Spatial Modelling (NMTP438)

May 4, 2023

Contents

Contents **1**

1. Spatial models on lattices **2**
1.1 Markov random fields 2
1.2 Gaussian models 9
1.3 Spatial autocorrelation 12

2. Random fields **14**
2.1 Basic definitions 14
2.2 Variogram and autocovariance function 17

3. Random measures and point processes **25**
3.1 Locally finite measures 26
3.2 Random measures 28
3.3 Simple point processes 29
3.4 Poisson point process 33
3.5 Moment measures 36
3.6 Palm distribution 39

4. Spatial point processes **42**
4.1 Stationary spatial point processes 42
4.2 Point process characteristics 44
4.3 Point process models 48
4.4 Finite point processes with density 54
4.5 Inhomogeneous point processes 61

5. Marked point processes **66**
5.1 Basic definitions 66
5.2 Marking models 69
5.3 Multivariate point processes 71
5.4 Processes with quantitative marks 75

6. Appendix **78**
6.1 Gaussian distribution 78
6.2 Bessel functions 79
6.3 Measure theory 80

References **80**

1. Spatial models on lattices

1.1 Markov random fields

Let $L \subseteq \mathbb{R}^d$ be a finite non-empty set that will be referred to as a *lattice*. It could be a regular grid, e.g. $L = \{1, 2, \ldots, N\}^d$ (for $N = 10$ and $d = 2$ see Figure 1 left). This could be useful when studying a type of spatial data in the form of images. In this case, the elements of L (called *sites*) represent pixels ($d = 2$) or voxels ($d = 3$). The sites can also represent some geographic regions (an example is shown in Figure 1 right). Spatial data given by the observations related to these regions are known as *areal unit data*.

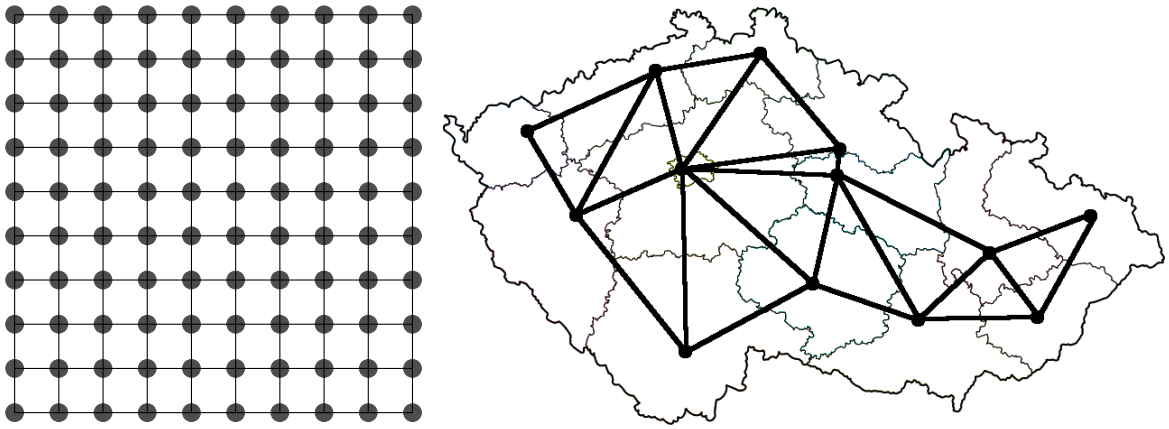

Figure 1. Two examples of lattices in the plane. Left: regular square grid 10×10. Right: capitals of 13 regions of the Czech Republic, two sites are connected if the regions share a common border.

Definition 1. A family of random variables $\{Z_i : i \in L\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a *random field* on the lattice L. The state space of the random variables Z_i will be denoted by $S \subseteq \mathbb{R}$.

The simplest random field is obtained from independent random variables. It serves as the basis for the definition of more interesting random fields that allow spatial dependence.

Definition 2. A random field $\{Z_i : i \in L\}$ is said to be a *(spatial) strict white noise* if the random variables Z_i are centred, independent and identically distributed.

The number of sites will be denoted by $n = |L|$. Sometimes it is convenient to order the sites by numbers $1, \ldots, n$. Then the random field $\boldsymbol{Z} = \{Z_i : i \in L\}$ can be viewed as a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$. Its distribution is given by the density $p(\boldsymbol{z})$ w.r.t. some σ-finite measure ν^n on S^n. Here, ν^n denotes the n-th power of the measure ν on S in the sense of the product of measures. As a measure ν, we usually consider the counting measure (discrete states) or the Lebesgue measure (continuous states). Taking into account the isomorphism between L and $\{1, \ldots, n\}$, we use the same symbol $p(\boldsymbol{z})$ for the density on S^L, $\boldsymbol{z} = \{z_i : i \in L\} \in S^L$, and by ν^L we denote the corresponding product measure on S^L. For $A \subseteq L$, we write shortly $\boldsymbol{z}_A = \{z_j : j \in A\}$. For disjoint sets $A, B \subseteq L$, we denote $\boldsymbol{z}_A \boldsymbol{w}_B = \{y_j : j \in A \cup B\}$, where

$$y_j = \begin{cases} z_j, & \text{for } j \in A, \\ w_j, & \text{for } j \in B. \end{cases}$$

Let $\sim$ be a symmetric relation on $L \times L$. We say that two sites are *neighbours* if they are in this relation. For simplicity, we use the following notation: $\partial i = \{j \in L : j \sim i, j \neq i\}$, $-i = L \setminus \{i\}$ for $i \in L$, and $-A = L \setminus A$ for $A \subseteq L$. The set L and the relation $\sim$ generate an unoriented graph where the set of nodes is L and two nodes $i, j \in L$ are connected by an edge if and only if $i \sim j$. On the other hand, every unoriented graph determines the system of neighbours. Figure 1 shows two examples of such unoriented graphs: a regular square lattice with the nearest-neighbour relation and an irregular lattice with the geographic adjacency neighbourhood.

Definition 3. A set $A \subseteq L$ is called a *clique* w.r.t. $\sim$ if $i \sim j$ for any $i, j \in A$, $i \neq j$. We denote by $\mathcal{C} = \{C \subseteq L : C \text{ is a clique}\}$ the system of all cliques. Moreover, put $\mathcal{C}_i = \{C \in \mathcal{C} : i \in C\}$ for $i \in L$.

Remark 1. The empty set and all singletons $\{i\}$ are cliques.

To simplify the notation, we use the symbol p not only for joint density but also for marginal and conditional densities. The type of density will be clear from the arguments of p.

Definition 4. A random field $\{Z_i : i \in L\}$ is called *Markov* w.r.t. $\sim$ if the conditional distribution of $Z_i \mid \boldsymbol{Z}_{-i}$ coincides with the conditional distribution of $Z_i \mid \boldsymbol{Z}_{\partial i}$ for every $i \in L$. In the language of conditional densities, it means that $p(z_i|\boldsymbol{z}_{-i}) = p(z_i|\boldsymbol{z}_{\partial i})$ for ν^L-a.a. $\boldsymbol{z} = \{z_i : i \in L\} \in S^L$ satisfying $p(\boldsymbol{z}) > 0$. This condition is known as the *local Markov property*. The densities $p(z_i|\boldsymbol{z}_{\partial i})$ are called *local characteristics*.

Remark 2. If $\partial i = \emptyset$, the local Markov property means that the conditional distribution of $Z_i \mid \boldsymbol{Z}_{-i}$ is the same as the distribution of Z_i. In other words, Z_i and $\boldsymbol{Z}_{-i}$ are independent.

Remark 3. The local Markov property is equivalent to the condition $p(z_i|\boldsymbol{z}_A) = p(z_i|\boldsymbol{z}_{\partial i})$ for ν^L-a.a. $\boldsymbol{z}$ satisfying $p(\boldsymbol{z}_A) > 0$ and for any $i \in L$ and $\partial i \subseteq A \subseteq -i$.

Example: It is obvious that the strict white noise is an example of a Markov random field.

Example: A Markov chain $\{Z_1, \dots, Z_n\}$ is a one-dimensional Markov random field ($d = 1$ and $L = \{1, \dots, n\}$) w.r.t. the relation $i \sim j \Leftrightarrow |i - j| \leq 1$ (see Exercise class).

Definition 5. Let A, B, C be pairwise disjoint subsets of L. We say that $\boldsymbol{Z}_A$ and $\boldsymbol{Z}_B$ are *conditionally independent* given $\boldsymbol{Z}_C$ if

$$p(\boldsymbol{z}_A\boldsymbol{z}_B|\boldsymbol{z}_C) = p(\boldsymbol{z}_A|\boldsymbol{z}_C)\, p(\boldsymbol{z}_B|\boldsymbol{z}_C)$$

for ν^L-a.a. $\boldsymbol{z}$ satisfying $p(\boldsymbol{z}_C) > 0$.

Lemma 1. *Let $\{Z_i : i \in L\}$ be a random field and let A, B, C be pairwise disjoint subsets of L. Then $\boldsymbol{Z}_A$ and $\boldsymbol{Z}_B$ are conditionally independent given $\boldsymbol{Z}_C$ if and only if*

$$p(\boldsymbol{z}_A|\boldsymbol{z}_B\boldsymbol{z}_C) = p(\boldsymbol{z}_A|\boldsymbol{z}_C)$$

for ν^L-a.a. $\boldsymbol{z}$ satisfying $p(\boldsymbol{z}_B\boldsymbol{z}_C) > 0$, which happens if and only if

$$p(\boldsymbol{z}_B|\boldsymbol{z}_A\boldsymbol{z}_C) = p(\boldsymbol{z}_B|\boldsymbol{z}_C)$$

for ν^L-a.a. $\boldsymbol{z}$ satisfying $p(\boldsymbol{z}_A\boldsymbol{z}_C) > 0$.

Proof: By simple manipulation, it follows from the definition of the conditional density that

$$p(\boldsymbol{z}_A|\boldsymbol{z}_B\boldsymbol{z}_C) = \frac{p(\boldsymbol{z}_A\boldsymbol{z}_B\boldsymbol{z}_C)}{p(\boldsymbol{z}_B\boldsymbol{z}_C)} = \frac{p(\boldsymbol{z}_A\boldsymbol{z}_B|\boldsymbol{z}_C)}{p(\boldsymbol{z}_B|\boldsymbol{z}_C)}.$$

The second equation is obtained by interchanging A and B.

□

Hence, the local Markov property from Definition 4 is equivalent to the fact that Z_i and $\boldsymbol{Z}_{-(\{i\}\cup\partial i)}$ are conditionally independent given $\boldsymbol{Z}_{\partial i}$.

Remark 4. Instead of an unoriented graph, we can work with a directed acyclic graph. Its nodes are sites from L. If there is a directed edge from i to j, then i is referred to as a *parent* of j and j is referred to as a *child* of i. We say that $\{Z_i : i \in L\}$ is a *Bayesian network*, if Z_i and $\boldsymbol{Z}_{-(\{i\}\cup\mathrm{de}(i)\cup\mathrm{pa}(i))}$ are conditionally independent given $\boldsymbol{Z}_{\mathrm{pa}(i)}$ for every $i \in L$. Here, $\mathrm{de}(i)$ is the set of all descendants of i (i.e. all nodes that can be reached by a direct path from i) and $\mathrm{pa}(i)$ is the set of parents of i. Since the graph is acyclic, we have $\mathrm{pa}(i) \subseteq -(\{i\} \cup \mathrm{de}(i))$. The joint density is then given by the relation

$$p(\boldsymbol{z}) = \prod_{i \in L} p(z_i \mid \boldsymbol{z}_{\mathrm{pa}(i)}).$$

Bayesian networks are widely used in the field of artificial intelligence.

The number of neighbours is usually much smaller than the number of sites. While the full conditional distributions $Z_i \mid \boldsymbol{Z}_{-i}$ can be very complicated, the local characteristics depend only on the neighbours of a given site. The structure of the random field is then simpler thanks to the local Markov property. This is used in MCMC methods, where the steps in Gibbs sampler are typically much easier.

We know that for Markov chains, the transition probabilities (or transition densities) together with the initial distribution determine the joint distribution of the chain. We can ask when the system of local characteristics determines the joint density of the random field. As opposed to the case of Markov chains, we can not expect that the local characteristics may be chosen arbitrarily so that the joint density exists and is unique (see Exercise class). The following theorem states that the conditional distributions well define the joint distribution if they are derived from a joint density of a particular form.

Definition 6. We say that a random field with the joint density $p(\boldsymbol{z})$ satisfies the *positivity condition* if $p(\boldsymbol{z}) > 0$ for all $\boldsymbol{z} \in S^L$.

