{[-n,n]^d} 1\{\theta_x \omega \in A_2\}\, \mathrm{d}x \right]$$

Fubini's Theorem

$$= \lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{E}\left[1\{\omega \in A_1\} 1\{\theta_x \omega \in A_2\}\right] \mathrm{d}x$$

$$= \lim_{n\to\infty} \frac{1}{(2n)^d} \int_{[-n,n]^d} \mathbf{P}\left(A_1 \cap \theta_x A_2\right) \mathrm{d}x\,,$$

which completes the proof.

Ergodicity and/or mixing for point processes

- Sometimes one says that a stationary point process $\mathbf{\Phi}$, meaning its distribution is ergodic or mixing.
- By this we mean that the canonical space $(\mathbb{M}, \mathcal{M}, \{S_x\}, \mathbf{P_\Phi})$ with the distribution of $\mathbf{\Phi}$ as the probability measure is ergodic or mixing, respectively. The following result simplifies verification of these condition.

□ PROPOSITION 30. Let Φ be a stationary point process with Laplace transform $\mathcal{L} = \mathcal{L}_\Phi$. Then

(i) Φ is ergodic if and only if

$$\lim_{a\to\infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathcal{L}_\Phi(f_1 + S_x f_2)\mathrm{d}x = \mathcal{L}_\Phi(f_1)\mathcal{L}_\Phi(f_2)$$

for any measurable $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}_+$ bounded with bounded support.

(ii) Φ is mixing if and only if

$$\lim_{|x|\to\infty} \mathcal{L}_\Phi(f_1 + S_x f_2) = \mathcal{L}_\Phi(f_1)\mathcal{L}_\Phi(f_2)$$

for any measurable $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}_+$ bounded with bounded support.

□ Proof: Cf. Proposition 12.3.VI in Daley and Vere-Jones, 2007 ([3])

[3] *An introduction to the theory of point processes: volume II: general theory and structure.* Springer, 2007

Poisson process is mixing hence ergodic

- COROLLARY 31. Homogeneous Poisson point process $\boldsymbol{\Phi}$ on $\mathbb{R}^{\boldsymbol{d}}$ is mixing and hence ergodic.
- Proof: Use Proposition 30 (ii).

Discrete ergodicity

- THEOREM 32. Let $(\Omega, \mathcal{A}, \{\theta_t\}_{t\in\mathbb{R}^d}, \mathbf{P})$ be a stationary and ergodic framework, Φ a point process on $\mathbb{R}^d$ compatible with the flow $\{\theta_t\}_{t\in\mathbb{R}^d}$ with finite and non-null intensity λ and Palm probability $\mathbf{P}^0$. Let $(B_n)_{n\geq 1}$ be a convex averaging sequence in $\mathbb{R}^d$. For a measurable function f on $(\Omega, \mathcal{A})$, such that $\mathbf{E}^0[|f|] < \infty$

$$\lim_{n\to\infty} \frac{1}{|B_n|} \int_{B_n} f \circ \theta_x \, \Phi(\mathrm{d}x) = \lambda \mathbf{E}^0[f], \quad \mathbf{P}\text{-a.s.} \tag{27}$$

- COROLLARY 33. Under assumptions of Theorem 32 we have

$$\lim_{n\to\infty} \frac{1}{\Phi(B_n)} \int_{B_n} f \circ \theta_x \, \Phi(\mathrm{d}x) = \mathbf{E}^0[f] \quad \mathbf{P}\text{-a.s.}$$

- Indeed, use (27) with $f = 1$ to observe that $\mathbf{P}$-a.s. $\lim_{n\to\infty} |B_n|/\Phi(B_n) = 1/\lambda$.

Can’t use discrete Birkhoff’s result, except in 1D

- REMARK 34: For the proof of Theorem 32, one might want use a discrete version of Birkhoff’s individual ergodic result.
- This is possible only in one dimension, i.e.; for stationary, ergodic framework with one dimensional flow $\{\theta_t\}_{t\in\mathbb{R}}$.
- The reason is that Palm probability $\mathbf{P}^0$ is invariant with respect to natural discrete point shifts only in dimension $d = 1$. (To be explained.)
- Proof: **[of Theorem 32]** The idea: approximating the discrete sum by the integral of some stochastic process:

$$h(\omega) = \int_{\mathbb{R}^d} g_\epsilon(x) f \circ \theta_x \Phi(\,\mathrm{d}x)$$

with some non-negative, continuous function $g_\epsilon(x)$ with bounded support around the origin and $\int_{\mathbb{R}^d} g_\epsilon(x)\,\mathrm{d}u = 1$. Then use Theorem 27. (continuum Birkhoff’s result). See details in Lecture Notes.