Theorem 2. *(Hammersley–Clifford theorem) A random field satisfying the positivity condition is Markov if and only if there exist functions $g_C : S^C \to \mathbb{R}^+$ such that*

$$p(\boldsymbol{z}) = \prod_{C \in \mathcal{C}} g_C(\boldsymbol{z}_C), \quad \boldsymbol{z} \in S^L.$$

Proof: The simpler implication is from right to left. If the density has the required form then

$$p(z_i|\boldsymbol{z}_{-i}) = \frac{p(\boldsymbol{z})}{p(\boldsymbol{z}_{-i})} = \frac{\prod_{C \in \mathcal{C}} g_C(\boldsymbol{z}_C)}{\int_S \prod_{C \in \mathcal{C}_i} g_C(w_i \boldsymbol{z}_{-i \cap C}) \prod_{C \in \mathcal{C} \setminus \mathcal{C}_i} g_C(\boldsymbol{z}_C)\, \nu(\mathrm{d}w_i)} \propto \prod_{C \in \mathcal{C}_i} g_C(\boldsymbol{z}_C),$$

where the symbol $\propto$ means that $p(z_i|\boldsymbol{z}_{-i})$ is proportional to $\prod_{C \in \mathcal{C}_i} g_C(\boldsymbol{z}_C)$. Since $i \in C$ implies $C \subseteq \{i\} \cup \partial i$, we also have

$$p(z_i|\boldsymbol{z}_{\partial i}) = \frac{p(\boldsymbol{z}_{\{i\} \cup \partial i})}{p(\boldsymbol{z}_{\partial i})} \propto \prod_{C \in \mathcal{C}_i} g_C(\boldsymbol{z}_C).$$

Now assume that the random field is Markov. Fix some configuration $\boldsymbol{w} \in S^L$ and define

$$\Psi_A(\boldsymbol{z}_A) = -\log p(\boldsymbol{z}_A \boldsymbol{w}_{-A}) \quad \text{and} \quad \Phi_A(\boldsymbol{z}_A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \Psi_B(\boldsymbol{z}_B), \quad A \subseteq L.$$

From Lemma 3 it follows that $\Psi_A(\boldsymbol{z}_A) = \sum_{B \subseteq A} \Phi_B(\boldsymbol{z}_B)$. For the density $p(\boldsymbol{z})$ we get

$$p(\boldsymbol{z}) = \exp\{-\Psi_L(\boldsymbol{z}_L)\} = \exp\left\{ - \sum_{B \subseteq L} \Phi_B(\boldsymbol{z}_B) \right\} = \prod_{B \subseteq L} g_B(\boldsymbol{z}_B),$$

where $g_B(\boldsymbol{z}_B) = \exp\{-\Phi_B(\boldsymbol{z}_B)\}$. It remains to show that if B is not a clique, then $g_B(\boldsymbol{z}_B) = 1$, which is equivalent to $\Phi_B(\boldsymbol{z}_B) = 0$. If B is not a clique, then there exist two sites $i, j \in B$ such that $i \not\sim j$. For $A \subseteq B \setminus \{i, j\}$ let us denote $A_i = A \cup \{i\}$, $A_j = A \cup \{j\}$, $A_{ij} = A \cup \{i, j\}$. Then

$$\begin{aligned}
\Phi_B(\boldsymbol{z}_B) &= \sum_{A \subseteq B} (-1)^{|B|-|A|} \Psi_A(\boldsymbol{z}_A) \\
&= \sum_{A \subseteq B \setminus \{i,j\}} (-1)^{|B|-|A|} \left[\Psi_A(\boldsymbol{z}_A) - \Psi_{A_i}(\boldsymbol{z}_{A_i}) - \Psi_{A_j}(\boldsymbol{z}_{A_j}) + \Psi_{A_{ij}}(\boldsymbol{z}_{A_{ij}}) \right] \\
&= \sum_{A \subseteq B \setminus \{i,j\}} (-1)^{|B|-|A|} \log \frac{p(\boldsymbol{z}_{A_i} \boldsymbol{w}_{-A_i}) p(\boldsymbol{z}_{A_j} \boldsymbol{w}_{-A_j})}{p(\boldsymbol{z}_{A_{ij}} \boldsymbol{w}_{-A_{ij}}) p(\boldsymbol{z}_A \boldsymbol{w}_{-A})} \\
&= \sum_{A \subseteq B \setminus \{i,j\}} (-1)^{|B|-|A|} \log \frac{p(z_i|\boldsymbol{z}_A \boldsymbol{w}_{-A_i}) p(w_i|\boldsymbol{z}_{A_j} \boldsymbol{w}_{-A_{ij}})}{p(z_i|\boldsymbol{z}_{A_j} \boldsymbol{w}_{-A_{ij}}) p(w_i|\boldsymbol{z}_A \boldsymbol{w}_{-A_i})} \\
&= \sum_{A \subseteq B \setminus \{i,j\}} (-1)^{|B|-|A|} \log \frac{p(z_i|\boldsymbol{z}_A \boldsymbol{w}_{-A_{ij}}) p(w_i|\boldsymbol{z}_A \boldsymbol{w}_{-A_{ij}})}{p(z_i|\boldsymbol{z}_A \boldsymbol{w}_{-A_{ij}}) p(w_i|\boldsymbol{z}_A \boldsymbol{w}_{-A_{ij}})} = 0.
\end{aligned}$$

The positivity condition assures that the conditional densities are well-defined. We have used the relation $p(z_i|\boldsymbol{z}_A\boldsymbol{w}_{-A_i}) = p(z_i|\boldsymbol{z}_A\boldsymbol{w}_{-A_{ij}}) = p(z_i|\boldsymbol{z}_{A_j}\boldsymbol{w}_{-A_{ij}})$, which follows from the local Markov property (see Remark 3) because $i \not\sim j$.

Note that $\Phi_\emptyset = \Psi_\emptyset = -\log p(\boldsymbol{w})$ and $g_\emptyset = \mathrm{e}^{-\Phi_\emptyset} = p(\boldsymbol{w})$ is a normalizing constant, which is generally difficult to compute.

□

Lemma 3. *(Möbius inversion formula) Let Φ and Ψ be real-valued functions defined on the power set of a finite set L. Then*

$$\Phi(A) = \sum_{B\subseteq A} (-1)^{|A|-|B|}\Psi(B) \quad \forall A \subseteq L \qquad \Longleftrightarrow \qquad \Psi(A) = \sum_{B\subseteq A} \Phi(B) \quad \forall A \subseteq L.$$

Proof: First we show the implication from left to right:

$$\sum_{B\subseteq A} \Phi(B) = \sum_{B\subseteq A}\sum_{D\subseteq B} (-1)^{|B|-|D|}\Psi(D) = \sum_{D\subseteq A}\sum_{C\subseteq A\setminus D} (-1)^{|C|}\Psi(D) = \Psi(A)$$

because $\sum_{C\subseteq A\setminus D}(-1)^{|C|}$ is distinct from zero only if $A \setminus D = \emptyset$. This can be seen from the identity $\sum_{k=0}^{n}\binom{n}{k}(-1)^k = 0$ for $n \in \mathbb{N}$, which follows from the binomial theorem.

The reverse implication is shown analogously:

$$\sum_{B\subseteq A} (-1)^{|A|-|B|}\Psi(B) = \sum_{B\subseteq A}\sum_{D\subseteq B} (-1)^{|A|-|B|}\Phi(D) = \sum_{D\subseteq A}\sum_{C\subseteq A\setminus D} (-1)^{|A|-|D|-|C|}\Phi(D) = \Phi(A).$$

□

Definition 7. The distribution of a random field $\{Z_i : i \in L\}$ with density

$$p(\boldsymbol{z}) = \exp\left\{-\sum_{C\in\mathcal{C}} \Phi_C(\boldsymbol{z}_C)\right\}, \quad \boldsymbol{z} \in S^L, \tag{1}$$

is called the *Gibbs distribution*. The random field $\{Z_i : i \in L\}$ is then called the *Gibbs random field*. It plays an important role in statistical mechanics where its density is usually written as $p(\boldsymbol{z}) = \frac{1}{\mathcal{Z}}\exp\{-E/T\}$. The term E is interpreted as the total energy (*Hamiltonian*) of the configuration $\boldsymbol{z}$. It is given as the sum of potentials $V_C(\boldsymbol{z}_C)$ over all non-empty cliques,

$$E = \sum_{C\in\mathcal{C}\setminus\{\emptyset\}} V_C(\boldsymbol{z}_C).$$

The parameter T is a constant called the *temperature*, and $\mathcal{Z}$ is a normalizing constant called the *partition function*,

$$\mathcal{Z} = \int_{S^L} \exp\left\{-\frac{1}{T}\sum_{C\in\mathcal{C}\setminus\{\emptyset\}} V_C(\boldsymbol{z}_C)\right\} \nu^L(\mathrm{d}\boldsymbol{z}).$$

In this context, $V_C(\boldsymbol{z}_C) = T\Phi_C(\boldsymbol{z}_C)$ is the potential of the configuration $\boldsymbol{z}_C$, $T\Psi_C(\boldsymbol{z}_C)$ is the energy of $\boldsymbol{z}_C$, and the partition function is $\mathcal{Z} = \mathrm{e}^{\Phi_\emptyset}$.

The Hammersley–Clifford theorem says that every Markov random field satisfying the positivity condition is a Gibbs random field where $g_C(\boldsymbol{z}_C) = \exp\{-\Phi_C(\boldsymbol{z}_C)\}$. We have already mentioned that the conditional distributions do not determine the joint distribution. Therefore, we can not specify the conditional distributions directly. However, we may instead construct them by choice of the potential functions Φ_C. Since the expression

$$p(\boldsymbol{z}) = \prod_{C\in\mathcal{C}} g_C(\boldsymbol{z}_C)$$

is not unique, also the potentials are not uniquely determined. For given $g_C(\boldsymbol{z}_C)$, $C \neq \emptyset$, the normalizing constant $g_\emptyset = \mathrm{e}^{-\Phi_\emptyset}$ is already uniquely determined. It is given by

$$g_\emptyset = \left(\int_{S^L} \prod_{C \in \mathcal{C} \setminus \{\emptyset\}} g_C(\boldsymbol{z}_C)\, \nu^L(\mathrm{d}\boldsymbol{z}) \right)^{-1}$$

provided that the integral is finite and positive. Therefore, it is enough to specify the functions g_C or Φ_C (or equivalently V_C and the temperature T) for non-empty cliques C. We can write the local characteristics in terms of the potentials in the following way:

$$p(z_i|\boldsymbol{z}_{-i}) \propto \exp\left\{ -\sum_{C \in \mathcal{C}_i} \Phi_C(\boldsymbol{z}_C) \right\}. \tag{2}$$

Let us give some examples of Markov random fields.

Example: The simplest non-trivial situation is when the state space has only two elements. Consider $S = \{0, 1\}$. In image analysis, the sites of L represent pixels, $z_i = 1$ usually denotes black colour and $z_i = 0$ white colour of the pixel i. When dealing with areal unit data, we may put $z_i = 1$ if an event of interest occurs in the region i and $z_i = 0$ otherwise. Define (for $C \neq \emptyset$)

$$\Phi_C(\boldsymbol{z}_C) = \begin{cases} -\beta, & \text{if } C = \{i, j\},\ i \sim j \text{ and } z_i = z_j, \\ 0, & \text{otherwise,} \end{cases}$$

where $\beta \geq 0$ is a parameter (in statistical mechanics it is referred to as the *inverse temperature*). Then we get the joint density (w.r.t. the counting measure)

$$p(\boldsymbol{z}) = \frac{1}{\mathcal{Z}(\beta)} \exp\left\{ \beta \sum_{\{i,j\}: i \sim j} \mathbf{1}_{[z_i = z_j]} \right\},$$

where

$$\mathcal{Z}(\beta) = \sum_{\boldsymbol{z} \in \{0,1\}^L} \exp\left\{ \beta \sum_{\{i,j\}: i \sim j} \mathbf{1}_{[z_i = z_j]} \right\} = \mathrm{e}^{\Phi_\emptyset}$$

is the partition function (here also sometimes called the *partition sum*), which is finite because S is finite. The local characteristics satisfy

$$p(z_i|\boldsymbol{z}_{\partial i}) = \mathbb{P}(Z_i = z_i \mid \boldsymbol{Z}_{\partial i} = \boldsymbol{z}_{\partial i}) = \frac{\exp\left\{\beta \sum_{j \in \partial i} \mathbf{1}_{[z_j = z_i]}\right\}}{\exp\left\{\beta \sum_{j \in \partial i} \mathbf{1}_{[z_j = 1]}\right\} + \exp\left\{\beta \sum_{j \in \partial i} \mathbf{1}_{[z_j = 0]}\right\}}.$$

For $\beta = 0$, every configuration has the same probability. It means that the values 0 and 1 are independently and uniformly randomly assigned to the sites. For $\beta > 0$, the configurations with attractive forces among the neighbours are more probable. The probability that a given pixel is black, given that it has k black neighbours and $m - k$ white neighbours, is $\mathrm{e}^{\beta k}/(\mathrm{e}^{\beta k} + \mathrm{e}^{\beta(m-k)})$. For $\beta \to \infty$, one state prevails in the whole configuration. Figure 2 shows simulated realizations of random fields on a regular grid 25×25 for different choices of the parameter β. Theoretically, it is also possible to consider $\beta < 0$. Then the neighbouring pixels of a black site will more likely tend to be white.

This random field is known as the *Ising model* [8]. It has been proposed as a mathematical model of ferromagnetism in statistical mechanics. The value Z_i represents the atomic spin at the site i, usually $+1$ is used for the upward orientation and -1 is used for the downward orientation.

The Ising model (as well as other Markov random fields) can be extended to an infinite lattice. The problem is that $\mathcal{C}$ may be uncountable and the expression (1) does not have sense. However, we can still consider local characteristics (if every site has finitely many neighbours) of the form (2). A Gibbs distribution can be defined so that its conditional distributions are determined by given local characteristics. The question is the existence and uniqueness of such distribution. It turns out that the

Gibbs distribution on an infinite lattice exists (if the state space is compact), but generally it is not unique. In this case, we speak about a *phase transition*. In particular, consider a regular planar grid $L_N = \{-N, \dots, N\}^2$. We are interested in the behaviour of the Ising model for $N \to \infty$, i.e. $L_N \nearrow \mathbb{Z}^2$. There exists a critical value $\beta_c = \log(1+\sqrt{2}) \doteq 0.881$ (analytically computed by Onsager [12]), at which the phase transition occurs. For $\beta \leq \beta_c$, the Gibbs distribution is unique, whereas for $\beta > \beta_c$, it is not. In the case $\beta > \beta_c$, the values at the boundary of lattice L_N influence the marginal distribution of $Z_{(0,0)}$, when $N \to \infty$. It means that there are long-range interactions in the configuration. From the physics point of view, the particle is magnetized.

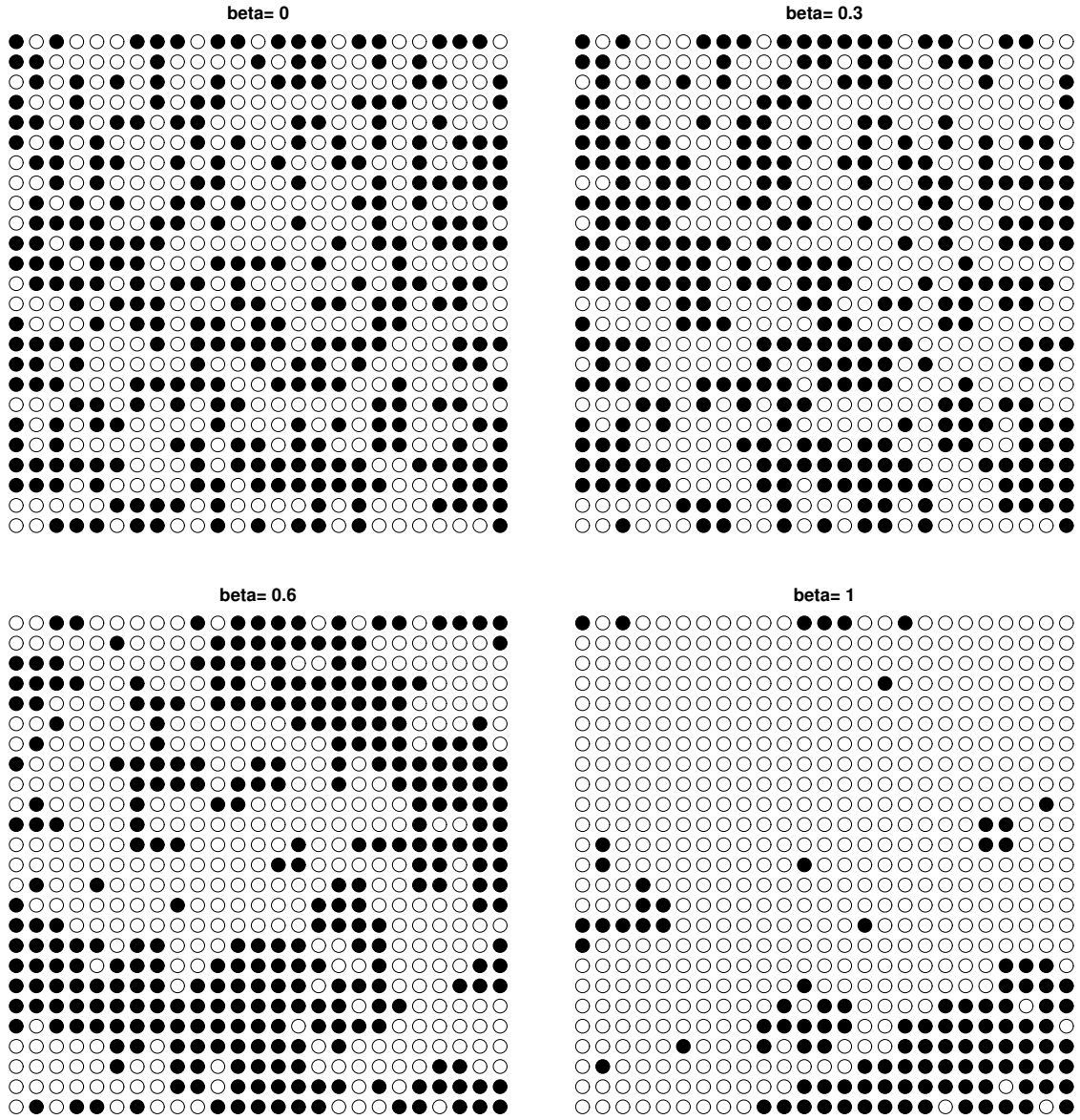

Figure 2. Simulations of the Ising model on the rectangular lattice 25×25 for $\beta \in \{0; 0.3; 0.6; 1\}$.

In the definition of Φ_C, we can allow non-zero values for one-point cliques: $\Phi_{\{i\}}(z_i) = -\beta h_i$. Then the total energy is

$$E = -\sum_{i \in L} h_i - \sum_{\{i,j\}: i \sim j} \mathbf{1}_{[z_i = z_j]}$$

and the joint density is $p(\boldsymbol{z}) \propto \exp\{-E/T\} = \exp\{-\beta E\}$. The values h_i can be interpreted as the influence of an external magnetic field. Further possible generalization is to admit the dependence of the interaction strength on the sites or the values of the field in these sites. It means that β is then a function of i, j, z_i and z_j. We can also consider the interactions of higher order than just pair interactions.

Even though the local structure is simple, the Ising model is already quite complex. The joint density contains a computationally demanding normalizing constant. Therefore, direct simulations from the model are practically unfeasible and one has to exploit MCMC or other methods.

Example: The *Potts model* is a multicolour generalization of the Ising model. Let

$$S = \{0, 1, \dots, n_c - 1\},$$

where $n_c > 2$ denotes the number of colours. The potential functions Φ_C are defined exactly as in the case of the Ising model. The local characteristics become

$$p(z_i|\boldsymbol{z}_{\partial i}) = \mathbb{P}(Z_i = z_i \mid \boldsymbol{Z}_{\partial i} = \boldsymbol{z}_{\partial i}) = \frac{\exp\left\{\beta \sum_{j\in\partial i} \mathbf{1}_{[z_j=z_i]}\right\}}{\sum_{k=0}^{n_c-1} \exp\left\{\beta \sum_{j\in\partial i} \mathbf{1}_{[z_j=k]}\right\}}, \quad z_i \in \{0, \dots, n_c - 1\}.$$

In statistical mechanics, the Potts model is a model of interacting spins.

Example: In the Potts model, the arrangement of the state space S does not play any role. There is equal strength of attraction for all colours. However, we can imagine that 0 corresponds to white colour, $n_c - 1$ to black colour, and other states to different grey shades. Then it would be desirable to take the ordering into account. The attraction could be larger for similar colours. For example, we can consider the model with the following local characteristics:

$$\begin{aligned} p(z_i|\boldsymbol{z}_{\partial i}) = \mathbb{P}(Z_i = z_i \mid \boldsymbol{Z}_{\partial i} = \boldsymbol{z}_{\partial i}) &= \binom{n_c - 1}{z_i} \pi(\boldsymbol{z}_{\partial i})^{z_i} \big(1 - \pi(\boldsymbol{z}_{\partial i})\big)^{n_c-1-z_i} \\ &= \binom{n_c - 1}{z_i} \big(1 - \pi(\boldsymbol{z}_{\partial i})\big)^{n_c-1} \exp\left\{z_i \log \frac{\pi(\boldsymbol{z}_{\partial i})}{1 - \pi(\boldsymbol{z}_{\partial i})}\right\}, \end{aligned}$$

where $\pi(\boldsymbol{z}_{\partial i})$ are prescribed probabilities. It means that $Z_i \mid \boldsymbol{Z}_{\partial i} = \boldsymbol{z}_{\partial i}$ has a binomial distribution with parameters $n_c - 1$ and $\pi(\boldsymbol{z}_{\partial i})$. If we assume that $\pi(\boldsymbol{z}_{\partial i})$, $i \in L$, satisfy

$$\log \frac{\pi(\boldsymbol{z}_{\partial i})}{1 - \pi(\boldsymbol{z}_{\partial i})} = -\beta_i - \sum_{j\in\partial i} \beta_{ij} z_j, \tag{3}$$

then we get the Markov random field with potentials $\Phi_{\{i\}}(z_i) = \beta_i z_i - \log \binom{n_c-1}{z_i}$, $\Phi_{\{i,j\}}(z_i z_j) = \beta_{ij} z_i z_j$ and $\Phi_C(\boldsymbol{z}_C) = 0$ for $|C| > 2$. The relation (3) is analogous to the logistic regression model. Therefore, such model for a random field $\{Z_i : i \in L\}$ is known as an *autologistic model*.

Example: Let us move to the countable state space: $S = \mathbb{N}_0$. In practice, this situation appears when we deal with counts data, e.g. disease-incidence counts in some regions. Consider a model where these counts have the Poisson distribution with intensity $\lambda(\boldsymbol{z}_{\partial i})$ that depends on the counts in neighbouring sites:

$$p(z_i|\boldsymbol{z}_{\partial i}) = \mathbb{P}(Z_i = z_i \mid \boldsymbol{Z}_{\partial i} = \boldsymbol{z}_{\partial i}) = \exp\{-\lambda(\boldsymbol{z}_{\partial i})\} \frac{\lambda(\boldsymbol{z}_{\partial i})^{z_i}}{z_i!} = \exp\{-\lambda(\boldsymbol{z}_{\partial i}) + z_i \log \lambda(\boldsymbol{z}_{\partial i}) - \log z_i!\}.$$

These local characteristics are called *auto-Poisson*. In order to get a Gibbs distribution of $\{Z_i : i \in L\}$, we require

$$\log \lambda(\boldsymbol{z}_{\partial i}) = -\beta_i - \sum_{j\in\partial i} \beta_{ij} z_j.$$

Moreover, we have to make sure that the normalizing constant is finite:

$$\mathcal{Z} = \sum_{\boldsymbol{z}\in S^L} \exp\left\{-\sum_{i\in L} (\log z_i! + \beta_i z_i) - \sum_{\{i,j\}\in\mathcal{C}} \beta_{ij} z_i z_j\right\} < \infty.$$

It can be shown that the sum is finite if and only if $\beta_{ij} \geq 0$ for all $i, j \in L$ such that $i \sim j$, $i \neq j$ (see Exercise class). The condition $\beta_{ij} \geq 0$ means that large values of neighbours of the site i result in a higher probability of smaller values in i. This restricts the practical application of the auto-Poisson model.

1.2 Gaussian models

Gaussian random fields are the most frequently used examples of random fields with continuous states. Let $\{Z_i : i \in L\}$ have an n-dimensional Gaussian distribution with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$, i.e. the joint density has the form

$$p(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right\}, \quad \boldsymbol{z} \in \mathbb{R}^L.$$

The inverse of the covariance matrix is commonly referred to as the *precision matrix*. We denote it by $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ and its elements by q_{ij}, $i, j \in L$. Then the conditional distributions $Z_i \mid \boldsymbol{Z}_{-i} = \boldsymbol{z}_{-i}$ are Gaussian with mean

$$\mu_i - \frac{1}{q_{ii}} \sum_{j \neq i} q_{ij}(z_j - \mu_j)$$

and variance $1/q_{ii}$ (cf. Theorem 92). This suggests a convenient choice of neighbourhood relation that ensures the local Markov property. If we put $i \sim j \Leftrightarrow q_{ij} \neq 0$, then we get a Markov random field because the conditional distribution $Z_i \mid \boldsymbol{Z}_{-i}$ does not depend on Z_j for j that is not a neighbour of i.

Definition 8. Let $\sim$ be a neighbourhood relation. A random field $\boldsymbol{Z} = \{Z_i : i \in L\}$ is called a *Gaussian Markov random field* if it has n-dimensional normal distribution with positive definite covariance matrix $\boldsymbol{\Sigma}$ satisfying $q_{ij} \neq 0 \Leftrightarrow i \sim j$, where q_{ij} are elements of the matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$.

The following theorem states that the random variables corresponding to non-neighbouring sites are conditionally independent.

Theorem 4. *Let $\boldsymbol{Z} = \{Z_i : i \in L\}$ be a Gaussian Markov random field. Then, for $i \neq j$, Z_i and Z_j are conditionally independent given $\boldsymbol{Z}_{-\{i,j\}}$ if and only if $i \not\sim j$ (i.e. $q_{ij} = 0$).*

Proof: We could simply use Lemma 1 and the knowledge of the distribution of $Z_i \mid \boldsymbol{Z}_{-i}$. However, let us proceed directly from the definition. Recall that the conditional independence of Z_i and Z_j given $\boldsymbol{Z}_{-\{i,j\}}$ means that $p(z_i, z_j|\boldsymbol{z}_{-\{i,j\}}) = p(z_i|\boldsymbol{z}_{-\{i,j\}})p(z_j|\boldsymbol{z}_{-\{i,j\}})$. Since the joint distribution of $\boldsymbol{Z}$ is Gaussian, also the conditional densities are Gaussian densities. From the relation for the joint density

$$p(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}\sum_{k,l}(z_k - \mu_k)q_{kl}(z_l - \mu_l)\right\}$$

we deduce that

$$p(z_i, z_j \mid \boldsymbol{z}_{-\{i,j\}}) \propto \exp\Bigg\{-(z_i-\mu_i)(z_j-\mu_j)q_{ij} - \frac{1}{2}(z_i-\mu_i)^2 q_{ii} - \sum_{k\neq i,j}(z_i-\mu_i)(z_k-\mu_k)q_{ik} - \frac{1}{2}(z_j-\mu_j)^2 q_{jj} - \sum_{l\neq i,j}(z_j-\mu_j)(z_l-\mu_l)q_{jl}\Bigg\}. \quad (4)$$

If $q_{ij} = 0$, then

$$p(z_i, z_j|\boldsymbol{z}_{-\{i,j\}}) \propto \exp\left\{-\frac{1}{2}(z_i-\mu_i)^2 q_{ii} - (z_i-\mu_i)\sum_{k\neq i,j}(z_k-\mu_k)q_{ik}\right\} \times \exp\left\{-\frac{1}{2}(z_j-\mu_j)^2 q_{jj} - (z_j-\mu_j)\sum_{l\neq i,j}(z_l-\mu_l)q_{jl}\right\},$$

where, apart from the normalizing constant, the first term is $p(z_i \mid \boldsymbol{z}_{-\{i,j\}})$ and the second term is $p(z_j|\boldsymbol{z}_{-\{i,j\}})$.

Conversely, if $p(z_i, z_j|\boldsymbol{z}_{-\{i,j\}}) = p(z_i|\boldsymbol{z}_{-\{i,j\}})p(z_j|\boldsymbol{z}_{-\{i,j\}})$, the right-hand side of (4) does not contain the term with $(z_i - \mu_i)(z_j - \mu_j)$. Hence, $q_{ij} = 0$.

□

The simplest example of a Gaussian Markov random field is obtained when any two distinct sites are not neighbours. It means that both $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ are diagonal matrices. A particular case is $\Sigma = \sigma^2 \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.

Definition 9. We say that a random field $\{Z_i : i \in L\}$ is a *Gaussian (spatial) white noise* if the random variables Z_i form a strict white noise and have normal distribution $N(0, \sigma^2)$.

Another simple example of a Gaussian Markov random field is a Gaussian autoregressive sequence of order 1 (see Exercise class). More information on the theory and applications of Gaussian Markov random fields can be found in the monograph [15].

The Gaussian Markov random fields were defined by their joint density, determined by the mean $\boldsymbol{\mu}$ and the precision matrix $\boldsymbol{Q}$. Alternatively, we can specify the full conditional distributions $Z_i \mid \boldsymbol{Z}_{-i}$. Obviously, we are not allowed to choose the conditional distributions arbitrarily (see Exercise class).

Lemma 5. *(Brook's lemma) Let p be the density of an n-dimensional random vector. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ satisfying $p(\boldsymbol{x}), p(\boldsymbol{y}) > 0$, the following relation holds:*

$$\begin{aligned}\frac{p(\boldsymbol{x})}{p(\boldsymbol{y})} &= \prod_{i=1}^{n} \frac{p(x_i|x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{p(y_i|x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)} \\ &= \prod_{i=1}^{n} \frac{p(x_i|y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n)}{p(y_i|y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n)}.\end{aligned}$$

Proof: From the relation

$$\frac{p(x_n|x_1, \ldots, x_{n-1})\, p(x_1, \ldots, x_{n-1})}{p(y_n|x_1, \ldots, x_{n-1})\, p(x_1, \ldots, x_{n-1})} = \frac{p(x_1, \ldots, x_{n-1}, x_n)}{p(x_1, \ldots, x_{n-1}, y_n)},$$

it follows that

$$p(\boldsymbol{x}) = \frac{p(x_n|x_1, \ldots, x_{n-1})}{p(y_n|x_1, \ldots, x_{n-1})}\, p(x_1, \ldots, x_{n-1}, y_n).$$

Now we can similarly express the last term on the right-hand side:

$$p(x_1, \ldots, x_{n-1}, y_n) = \frac{p(x_{n-1}|x_1, \ldots, x_{n-2}, y_n)}{p(y_{n-1}|x_1, \ldots, x_{n-2}, y_n)}\, p(x_1, \ldots, x_{n-2}, y_{n-1}, y_n).$$

In this way, we inductively obtain the desired equation. The second equation is obtained analogously by adjusting the formula

$$p(\boldsymbol{x}) = \frac{p(x_1|x_2, \ldots, x_n)}{p(y_1|x_2, \ldots, x_n)}\, p(y_1, x_2 \ldots, x_n).$$

□

Brook's lemma gives instructions on how to get the joint density from the full conditional densities. We fix some $\boldsymbol{y}$ and apply Lemma 5 to compute $p(\boldsymbol{x})$ up to the normalizing constant $p(\boldsymbol{y})$. The normalizing constant is determined so that the integral of the joint density is equal to one. If we obtain a function that is not integrable, then our system of conditional densities does not lead to a proper joint density. The system of full conditional densities that gives a proper joint density is called consistent.

Definition 10. The system of conditional densities $\{p(z_i|\boldsymbol{z}_{-i}) : i \in L\}$ is called *consistent* if there exists a joint density $p(\boldsymbol{z})$ of a random field $\{Z_i : i \in L\}$ such that $p(z_i|\boldsymbol{z}_{-i})$ are the corresponding full conditional densities. The random field $\{Z_i : i \in L\}$ is called the *conditional autoregressive model* which is abbreviated as CAR.

A particular CAR model is obtained by choosing Gaussian conditional densities. Let $\boldsymbol{B} = (b_{ij})_{i,j \in L}$ be a zero-diagonal matrix ($b_{ii} = 0$ for all $i \in L$). Let τ_i^2, $i \in L$, be positive parameters. Denote by $\boldsymbol{D}$ a diagonal matrix with elements $d_{ii} = \tau_i^2$ in its diagonal. Consider the system of conditional distributions such that $Z_i \mid \boldsymbol{Z}_{-i}$ has a Gaussian distribution with mean $\sum_{j \in L} b_{ij} Z_j$ and variance τ_i^2. Enumerate the sites of L by $1, \ldots, n$ and fix $\boldsymbol{y} = o$ as the null vector in $\mathbb{R}^n$. Then by Lemma 5 we have

$$\frac{p(\boldsymbol{z})}{p(o)} = \exp\left\{ -\sum_{i=1}^{n} \frac{z_i^2}{2\tau_i^2} + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{b_{ij}}{\tau_i^2} z_i z_j \right\} = \exp\left\{ -\sum_{i=1}^{n} \frac{z_i^2}{2\tau_i^2} + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{b_{ij}}{\tau_i^2} z_i z_j \right\}, \quad \boldsymbol{z} \in \mathbb{R}^n.$$

Comparing these expressions, we get the necessary conditions

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2} \quad \text{for all } i, j. \tag{5}$$

Under these conditions,

$$p(\boldsymbol{z}) = p(o) \exp\left\{ -\sum_{i=1}^{n} \frac{z_i^2}{2\tau_i^2} + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{b_{ij}}{2\tau_i^2} z_i z_j \right\} = p(o) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} z_i z_j \right\},$$

where q_{ij} are the elements of the matrix $\boldsymbol{Q} = \boldsymbol{D}^{-1}(\boldsymbol{I} - \boldsymbol{B})$. From this we see that under the assumption of the positive definiteness of $\boldsymbol{Q}$ (which is equivalent to the positive definiteness of $\boldsymbol{I} - \boldsymbol{B}$), the density $p(\boldsymbol{z})$ is the density of a centred n-dimensional normal distribution with precision matrix $\boldsymbol{Q}$. The normalizing constant $p(o)$ is equal to $(2\pi)^{-n/2}\sqrt{\det \boldsymbol{Q}}$. The matrix $\boldsymbol{Q}$ is symmetric due to (5). Note that the matrix $\boldsymbol{B}$ is not symmetric unless all τ_i^2 are equal. Our choice of n conditional normal distributions is consistent under the condition (5) and if $\boldsymbol{I} - \boldsymbol{B}$ is assumed to be positive definite.

Definition 11. Consider a matrix $\boldsymbol{B} = (b_{ij})_{i,j \in L}$ such that $b_i i = 0$ for all $i \in L$ and $\boldsymbol{I} - \boldsymbol{B}$ is positive definite. Let τ_i^2, $i \in L$, be positive parameters satisfying (5). Then there exists a random field $\{Z_i : i \in L\}$ for which $Z_i \mid \boldsymbol{Z}_{-i}$ has a Gaussian distribution $N(\sum_{j \in L} b_{ij} Z_j, \tau_i^2)$. We will refer to it as the *Gaussian CAR model*.

The Gaussian CAR model can be viewed as the Gaussian Markov random field w.r.t. the relation $i \sim j \Leftrightarrow q_{ij} \neq 0$. The whole system can be briefly written as $\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{Z} + \boldsymbol{\varepsilon}$, which is equivalent to the expression $(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{Z} = \boldsymbol{\varepsilon}$. Since $\boldsymbol{Z}$ has a centred n-dimensional normal distribution with the covariance matrix $\boldsymbol{Q}^{-1} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{D}$, the vector $\boldsymbol{\varepsilon}$ has a centred n-dimensional normal distribution with the covariance matrix $\boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B})^{\mathrm{T}}$. It means that the elements of $\boldsymbol{\varepsilon}$ are not independent. For simplicity, we have considered the centred case. However, we can easily incorporate the mean $\boldsymbol{\mu}$ in the model:

$$\boldsymbol{Z} = \boldsymbol{\mu} + \boldsymbol{B}(\boldsymbol{Z} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}. \tag{6}$$

Besides the approach using conditional distributions, it is possible to consider spatial Gaussian models where the random field is specified simultaneously. This approach is motivated by the generalization of autoregressive sequences from stochastic processes in time. The relation (6) can be rewritten as

$$Z_i - \mu_i = \sum_{j \in L} b_{ij}(Z_j - \mu_j) + \varepsilon_i, \quad i \in L.$$

While for the CAR models, $\boldsymbol{Z}$ induces the distribution of $\boldsymbol{\varepsilon}$, we now let $\boldsymbol{\varepsilon}$ induce the distribution of $\boldsymbol{Z}$.

Definition 12. Let $\boldsymbol{\varepsilon} = \{\varepsilon_i : i \in L\}$ be a Gaussian white noise. We assume that $\boldsymbol{B}$ is a matrix whose diagonal elements are zero and that $(\boldsymbol{I} - \boldsymbol{B})^{-1}$ exists. The matrix $\boldsymbol{B}$ is not necessarily symmetric. We define the random field $\boldsymbol{Z} = \{Z_i : i \in L\}$ by the relation

$$(\boldsymbol{I} - \boldsymbol{B})(\boldsymbol{Z} - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}. \tag{7}$$

We speak about the *simultaneous autoregressive model* with mean $\boldsymbol{\mu}$ and abbreviate it by SAR.

Clearly, $\mathbb{E}\boldsymbol{Z} = \boldsymbol{\mu}$ and the covariance matrix of $\boldsymbol{Z}$ is

$$\mathbb{E}(\boldsymbol{Z} - \boldsymbol{\mu})(\boldsymbol{Z} - \boldsymbol{\mu})^{\mathrm{T}} = \sigma^2(\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{I} - \boldsymbol{B}^{\mathrm{T}})^{-1}.$$

Since $\boldsymbol{Z}$ is a linear transformation of $\boldsymbol{\varepsilon}$, the distribution of $\boldsymbol{Z}$ is normal. The relation (7) coincides with (6). The difference is that now we considered (in analogy with the time series autoregressive model) $\boldsymbol{\varepsilon}$ to be a white noise. The elements of matrix $\boldsymbol{B}$ determine the spatial dependence. If $b_{ij} = 0$, then Z_i and Z_j are conditionally independent given $\boldsymbol{Z}_{-\{i,j\}}$. The joint density has the form

$$p(\boldsymbol{z}) = \frac{\det(\boldsymbol{I} - \boldsymbol{B})}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{z} - \boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{B}^{\mathrm{T}})(\boldsymbol{I} - \boldsymbol{B})(\boldsymbol{z} - \boldsymbol{\mu}) \right\}, \quad \boldsymbol{z} \in \mathbb{R}^n.$$

Note that $\operatorname{cov}(\boldsymbol{\varepsilon}, \boldsymbol{Z}) = \sigma^2(\boldsymbol{I} - \boldsymbol{B}^{\mathrm{T}})^{-1}$, and so (as opposed to a causal autoregressive sequence in time) the ε_i may depend on the autoregressors.

Similarly, we can consider generalizations of the moving average (MA) models or ARMA models to spatial models.

Definition 13. Let $\boldsymbol{\varepsilon} = \{\varepsilon_i : i \in L\}$ be a Gaussian white noise, let $\boldsymbol{\mu}$ be a vector of means and let $\boldsymbol{E}$ be a matrix of real coefficients. The random field $\boldsymbol{Z} = \{Z_i : i \in L\}$ given by $\boldsymbol{Z} = \boldsymbol{\mu} + (\boldsymbol{I} - \boldsymbol{E})\boldsymbol{\varepsilon}$ is called a *spatial moving average* and is abbreviated by SMA. If we moreover consider a zero-diagonal matrix $\boldsymbol{B}$ such that $\boldsymbol{I} - \boldsymbol{B}$ is invertible, then we define a *SARMA model* by

$$(\boldsymbol{I} - \boldsymbol{B})(\boldsymbol{Z} - \boldsymbol{\mu}) = (\boldsymbol{I} - \boldsymbol{E})\boldsymbol{\varepsilon}.$$

For the SMA model, $\boldsymbol{Z}$ has an n-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\sigma^2(\boldsymbol{I} - \boldsymbol{E})(\boldsymbol{I} - \boldsymbol{E}^{\mathrm{T}})$. The covariance matrix for the SARMA model has the form $\sigma^2(\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{I} - \boldsymbol{E})(\boldsymbol{I} - \boldsymbol{E}^{\mathrm{T}})(\boldsymbol{I} - \boldsymbol{B}^{\mathrm{T}})^{-1}$.

1.3 Spatial autocorrelation

A random field $\{Z_i : i \in L\}$ is used for modelling dependent data. Typically the variables at neighbouring sites appear to be correlated (either positively or negatively). Our aim is to quantify this *spatial autocorrelation*. We introduce some statistical measures that are used to determine the degree of autocorrelation in spatial data.

Since closer observations are usually more related than distant observations, we would like to take into account the "closeness" of individual sites. For Markov random fields, this is expressed by the relation $\sim$. We assign a non-negative weight w_{ij} to each pair (i, j) of sites from L. We only require that $w_{ij} = 0$ if $i = j$ or $i \not\sim j$. The weights w_{ij} are called the *spatial proximity weights* or *spatial connectivity weights*. The simplest example is given by the *binary weights*

$$w_{ij} = \begin{cases} 1, & \text{if } i \sim j,\ i \neq j, \\ 0, & \text{otherwise.} \end{cases}$$

Another popular choice is obtained by the *normalized binary weights*

$$w_{ij} = \begin{cases} \frac{1}{|\partial i|}, & j \in \partial i, \\ 0, & j \notin \partial i, \end{cases}$$

where $|\partial i|$ denotes the cardinality of ∂i. Notice that the weights do not have to be symmetric, i.e. the relation $w_{ij} = w_{ji}$ may not hold. Denote by $\boldsymbol{W}$ the matrix with entries w_{ij}, $i, j \in L$. In case of the normalized binary weights, this matrix is stochastic (the row sums are 1) if we assume that each site has at least one neighbour.

First consider binary random fields, i.e. the state space $S = \{0, 1\}$ has only two elements. The states often represent whether the event of interest occurred at site i ($Z_i = 1$) or not ($Z_i = 0$). In image analysis, 1 typically corresponds to black colour of the pixel i and 0 to white colour.

Definition 14. Let $\boldsymbol{Z} = \{Z_i, i \in L\}$ be a random field with the state space $S = \{0, 1\}$. Define the *black-black join count statistic* as

$$BB = \frac{1}{2} \sum_{i \in L} \sum_{j \in L} w_{ij} Z_i Z_j$$

and the *black-white join count statistic* as

$$BW = \frac{1}{2} \sum_{i \in L} \sum_{j \in L} w_{ij} (Z_i - Z_j)^2.$$

Remark 5. In these statistics we make a weighted sum over those pairs of neighbours that are both black ($Z_i = Z_j = 1$) or one is black and the second is white ($Z_i = 1$ and $Z_j = 0$ or the other way round). For the binary weights, BB is directly equal to the number of neighbours that are both black. Similarly, BW is the number of neighbours with different colours.

Assume that n sites have m values 1 (black) and $n-m$ values 0 (white). No spatial autocorrelation in the data can be understood so that black and white colours are assigned to the sites completely at random. There are several ways how to make such an assignment. The most natural are the *binomial* and *hypergeometric* samplings. For the binomial sampling, we assume that each site gets a colour independently of the other sites, it is black with probability π (which we would estimate from the data as m/n) or white with probability $1-\pi$. Then $\mathbb{P}(Z_i = 1) = \pi$ for every $i \in L$ and the number of black sites has the binomial distribution with parameters n and π. Obviously, $\mathbb{E}BB = \frac{1}{2}\pi^2 w$ and $\mathbb{E}BW = \pi(1-\pi)w$, where $w = \mathbf{1}^{\mathrm{T}}\boldsymbol{W}\mathbf{1} = \sum_{i\in L}\sum_{j\in L} w_{ij}$ and $\mathbf{1} = (1,\dots,1)^{\mathrm{T}}$. The hypergeometric sampling is appropriate if we want to guarantee that the number of black sites is exactly m. From n sites we randomly select (without replacement) m sites that obtain black colour. The remaining $n-m$ sites are white. In this case, $\mathbb{P}(Z_i = 1) = m/n$, $\mathbb{P}(Z_i = 1, Z_j = 1) = \frac{m(m-1)}{n(n-1)}$, and $\mathbb{P}((Z_i - Z_j)^2 = 1) = \frac{2m(n-m)}{n(n-1)}$. Hence,

$$\mathbb{E}BB = \frac{w}{2}\cdot\frac{m(m-1)}{n(n-1)} \qquad \text{and} \qquad \mathbb{E}BW = w\frac{m(n-m)}{n(n-1)}.$$

The formulas for the variance are a little bit more complicated in both models (see [3]). If the statistics BB computed from the data is considerably larger than $\mathbb{E}BB$, it indicates the presence of positive autocorrelation. Neighbouring sites tend to have the same colour. On the contrary, large values of BW correspond to negative autocorrelation because neighbouring sites have rather distinct colours. For realizations of the Ising model from Figure 2, the values of BW as β increases are 588, 489, 352 and 133. At the same time, the theoretical expectation for the case $\beta = 0$ (no spatial autocorrelation – independent uniform assignment of 0 and 1 to the sites) is 600. With increasing β, there is stronger positive autocorrelation.

For continuous data, the similarity of variables at sites i and j is often measured by $(Z_i - \bar{Z})(Z_j - \bar{Z})$ or $(Z_i - Z_j)^2$. If we sum all these contributions over the pairs of neighbours and normalize by the variance estimate, we get the following indices.

Definition 15. Let $\{Z_i : i \in L\}$ be a random field with constant mean $\mathbb{E}Z_i = \mu$ and constant variance $\operatorname{var} Z_i = \sigma^2$. The *Moran index* is defined as

$$I = \frac{n}{w}\frac{\sum_{i\in L}\sum_{j\in L} w_{ij}(Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i\in L}(Z_i - \bar{Z})^2},$$

where $\bar{Z} = \frac{1}{n}\sum_{i\in L} Z_i$. The *Geary index* is given by the formula

$$c = \frac{n-1}{2w}\frac{\sum_{i\in L}\sum_{j\in L} w_{ij}(Z_i - Z_j)^2}{\sum_{i\in L}(Z_i - \bar{Z})^2}.$$

For the normalized binary weights (under the assumption that every site has at least one neighbour), we have $w = n$ and the normalizing constant for the Moran index is equal to one.

We will consider two different assumptions that both correspond to no spatial autocorrelation:

1. normality assumption: the random field $\boldsymbol{Z}$ is obtained by independent random variables with normal distribution $N(\mu, \sigma^2)$;
2. randomization assumption: each of $n!$ permutations of observed values at n sites is equally probable.

Lemma 6. *Let $\mathbb{E}_g$ and $\mathbb{E}_r$ denote the expectation under the normality assumption and the randomization assumption, respectively. Then*

$$\mathbb{E}_g I = \mathbb{E}_r I = -\frac{1}{n-1} \quad \textit{and} \quad \mathbb{E}_g c = \mathbb{E}_r c = 1.$$

Proof: Denote $Y_i = Z_i - \bar{Z}$, $M_2 = \sum_{i\in L} Y_i^2$, $R = \sum_{i\in L}\sum_{j\in L} w_{ij}Y_iY_j$, and $V = \sum_{i\in L}\sum_{j\in L} w_{ij}(Z_i - Z_j)^2$.

Obviously, $\mathbb{E}_g Y_i = 0$, $\mathbb{E}_g Y_i^2 = \mathbb{E}_g(Z_i - \bar{Z})^2 = (n-1)\sigma^2/n$, $\mathbb{E}_g Y_iY_j = \mathbb{E}_g(Z_i - \bar{Z})(Z_j - \bar{Z}) = -\sigma^2/n$ and $\mathbb{E}_g(Z_i - Z_j)^2 = 2\sigma^2$ for $i \neq j$. Recalling that $w_{ii} = 0$ we obtain

$$\mathbb{E}_g M_2 = \sum_{i\in L}\mathbb{E}_g Y_i^2 = (n-1)\sigma^2,$$

$$\mathbb{E}_g R = \sum_{i\in L}\sum_{j\in L} w_{ij}\mathbb{E}_g Y_iY_j = -\frac{\sigma^2}{n}w,$$

$$\mathbb{E}_g V = \sum_{i\in L}\sum_{j\in L} w_{ij}\mathbb{E}_g(Z_i - Z_j)^2 = 2\sigma^2 w.$$

Since the index I remains unchanged when we multiply the Z_i by a non-zero constant, M_2 and I are independent under the normality assumption by Theorem 93. Therefore,

$$\frac{n}{w}\mathbb{E}_g R = \mathbb{E}_g I M_2 = \mathbb{E}_g I \cdot \mathbb{E}_g M_2 = (n-1)\sigma^2 \mathbb{E}_g I,$$

which gives $\mathbb{E}_g I = -\frac{1}{n-1}$. Similarly, Theorem 93 guarantees the independence of M_2 and c. This leads to

$$\frac{n-1}{2w}\mathbb{E}_g V = \mathbb{E}_g c M_2 = \mathbb{E}_g c \cdot \mathbb{E}_g M_2 = (n-1)\sigma^2 \mathbb{E}_g c,$$

and thus $\mathbb{E}_g c = 1$.

Under the randomization assumption, the values $\bar{Z} = \bar{z}$ and $M_2 = \sum_{i\in L}(z_i - \bar{z})^2 = m_2$ are constant (every permutation leads to the same value of the mean and sample variance). For $i \neq j$, we have

$$\mathbb{E}_r Y_i Y_j = \sum_{k\in L}\sum_{l\in L: l\neq k} \frac{1}{n(n-1)}(z_k - \bar{z})(z_l - \bar{z}) = -\frac{1}{n(n-1)}\sum_{k\in L}(z_k - \bar{z})^2 = -\frac{m_2}{n(n-1)}$$

and

$$\mathbb{E}_r (Z_i - Z_j)^2 = \sum_{k\in L}\sum_{l\in L: l\neq k} \frac{1}{n(n-1)}(z_k - z_l)^2 = \frac{1}{n(n-1)}\sum_{k\in L}\sum_{l\in L}(y_k - y_l)^2 = \frac{2m_2}{n-1},$$

where $y_i = z_i - \bar{z}$ and we used that $\sum_{i\in L} y_i = 0$. This implies

$$\mathbb{E}_r I = \frac{n}{w m_2}\sum_{i\in L}\sum_{j\in L} w_{ij}\mathbb{E}_r Y_i Y_j = -\frac{1}{n-1},$$
$$\mathbb{E}_r c = \frac{n-1}{2w m_2}\sum_{i\in L}\sum_{j\in L} w_{ij}\mathbb{E}_r (Z_i - Z_j)^2 = 1.$$

□

We have shown that the expectations under the normality and randomization assumptions coincide for both Moran and Geary index. However, the variances are distinct (see [3]). The interpretation of Moran and Geary statistics is the following: if $I > \mathbb{E}I$ or $c < \mathbb{E}c$, the site has a tendency to be connected to the site with a similar value of the field, i.e. there is a positive spatial autocorrelation. Conversely, for $I < \mathbb{E}I$ or $c > \mathbb{E}c$, the values at two neighbouring sites are more likely to be dissimilar.

The assumption of constant mean and variance of the random field is important. Otherwise, the values at neighbouring sites could be similar not due to positive spatial autocorrelation but because they are independent realizations from distributions with similar expectations. Similarly, the values at distant sites could appear distinct because the mean of the random field is changing.

2. Random fields

By a random field, we understand a stochastic process with d-dimensional index set D. In this section, we consider that D is connected and has a positive d-dimensional Lebesgue measure. The basic definitions and propositions are analogous to the one-dimensional case. For $d = 1$ they can be found in [13].

2.1 Basic definitions

Definition 16. Let D be a fixed subset of $\mathbb{R}^d$. A *random field* is a collection of real random variables $\{Z(x) : x \in D\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Finite-dimensional distributions of the random field are described by the distribution functions

$$F_{x_1,\dots,x_n}(t_1,\dots,t_n) = \mathbb{P}(Z(x_1) \leq t_1, \dots, Z(x_n) \leq t_n), \quad t_1,\dots,t_n \in \mathbb{R}, \tag{8}$$

where $n \in \mathbb{N}$ and $x_1,\dots,x_n \in D$. The family of finite-dimensional distributions uniquely determines the distribution of $\{Z(x) : x \in D\}$.

Definition 17. We say that the system of distribution functions $\{F_{x_1,\dots,x_n} : n \in \mathbb{N}, x_1, \dots, x_n \in D\}$ is *projective* if for any $n \in \mathbb{N}$, $x_1, \dots, x_n \in D$, $t_1, \dots, t_n \in \mathbb{R}$ and a permutation $i_1, \dots, i_n$ of the numbers $1, \dots, n$ we have

$$F_{x_1,\dots,x_n}(t_1, \dots, t_n) = F_{x_{i_1},\dots,x_{i_n}}(t_{i_1}, \dots, t_{i_n})$$

and

$$F_{x_1,\dots,x_n}(t_1, \dots, t_n) = \lim_{t_{n+1}\to\infty} F_{x_1,\dots,x_n,x_{n+1}}(t_1, \dots, t_n, t_{n+1}).$$

Clearly, the distribution functions of finite-dimensional distributions of the random field form a projective system. Conversely, we have the following result.

Theorem 7. *(Daniell–Kolmogorov existence theorem) Let $\{F_{x_1,\dots,x_n} : n \in \mathbb{N}, x_1, \dots, x_n \in D\}$ be a projective system of distribution functions. Then there exists a random field $\{Z(x) : x \in D\}$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that (8) holds for any $n \in \mathbb{N}$, $x_1, \dots, x_n \in D$.*

Proof: [19], Theorem I.10.3 or [9], Theorem 6.16.

□

The most studied are Gaussian random fields, for which all their finite-dimensional distributions are Gaussian.

Definition 18. A random field $\{Z(x) : x \in D\}$ is called a *Gaussian random field* if the random vector $(Z(x_1), \dots, Z(x_n))^{\mathrm{T}}$ has n-dimensional normal distribution for every $n \in \mathbb{N}$ and $x_1, \dots, x_n \in D$.

Remark 6. The distribution of every Gaussian random field is completely determined by its mean $\mu(x) = \mathbb{E}Z(x)$ and autocovariance function $C(x, y) = \operatorname{cov}(Z(x), Z(y))$, $x, y \in D$.

Three realizations of Gaussian random fields with constant mean and different choices of autocovariance functions $C(x, y)$ are shown in Figure 3.

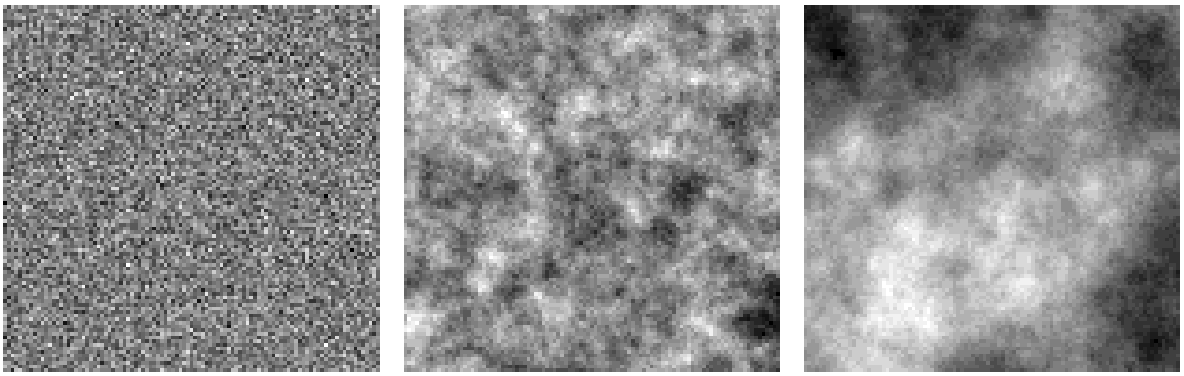

Figure 3. Simulated realizations of Gaussian random fields in the square $D = [0, 10]^2$. Constant mean and variance ($\mu(x) = 0$ and $\operatorname{var} Z(x) = 1$ for each $x \in D$) and three different autocovariance functions were considered. Left: there are no correlations ($C(x, y) = \mathbf{1}_{[x=y]}$). Middle: stronger correlations ($C(x, y) = \mathrm{e}^{-2\|x-y\|}$). Right: the strongest correlations ($C(x, y) = \mathrm{e}^{-0.1\|x-y\|}$).

In practice, we only observe one realization $\boldsymbol{z}$ of a random field $\boldsymbol{Z} = \{Z(x) : x \in D\}$ in finitely many points $x_1, \dots, x_n$. In order to make some statistical inference, we would need to impose further assumptions on the random field $\boldsymbol{Z}$.

Definition 19. We say that a random field $\{Z(x) : x \in D\}$ is *strict(ly) stationary*, if the finite-dimensional distributions of random vectors $(Z(x_1), \dots, Z(x_n))^{\mathrm{T}}$ and $(Z(x_1 + h), \dots, Z(x_n + h))^{\mathrm{T}}$ are equal for every $n \in \mathbb{N}$, $x_1, \dots, x_n \in D$, and $h \in \mathbb{R}^d$ such that $x_1 + h, \dots, x_n + h \in D$.

A random field with finite second moments is called *weak(ly) stationary* if it has a constant mean ($\mathbb{E}Z(x) = \mu$ for all $x \in D$) and its autocovariance function $C(x, y) = \operatorname{cov}(Z(x), Z(y))$ is translation-invariant, i.e. $C(x + h, y + h) = C(x, y)$ for all $x, y \in D$ and $h \in \mathbb{R}^d$ satisfying $x + h, y + h \in D$. In this case, $C(x, y) = C(x - y)$ for all $x, y \in D$ and with a slight abuse of notation we use C also for the function of one argument in $\mathbb{R}^d$.

If only the condition on the autocovariance function is satisfied (the expectation is not necessarily constant), then the random field is *covariance stationary*.

Remark 7. A strictly stationary random field with finite second moments is weakly stationary. For Gaussian random fields, weak stationarity implies strict stationarity. Realizations in Figure 3 are examples of strictly stationary Gaussian random fields.

We will need one other kind of stationarity.

Definition 20. We say that a random field $\{Z(x) : x \in D\}$ is *intrinsic(ally) stationary* if for each $x, y \in D$ we have $\mathbb{E}(Z(x) - Z(y)) = 0$ and $\operatorname{var}(Z(x) - Z(y))$ is a function of $x - y$.

Remark 8. For a weak stationary random field, it follows that $\mathbb{E}(Z(x) - Z(y)) = \mu - \mu = 0$ and

$$\operatorname{var}(Z(x) - Z(y)) = \operatorname{var} Z(x) + \operatorname{var} Z(y) - 2\operatorname{cov}(Z(x), Z(y)) = 2\big(C(o) - C(x - y)\big). \tag{9}$$

It means that every weak stationary random field is also intrinsic stationary. However, the converse is not true. For example, for $d = 1$, the Wiener process is intrinsic stationary ($\operatorname{var}(Z(x + h) - Z(x)) = |h|$) but it is not weak stationary ($\operatorname{var} Z(x) = |x|$).

Stationarity expresses translation-invariance of distributions or moments. In our spatial case, we can also consider rotation-invariance.

Definition 21. A random field $\{Z(x) : x \in D\}$ is called *strict(ly) isotropic* if the finite-dimensional distributions of random vectors $(Z(x_1), \dots, Z(x_n))^{\mathrm{T}}$ and $(Z(\mathcal{O}x_1), \dots, Z(\mathcal{O}x_n))^{\mathrm{T}}$ coincide for each $n \in \mathbb{N}$, $x_1, \dots, x_n \in D$, and rotation $\mathcal{O}$ around the origin such that $\mathcal{O}x_1, \dots, \mathcal{O}x_n \in D$.

A random field with finite second moments is *weak(ly) isotropic* if for every $x, y \in D$ and rotation $\mathcal{O}$ around the origin satisfying $\mathcal{O}x, \mathcal{O}y \in D$ we have $\mathbb{E}Z(x) = \mathbb{E}Z(\mathcal{O}x)$ and $\operatorname{cov}(Z(x), Z(y)) = \operatorname{cov}(Z(\mathcal{O}x), Z(\mathcal{O}y))$.

Remark 9. A strictly isotropic random field with finite second moments is weakly isotropic. For Gaussian random fields, weak isotropy implies strict isotropy. Realizations in Figure 3 come from stationary and isotropic Gaussian random fields.

Definition 22. A random field is called *strictly (or weakly) motion-invariant* if it is both strictly (or weakly) stationary and strictly (or weakly) isotropic.

For a weakly motion-invariant random field $\{Z(x) : x \in \mathbb{R}^d\}$, we have $C(x, y) = C(\|x - y\|)$ for any $x, y \in \mathbb{R}^d$. Again we slightly abuse the notation and use C both for the function of two arguments x and y and the function of one argument $\|x - y\| \in \mathbb{R}^+$ as well.

Definition 23. We say that a random field $\{Z(x) : x \in D\}$ is L_2*-continuous* or *mean square continuous* at $x \in D$ if $\mathbb{E}(Z(x + h) - Z(x))^2 \to 0$ for $\|h\| \to 0+$. The field is L_2-continuous if it is L_2-continuous at each point $x \in D$.

Remark 10. It is good to realize that L_2-continuity does not mean continuity of realizations of the random field.

The Kolmogorov–Chentsov theorem gives the conditions for the existence of a sample continuous modification.

Definition 24. We say that a random field $\{\tilde{Z}(x) : x \in D\}$ is a *modification* of a random field $\{Z(x) : x \in D\}$ if $\mathbb{P}(Z(x) = \tilde{Z}(x)) = 1$ for every $x \in D$.

Theorem 8. *(Kolmogorov–Chentsov theorem) Let $\{Z(x) : x \in D\}$ be a random field, where $D = [a_1, b_1] \times \cdots \times [a_d, b_d]$ is a bounded rectangle. Suppose that there are positive constants α, β, C such that*

$$\mathbb{E}|Z(x) - Z(y)|^\alpha \le C\|x - y\|^{d+\beta}$$

for all $x, y \in D$. Then there exists a modification $\{\tilde{Z}(x) : x \in D\}$ such that the mapping $x \mapsto \tilde{Z}(x)$ is continuous almost surely.

Proof: [10], Problem 2.9.

□

The smoothness of a random field is studied via its differentiability.

Definition 25. Assume that D is open. A random field $\{Z(x) : x \in D\}$ is *L_2-differentiable* or *mean square differentiable* at $x \in D$ in direction $h \in \mathbb{R}^d$ if there exists an L_2-limit of $(Z(x+th) - Z(x))/t$ as $t \to 0$. If we denote this limit by $Z'(x,h)$, then the following relation must be satisfied,

$$\lim_{t\to 0} \mathbb{E}\left(\frac{Z(x+th) - Z(x)}{t} - Z'(x,h)\right)^2 = 0.$$

Let $\{e_1, \ldots, \mathrm{e}_d\}$ be the canonical basis of $\mathbb{R}^d$, then $Z'(x, e_j)$, $j = 1, \ldots, d$, are the *partial derivatives* of a random field Z at point x.

2.2 Variogram and autocovariance function

Definition and properties

Random fields are used as the models for geostatistical data (e.g. temperature, air quality, or soil mineral content). In geostatistics, a popular tool for describing the spatial correlation is the so-called variogram.

Definition 26. For an intrinsic stationary random field $\{Z(x) : x \in D\}$, we define its *variogram* as

$$2\gamma(h) = \mathrm{var}(Z(x+h) - Z(x)), \quad h \in D - D,$$

where $D - D = \{h \in \mathbb{R}^d : h = x - y, x \in D, y \in D\}$. The function $\gamma(h)$ itself is called a *semivariogram*. If the function $\gamma(h)$ depends only on $\|h\|$, we speak about an *isotropic* semivariogram or variogram. In this case, we use the letter γ also for the function $\gamma(\|h\|)$ which is defined on non-negative real numbers.

It is clear from the definition that $\gamma(h) = \gamma(-h)$, $\gamma(o) = 0$, and $\gamma(h) \geq 0$. A weakly motion-invariant random field has an isotropic (semi)variogram. The semivariogram of a weakly stationary random field is bounded and it is linked to the autocovariance function by the formula $\gamma(h) = C(o) - C(h)$, which follows from (9). In general, the function γ needs not to be bounded.

Lemma 9. *Let $\{Z(x) : x \in D\}$ be an intrinsic stationary random field with the semivariogram γ. Then $\gamma(h) \to 0$ for $\|h\| \to 0+$ if and only if the random field is L_2-continuous.*

Proof: The result follows directly from the definition as for an intrinsic stationary random field we have $2\gamma(h) = \mathrm{var}(Z(x+h) - Z(x)) = \mathbb{E}(Z(x+h) - Z(x))^2$.

If $\gamma(h)$ is not continuous at the origin, then we speak about the so-called *nugget effect*.

Definition 27. If there exists a limit $\lim_{\|h\|\to 0+} 2\gamma(h) = 2\tau^2 > 0$, it is called the *nugget*. If there exists a finite limit $\lim_{\|h\|\to\infty} 2\gamma(h) = 2(\tau^2 + \sigma^2)$, it is called the *sill*. In this case, we define the *range* as

$$r = \inf\{s \geq 0 : 2\gamma(h) = 2(\tau^2 + \sigma^2) \text{ for all } h \in \mathbb{R}^d : \|h\| \geq s\}.$$

The value $2\sigma^2$ is referred to as the *partial sill*.

An example of an isotropic variogram with nugget, sill, and finite range is depicted in Figure 4.

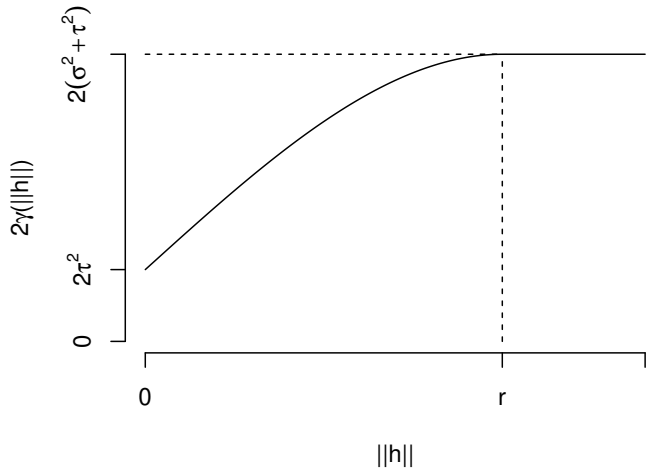

Figure 4. An illustration of the definition of the nugget, sill, and range.

The nugget appears when the repeated measurements at the same location give different values. It happens, for example, when we are not observing the realization of a random field directly, but the observations are affected by some error. Let $\{S(x) : x \in D\}$ be an intrinsically stationary random field (so-called signal) with variogram $2\gamma_S(h)$, which is continuous at the origin. Let $\{\varepsilon(x) : x \in D\}$ be an intrinsically stationary random field (so-called noise) that is independent of $\{S(x) : x \in D\}$. We observe a realization of the random field $\{Z(x) : x \in D\}$, where $Z(x) = S(x) + \varepsilon(x)$. If the $\varepsilon(x)$ are uncorrelated random variables with zero mean and variance τ^2 (so-called *white noise*), then the variogram of $\{Z(x) : x \in D\}$ is

$$2\gamma(h) = 2\gamma_S(h) + 2\tau^2 \mathbf{1}_{[h \neq o]}$$

and thus the nugget is equal to $2\tau^2$.

The autocovariance function of any weakly stationary random field has the following properties: $C(h) = C(-h)$; $C(o) = \operatorname{var} Z(x) \geq 0$; $|C(h)| \leq C(o)$, and so C is always a bounded function. The following theorem gives the connection between the L_2-continuity of the random field and the continuity of its autocovariance function.

Theorem 10. *Let $\{Z(x) : x \in D\}$ be a random field with finite second moments (i.e. $\mathbb{E}Z(x)^2 < \infty$ for each $x \in D$) such that its mean $\mu(x) = \mathbb{E}Z(x)$ is continuous on D. The random field is mean square continuous if and only if its autocovariance function $C(x, y)$ is continuous at the points satisfying $x = y$.*

Proof: We can use the relation

$$\begin{aligned}\mathbb{E}(Z(x+h) - Z(x))^2 &= \operatorname{var}(Z(x+h) - Z(x)) + (\mu(x+h) - \mu(x))^2 \\ &= C(x+h, x+h) - 2C(x+h, x) + C(x, x) + (\mu(x+h) - \mu(x))^2.\end{aligned}$$

If $C(x, y)$ is continuous at the points $x = y$, then the right-hand side tends to zero as $\|h\| \to 0+$. Therefore, the random field is L_2-continuous.

On the contrary, L_2-continuity of the random field implies

$$\lim_{\|h\| \to 0+} [C(x+h, x+h) - 2C(x+h, x) + C(x, x)] = 0, \tag{10}$$

which after taking squares gives

$$4 \lim_{\|h\| \to 0+} C(x+h, x)^2 = \left(\lim_{\|h\| \to 0+} C(x+h, x+h) + C(x, x) \right)^2.$$

From the Cauchy–Schwarz inequality we have

$$C(x+h, x)^2 \leq C(x+h, x+h) C(x, x),$$

and so

$$4 \lim_{\|h\| \to 0+} C(x+h, x+h) C(x, x) \geq \left(\lim_{\|h\| \to 0+} C(x+h, x+h) + C(x, x) \right)^2.$$

Denote $a = \lim_{\|h\| \to 0+} C(x+h, x+h)$ and $b = C(x, x)$. We have obtained the inequality $4ab \geq (a+b)^2$ which is possible only if $a = b$. It means that $\lim_{\|h\| \to 0+} C(x+h, x+h) = C(x, x)$. Due to (10), it also follows that $\lim_{\|h\| \to 0+} C(x+h, x) = C(x, x)$ for any $x \in D$. Finally,

$$|C(x+h, x+h') - C(x, x)| \leq |C(x+h, x+h') - C(x+h, x)| + |C(x+h, x) - C(x, x)|$$

and both terms on the right-hand side converge to zero as $\|h\|, \|h'\| \to 0+$.

□

Corollary 11. *A weak stationary random field is L_2-continuous if and only if its autocovariance function is continuous at the origin.*

Proof: The assertion follows immediately from Theorem 10. However, it is easy to prove it directly. It suffices to realize the relation (9) between the variogram and autocovariance function and use Lemma 9.

□

Similarly, it can be shown that L_2-differentiability of a weak stationary (or intrinsic stationary) random field is related to the differentiability of its autocovariance function (or variogram) at the origin.

Now we state an auxiliary result that will be useful for the calculation of the second-order characteristics of the random field.

Lemma 12. *For a covariance stationary random field $\{Z(x) : x \in D\}$, we have*

$$\operatorname{cov}\left(\sum_{j=1}^{n} \alpha_j Z(x_j), \sum_{j=1}^{n} \beta_j Z(x_j)\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_j \beta_k C(x_k - x_j)$$

for any $n \in \mathbb{N}$, $x_1, \dots, x_n \in D$ and $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \in \mathbb{R}$.

For an intrinsic stationary random field $\{Z(x) : x \in D\}$, we have

$$\operatorname{cov}\left(\sum_{j=1}^{n} \alpha_j Z(x_j), \sum_{j=1}^{n} \beta_j Z(x_j)\right) = -\sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_j \beta_k \gamma(x_k - x_j)$$

for any $n \in \mathbb{N}$, $x_1, \dots, x_n \in D$ and $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \in \mathbb{R}$ satisfying $\sum_{j=1}^{n} \alpha_j = \sum_{j=1}^{n} \beta_j = 0$.

Proof: The first relation is a well-known formula for the covariance of linear combinations of random variables. For the proof of the second relation we use

$$\sum_{j=1}^{n} \alpha_j Z(x_j) = \sum_{j=1}^{n} \alpha_j (Z(x_j) - Z(x_1)), \quad \sum_{j=1}^{n} \beta_j Z(x_j) = \sum_{j=1}^{n} \beta_j (Z(x_j) - Z(x_1))$$

and the identity

$$2\gamma(x_j - x_k) = 2\gamma(x_j - x_1) + 2\gamma(x_k - x_1) - 2\operatorname{cov}(Z(x_j) - Z(x_1), Z(x_k) - Z(x_1)),$$

which follows from $Z(x_j) - Z(x_k) = (Z(x_j) - Z(x_1)) - (Z(x_k) - Z(x_1))$ by computing the variance on both sides. Altogether we get

$$\begin{aligned}
\operatorname{cov}\left(\sum_{j=1}^{n} \alpha_j Z(x_j), \sum_{j=1}^{n} \beta_j Z(x_j)\right) &= \sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_j \beta_k \operatorname{cov}(Z(x_j) - Z(x_1), Z(x_k) - Z(x_1)) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_j \beta_k (\gamma(x_j - x_1) + \gamma(x_k - x_1) - \gamma(x_j - x_k)) \\
&= -\sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_j \beta_k \gamma(x_j - x_k).
\end{aligned}$$

□

Definition 28. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a symmetric function, i.e. $f(x) = f(-x)$ for every $x \in \mathbb{R}^d$. We say that f is *positive semidefinite* if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_j \alpha_k f(x_j - x_k) \geq 0$$

for every $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathbb{R}^d$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. The function f is called *conditional negative definite* if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} \beta_j \beta_k f(x_j - x_k) \leq 0$$

for every $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathbb{R}^d$ and $\beta_1, \dots, \beta_n \in \mathbb{R}$ satisfying $\sum_{j=1}^{n} \beta_j = 0$.

Corollary 13. *The autocovariance function of a covariance stationary random field is positive semidefinite. The variogram of an intrinsic stationary random field is conditional negative definite.*

Proof: From Lemma 12 it follows that

$$0 \le \operatorname{var} \sum_{j=1}^{n} \alpha_j Z(x_j) = \operatorname{cov}\left(\sum_{j=1}^{n} \alpha_j Z(x_j), \sum_{j=1}^{n} \alpha_j Z(x_j) \right) = \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \alpha_k C(x_j - x_k)$$

for arbitrary $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and

$$0 \le \operatorname{var} \sum_{j=1}^{n} \beta_j Z(x_j) = \operatorname{cov}\left(\sum_{j=1}^{n} \beta_j Z(x_j), \sum_{j=1}^{n} \beta_j Z(x_j) \right) = -\sum_{j=1}^{n} \sum_{k=1}^{n} \beta_j \beta_k \gamma(x_j - x_k)$$

for arbitrary $\beta_1, \ldots, \beta_n \in \mathbb{R}$ satisfying $\sum_{j=1}^{n} \beta_j = 0$.

□

Moreover, the class of all autocovariance functions of covariance stationary random fields coincides with the class of positive semidefinite functions. Similarly, the class of all variograms of intrinsic stationary random fields coincides with the class of conditional negative definite functions that attain value zero at the origin.

Theorem 14. *For each positive semidefinite function $C : \mathbb{R}^d \to \mathbb{R}$ there exists a covariance stationary random field such that C is its autocovariance function. For each conditional negative definite function $\gamma : \mathbb{R}^d \to \mathbb{R}$ satisfying $\gamma(o) = 0$ there exists an intrinsic stationary random field such that 2γ is its variogram.*

Proof: For arbitrary $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in \mathbb{R}^d$, the matrix $\Sigma = (C(x_i - x_j))_{i,j=1,\ldots,n}$ is positive semidefinite (it follows from the positive semidefiniteness of C) and we can consider n-dimensional centred normal distribution with the covariance matrix Σ. We get a projective system of finite-dimensional distributions. By Daniell–Kolmogorov theorem (Theorem 7), there exists a Gaussian random field $\{Z(x) : x \in \mathbb{R}^d\}$ that satisfies $\operatorname{cov}(Z(x), Z(y)) = C(x - y)$. In the same way, we can show that for any positive semidefinite function C on $\mathbb{R}^d \times \mathbb{R}^d$ there exists a centred Gaussian random field with the autocovariance function $C(x, y)$.

Let $\beta_0, \beta_1, \ldots, \beta_n \in \mathbb{R}$ satisfy $\sum_{j=0}^{n} \beta_j = 0$. Then for each $x_0, x_1, \ldots, x_n \in \mathbb{R}^d$, we have

$$0 \le -\sum_{j=0}^{n} \sum_{k=0}^{n} \beta_j \beta_k \gamma(x_j - x_k) = \sum_{j=0}^{n} \sum_{k=0}^{n} \beta_j \beta_k (\gamma(x_j) + \gamma(x_k) - \gamma(x_j - x_k)). \tag{11}$$

We consider the function $C(x, y) = \gamma(x) + \gamma(y) - \gamma(x - y)$. It satisfies $C(o, x) = C(x, o) = 0$ for any $x \in \mathbb{R}^d$. If we put $x_0 = o$, then for arbitrary $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$,

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \alpha_k C(x_j, x_k) = \sum_{j=0}^{n} \sum_{k=0}^{n} \alpha_j \alpha_k C(x_j, x_k),$$

which is non-negative due to (11) because we can take $\alpha_0 = -\sum_{j=1}^{n} \alpha_j$. We found out that $C(x, y)$ is positive semidefinite function on $\mathbb{R}^d \times \mathbb{R}^d$. Hence, by the first part of the proof, there exists a centred Gaussian random field $\{Z(x) : x \in \mathbb{R}^d\}$ such that $\operatorname{cov}(Z(x), Z(y)) = C(x, y)$. It remains to verify that it is intrinsic stationary and its variogram is 2γ:

$$\begin{aligned} \operatorname{var}(Z(x) - Z(y)) &= C(x, x) + C(y, y) - 2C(x, y) \\ &= 2\gamma(x) + 2\gamma(y) - 2\gamma(x) - 2\gamma(y) + 2\gamma(x - y) = 2\gamma(x - y). \end{aligned}$$

□

Spectral decomposition

Similarly to the stochastic processes, we can consider a spectral decomposition of the autocovariance function of a random field. It is based on the Bochner theorem which states that every continuous positive semidefinite function has a unique representation in the form of a Fourier transform of some finite measure.

Theorem 15. *A complex-valued function C on $\mathbb{R}^d$ is an autocovariance function of a weak stationary L_2-continuous complex-valued random field $\{Z(x) : x \in \mathbb{R}^d\}$ if and only if it can be expressed as*

$$C(h) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\omega^{\mathrm{T}}h} \,\mathrm{d}S(\omega), \quad h \in \mathbb{R}^d, \tag{12}$$

where S has the following properties:

1. $\lim_{\min_{i=1,\dots,d}\omega_i \to \infty} S(\omega_1, \dots, \omega_d) = C(o)$,
2. $\lim_{\omega_i \to -\infty} S(\omega_1, \dots, \omega_d) = 0$ *for every* $i = 1, \dots, d$,
3. S *is right continuous in every coordinate,*
4. S *is non-decreasing in* ω, *i.e. for each* $\omega, \vartheta \in \mathbb{R}^d$ *satisfying* $\omega_i < \vartheta_i$, $i = 1, \dots, d$, *we have*

$$\mu_S((\omega, \vartheta]) = \sum_{\delta_1=0}^{1} \cdots \sum_{\delta_d=0}^{1} (-1)^{d-\sum_{i=1}^d \delta_i} S(\omega_1 + \delta_1(\vartheta_1 - \omega_1), \dots, \omega_d + \delta_d(\vartheta_d - \omega_d)) \geq 0,$$

where $(\omega, \vartheta] = (\omega_1, \vartheta_1] \times \cdots \times (\omega_d, \vartheta_d]$.

Proof: [16], Theorem 1.9.6. □

The function S is called a *spectral distribution function*. It generates a finite Lebesgue-Stieltjes measure μ_S. The integral in (12) is understood as the integral w.r.t. this measure (instead of $\mathrm{d}S(\omega)$ we can write $\mu_S(\mathrm{d}\omega)$). Thus, it is the Lebesgue-Stieltjes integral. If there exists a density $s(\omega)$ of the function $S(\omega)$, then it is called a *spectral density*. The inverse formula for the spectral density has the form (if $\int_{\mathbb{R}^d} |C(h)| \,\mathrm{d}h < \infty$)

$$s(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{e}^{-\mathrm{i}\omega^{\mathrm{T}}h} C(h) \,\mathrm{d}h, \quad \omega \in \mathbb{R}^d.$$

For a real-valued random field, we get

$$C(h) = \int_{\mathbb{R}^d} \cos(\omega^{\mathrm{T}}h) \,\mathrm{d}S(\omega) = \int_{\mathbb{R}^d} \cos(\omega^{\mathrm{T}}h) s(\omega) \,\mathrm{d}\omega, \quad h \in \mathbb{R}^d$$

and

$$s(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \cos(\omega^{\mathrm{T}}h) C(h) \,\mathrm{d}h, \quad \omega \in \mathbb{R}^d,$$

if the spectral density exists.

Now let us add an assumption of the weak isotropy of the random field. Then the autocovariance function is isotropic, i.e. it satisfies $C(h) = C(\|h\|)$. Therefore, for $r \geq 0$ and $u \in \mathbb{S}^{d-1}$ an element of the unit sphere in $\mathbb{R}^d$, we have $C(r) = C(ru) = \int_{\mathbb{S}^{d-1}} C(ru) \,U(\mathrm{d}u)$, where $U(\cdot) = \mathcal{H}^{d-1}(\cdot)/\mathcal{H}^{d-1}(\mathbb{S}^{d-1})$ is the probability spherical measure. The symbol $\mathcal{H}^{d-1}$ stands for the $(d-1)$-dimensional Hausdorff measure. Plugging in the integral from the spectral decomposition yields

$$C(r) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} \cos(r\omega^{\mathrm{T}}u) \,\mathrm{d}S(\omega) \,U(\mathrm{d}u) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} \cos(r\omega^{\mathrm{T}}u) \,U(\mathrm{d}u) \,\mathrm{d}S(\omega).$$

The inner integral can be expressed as

$$\int_{\mathbb{S}^{d-1}} \cos(r\omega^{\mathrm{T}}u) \,U(\mathrm{d}u) = \Omega_d(r\|\omega\|),$$

where

$$\Omega_d(t) = \left(\frac{2}{t}\right)^{\nu} \Gamma(d/2) J_\nu(t),$$

$\nu = d/2 - 1$ and J_ν is the Bessel function of the first kind of order ν (see Subsection 6.2). The result depends on ω only through its norm $\|\omega\|$. In this way, we can replace the d-dimensional Fourier transform with a one-dimensional integral (so-called *Hankel* or *Bessel transform*):

$$C(\|h\|) = \int_0^\infty \Omega_d(\|h\|v) \,\mathrm{d}H(v). \tag{13}$$

The function $H(u)$ is non-decreasing on $\mathbb{R}^+$ with finite limit for $u \to \infty$ and it is connected with the spectral distribution function by the relation

$$H(u) = \mu_S(b(o,u)) = \int_{b(o,u)} \mathrm{d}S(\omega),$$

where $b(x,r)$ denotes the closed ball with centre x and radius r. The function Ω_d is called a *basis function* of the autocovariance function. Particular examples of the basis function are $\Omega_1(t) = \cos t$, $\Omega_2(t) = J_0(t)$, $\Omega_3(t) = \frac{\sin t}{t}$, and $\lim_{d\to\infty} \Omega_d(t) = \mathrm{e}^{-t^2}$. These functions are shown in Figure 5.

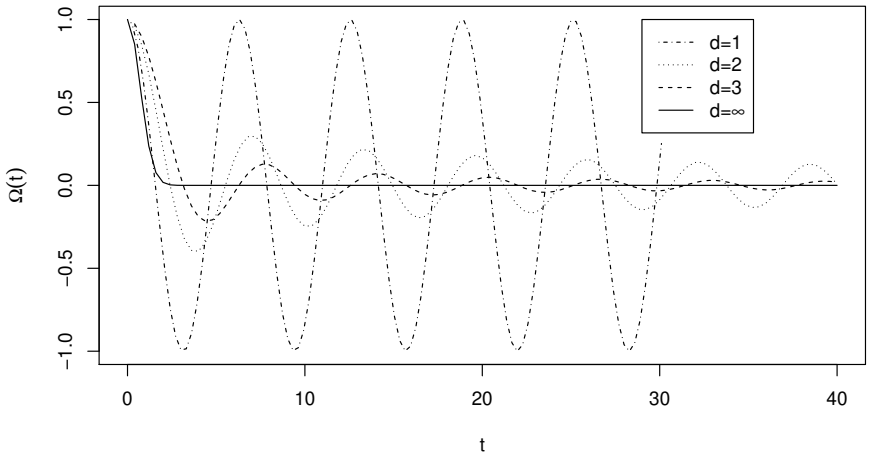

Figure 5. The basis functions $\Omega_d(t)$ for $d = 1$, $d = 2$, $d = 3$, and $d \to \infty$.

The variogram of an intrinsic stationary random field has the spectral representation as well:

$$2\gamma(h) = \int_{\mathbb{R}^d} \frac{1 - \cos(\omega^{\mathrm{T}} h)}{\|\omega\|^2} \,\mathrm{d}F(\omega), \quad h \in \mathbb{R}^d,$$

where F induces a measure on $\mathbb{R}^d$, which has no atom at the origin and satisfies $\int_{\mathbb{R}^d} (1+\|\omega\|^2)^{-1} \,\mathrm{d}F(\omega) < \infty$. In the isotropic case, we have

$$2\gamma(\|h\|) = \int_0^\infty \frac{1 - \Omega_d(v\|h\|)}{v^2} \,\mathrm{d}F(v),$$

where F does not have an atom at 0 and $\int_0^\infty (1+v^2)^{-1} \,\mathrm{d}F(v) < \infty$.

Lemma 16. *Every variogram of an intrinsic stationary random field satisfies*

$$\frac{2\gamma(h)}{\|h\|^2} \to 0 \quad \textit{for} \quad \|h\| \to \infty.$$

Proof: The spectral decomposition yields

$$\frac{2\gamma(h)}{\|h\|^2} = \int_{\mathbb{R}^d} \frac{1 - \cos(\omega^{\mathrm{T}} h)}{\|\omega\|^2 \|h\|^2} \,\mathrm{d}F(\omega).$$

Since for $\|h\| \geq 2$, we have

$$\left| \frac{1 - \cos(\omega^{\mathrm{T}} h)}{\|\omega\|^2 \|h\|^2} \right| \leq \frac{1}{2} \min\left(1, \frac{1}{\|\omega\|^2}\right) \leq \frac{1}{1 + \|\omega\|^2},$$

the assertion follows from the Lebesgue dominated convergence theorem.

□

Parametric models

We mention several basic parametric models for the isotropic semivariogram of an intrinsic stationary random field. In all these models, one of the parameters is the nugget $2\tau^2 \geq 0$.

0. *nugget*:

$$\gamma(h) = \tau^2 \mathbf{1}_{[h \neq o]},$$

this model is known as the *nugget model*. An example of a random field with this semivariogram is a white noise formed by uncorrelated random variables with constant mean and finite variance τ^2. One realization of the white noise is shown in Figure 3 left.

1. *power*:

$$\gamma(h) = \tau^2 \mathbf{1}_{[h \neq o]} + \sigma^2 \|h\|^\nu,$$

where $\sigma > 0$ and $0 < \nu < 2$. This model does not have sill and the corresponding random field is not weak stationary. The power model for $\nu = 1$ is called *linear* because it is a linear function of $\|h\|$. The semivariogram for different choices of ν is shown in Figure 6. For $d = 1$ and $\tau = 0$, the corresponding Gaussian process is known as the fractional Brownian motion with Hurst parameter $H = \nu/2, 0 < H < 1$. For $d > 1$ and $\tau = 0$, we have the so-called *fractional isotropic Brownian motion* in $\mathbb{R}^d$ or also *Lévy's fractional Brownian random field* [2], see Exercise class.

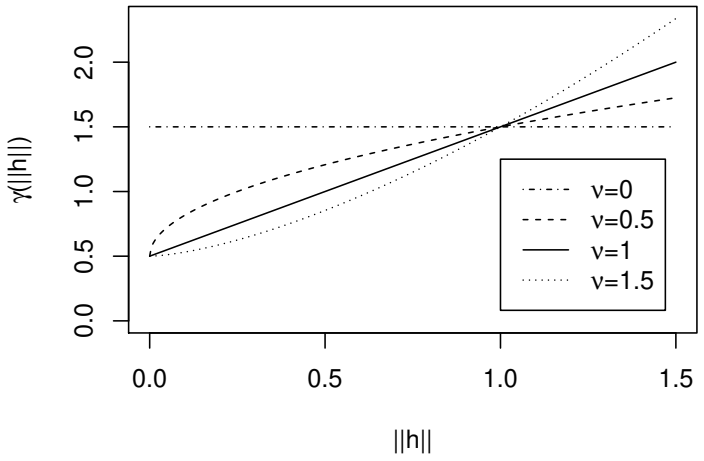

Figure 6. Power model of the isotropic semivariogram for $\tau^2 = 0.5$, $\sigma^2 = 1$, and different choices of the parameter ν.

2. *spherical*:

$$\gamma(h) = \tau^2 \mathbf{1}_{[h \neq o]} + \sigma^2 \left(1 - \frac{|b(o, \varrho) \cap b(h, \varrho)|}{|b(o, \varrho)|}\right),$$

where $\sigma > 0$ and $\varrho > 0$. The sill of this model is $2(\tau^2 + \sigma^2)$ and the range is $r = 2\varrho$, see Figure 7. The most commonly used is the spherical model for $d = 3$:

$$\gamma(h) = \begin{cases} 0 & \text{for } h = o, \\ \tau^2 + \sigma^2 \left(\frac{3\|h\|}{2r} - \frac{\|h\|^3}{2r^3}\right) & \text{for } 0 < \|h\| \leq r, \\ \tau^2 + \sigma^2 & \text{for } \|h\| \geq r. \end{cases}$$

This model is also valid in lower dimensions $d = 1$ and $d = 2$. However, the model that is valid in a lower dimension does not have to be valid in a higher dimension (see Exercise class). The explicit expression of the spherical semivariogram for $d = 2$ (also known as the *circular semivariogram*) contains goniometric functions and is left to the Exercise class. For $d = 1$ we obtain the so-called *triangular semivariogram*

$$\gamma(h) = \tau^2 + \sigma^2 \frac{h}{2\varrho}, \quad 0 < h \leq 2\varrho.$$

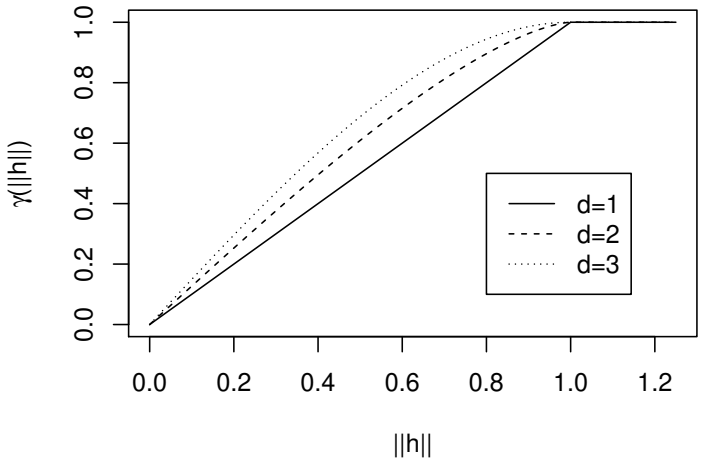

Figure 7. Spherical semivariogram in dimensions $d = 1$, $d = 2$, and $d = 3$ for the following choice of parameters: $\tau^2 = 0$, $\sigma^2 = 1$, and $r = 1$.

3. *generalized exponential*:

$$\gamma(h) = \tau^2 \mathbf{1}_{[h \neq o]} + \sigma^2 \left(1 - \exp\{-(\|h\|/a)^\nu\}\right),$$

where $\sigma > 0$, $a > 0$, and $0 < \nu \leq 2$. The sill of this model is $2(\tau^2 + \sigma^2)$ and the range is infinite. The corresponding graph of the function γ is depicted in Figure 8. Two special cases are the most often used: for $\nu = 1$ we speak about the *exponential* model and for $\nu = 2$ we get the *Gaussian* model.

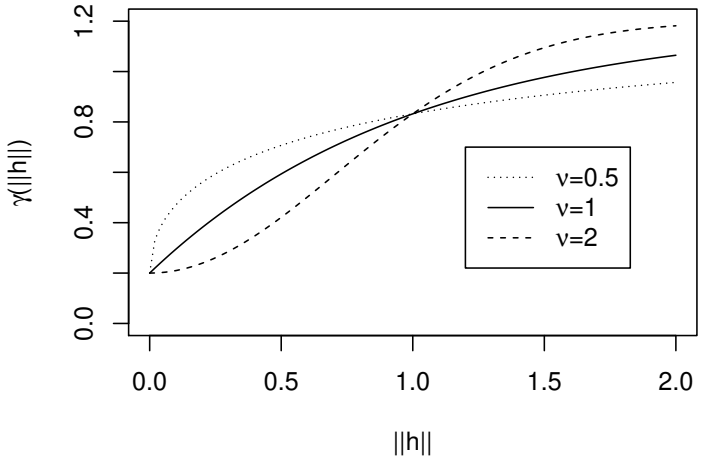

Figure 8. Generalized exponential model of the isotropic semivariogram for the following choice of parameters: $\tau^2 = 0.2$, $\sigma^2 = 1$, $a = 1$, and $\nu \in \{0.5, 1, 2\}$.

Both the spherical and the generalized exponential models have sill and thus they lead to a weak stationary random field with the autocovariance function $C(h) = \sigma^2 + \tau^2 - \gamma(h)$. A relatively wide and flexible class of parametric models for isotropic autocovariance functions is obtained by the *Whittle–Matérn model*:

$$C(h) = \tau^2 \mathbf{1}_{[h=o]} + \sigma^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\alpha\|h\|\right)^\nu K_\nu(\alpha\|h\|), \tag{14}$$

where $\tau^2 \geq 0$, $\nu > 0$, $\alpha > 0$, $\sigma^2 \geq 0$ are real parameters, and K_ν denotes the modified Bessel function of the second kind of order ν (see Subsection 6.2). The parameter τ^2 is half of the nugget, $\tau^2 + \sigma^2$ gives the variance of the random field, α is the scale parameter and the parameter ν is related to the differentiability of the autocovariance function and thus to the differentiability of the random field. For Gaussian random fields, it holds that their realizations are k-times differentiable if and only if $\nu > k$. The graph of this autocovariance function for several choices of ν can be found in Figure 9. For $\nu = 1/2$ we have an exponential autocovariance function $C(h) = \sigma^2 \exp\{-\alpha\|h\|\}$. The case $\nu = 1$ was proposed in Whittle's original paper [21]. The spectral density of the autocovariance function (14) has the following form:

$$s(\omega) = \sigma^2 \frac{\Gamma\left(\nu + \frac{d}{2}\right)}{\Gamma(\nu)\pi^{d/2}} \cdot \frac{\alpha^{2\nu}}{(\alpha^2 + \omega^2)^{\nu + \frac{d}{2}}}.$$

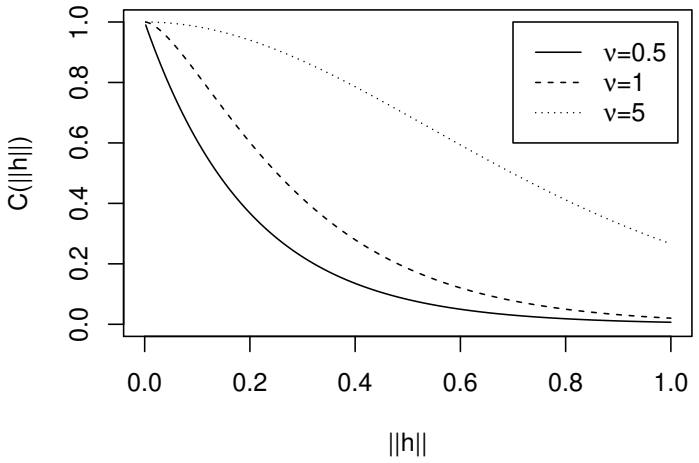

Figure 9. The Whittle–Matérn autocovariance function for $\alpha = 5$, $\sigma = 1$, and three values of ν.

3. Random measures and point processes

Let (E, ϱ) be a separable metric space. Recall that separability means that E contains a countable dense subset. We assume that (E, ϱ) has the Heine–Borel property, i.e. every closed and bounded subset is compact. It can be shown that E is complete and locally compact, i.e. every point $x \in E$ has a compact neighbourhood U_x. Then E is also σ-compact which means that it can be represented as a countable union of compact sets. Indeed, we can write $E = \cup_{x \in S} U_x$, where S is a countable dense subset that exists due to separability.

Example: The most important case for applications is the d-dimensional Euclidean space $E = \mathbb{R}^d$ with the Euclidean metric $\varrho(x, y) = \|x - y\|_d$. For geometric models, it is often considered that $E = \mathcal{K}'(\mathbb{R}^d)$ is the space of non-empty compact subsets of $\mathbb{R}^d$ with the Hausdorff metric

$$\varrho(K, L) = \max\left\{ \sup_{x \in K} \inf_{y \in L} \|x - y\|_d, \sup_{y \in L} \inf_{x \in K} \|x - y\|_d \right\}.$$

We will use the following notation for the systems of subsets of the space E:

$\mathcal{B}(E)$	... Borel sets,
$\mathcal{B}_0(E)$	... bounded Borel sets,
$\mathcal{F}(E)$	... closed sets,
$\mathcal{G}(E)$	... open sets,
$\mathcal{K}(E)$	... compact sets,
$\mathcal{K}'(E)$	... non-empty compact sets.

If it does not lead to confusion, we will omit the symbol E and write shortly $\mathcal{B}$, $\mathcal{B}_0$, $\mathcal{F}$, $\mathcal{G}$, $\mathcal{K}$, $\mathcal{K}'$.

3.1 Locally finite measures

Definition 29. A Borel measure μ on E is said to be
- *boundedly finite* if it is finite on $\mathcal{B}_0$, i.e. $\mu(B) < \infty$ for all $B \in \mathcal{B}_0$,
- *locally finite* if it is finite on $\mathcal{K}$, i.e. $\mu(K) < \infty$ for all $K \in \mathcal{K}$.

By $\mathcal{M} = \mathcal{M}(E)$ we denote the space of all locally finite measures on $(E, \mathcal{B})$ and by $\mathcal{M}_{\mathrm{f}} = \mathcal{M}_{\mathrm{f}}(E) = \{\mu \in \mathcal{M} : \mu(E) < \infty\}$ we denote the space of all finite measures on $(E, \mathcal{B})$. Moreover, we denote

$$\mathcal{N} = \mathcal{N}(E) = \{\mu \in \mathcal{M} : \mu(B) \in \mathbb{N} \cup \{0, \infty\}\ \forall B \in \mathcal{B}\}$$

the space of all locally finite counting measures on $(E, \mathcal{B})$ and $\mathcal{N}_{\mathrm{f}} = \mathcal{M}_{\mathrm{f}} \cap \mathcal{N}$ the space of all finite counting measures.

Under our requirements on the space E, a Borel measure is boundedly finite if and only if it is locally finite. In what follows, we use only the term locally finite measure.

Lemma 17. *Every locally finite measure on E is σ-finite.*

Proof: Since E is σ-compact, we can write $E = \cup_n K_n$ where the sets $K_n \in \mathcal{K}$ have finite measure. □

Definition 30. A set $A \subseteq E$ is called *locally finite* if $A \cap K$ is a finite set for every $K \in \mathcal{K}(E)$. Obviously, every locally finite set is closed and has at most countably many points. Denote by $\mathcal{F}_{\mathrm{lf}} = \{A \in \mathcal{F} : A \text{ locally finite}\}$ the family of all locally finite sets.

Lemma 18. *Every locally finite counting measure $\nu \in \mathcal{N}$ has the form $\nu = \sum_{x \in A} m_x \delta_x$, where A is a locally finite set, $m_x \in \mathbb{N}$, and δ_x is the Dirac measure at x.*

Proof: Put $A = \{x \in E : \nu(\{x\}) \geq 1\}$. Obviously, $\nu \in \mathcal{N}$ implies $A \in \mathcal{F}_{lf}$. It can be shown that any $B \in \mathcal{B}$ with $\nu(B) > 0$ contains $x \in A$ (see the proof of Lemma 29). Therefore, for $K \in \mathcal{K}$, $\nu(K)$ is given as a finite sum of $m_x = \nu(\{x\})$ over $x \in A \cap K$. □

Definition 31. For $B \in \mathcal{B}(E)$, a *one-dimensional projection* is the mapping $\pi_B : \mathcal{M} \to [0, \infty]$ defined as $\pi_B(\mu) = \mu(B)$. On the space $\mathcal{M}(E)$ we introduce a σ-algebra $\mathfrak{M}$ as the smallest σ-algebra for which all one-dimensional projections are measurable. We can write it shortly as

$$\mathfrak{M} = \sigma\{\pi_B \text{ measurable}, B \in \mathcal{B}\}.$$

The space $\mathcal{N} \subseteq \mathcal{M}$ is endowed with the σ-algebra $\mathfrak{N}$ defined as the trace of the σ-algebra $\mathfrak{M}$ on $\mathcal{N}$:

$$\mathfrak{N} = \{\mathcal{U} \cap \mathcal{N} : \mathcal{U} \in \mathfrak{M}\}.$$

Remark 11. If we denote $\mathcal{M}_{B,I} = \{\mu \in \mathcal{M} : \mu(B) \in I\}$ for $B \in \mathcal{B}$ and $I \in \mathcal{B}([0, \infty])$, then $\pi_B^{-1}(I) = \mathcal{M}_{B,I}$. Hence, $\mathfrak{M} = \sigma\{\mathcal{M}_{B,I} : B \in \mathcal{B}, I \in \mathcal{B}([0, \infty])\}$. Moreover, it is enough to consider $I = [0, r)$, $r \in [0, \infty]$. Then we denote $\mathcal{M}_{B,r} = \mathcal{M}_{B,[0,r)}$ and we have $\mathfrak{M} = \sigma\{\mathcal{M}_{B,r} : B \in \mathcal{B}, r \in [0, \infty]\}$. Since $\mathcal{M}_{\mathrm{f}} = \mathcal{M}_{E,\infty}$, it is clear that $\mathcal{M}_{\mathrm{f}} \in \mathfrak{M}$. Consequently, $\mathcal{N}_{\mathrm{f}} \in \mathfrak{N}$.

Lemma 19. *Let $\mathcal{S} \subseteq \mathcal{K}$ be a π-system that generates $\mathcal{B}$ (i.e. $\sigma\mathcal{S} = \mathcal{B}$) and let there exist the sets $A_n \in \mathcal{S}$ such that $A_n \nearrow E$. Then*

$$\mathfrak{M} = \sigma\{\pi_A \text{ measurable}, A \in \mathcal{S}\}.$$

Proof: Denote

$$\widetilde{\mathfrak{M}} = \sigma\{\pi_A \text{ measurable}, A \in \mathcal{S}\} = \sigma\{\pi_A^{-1}([0, r)) : A \in \mathcal{S}, r \in [0, \infty]\} = \sigma\{\mathcal{M}_{A,r} : A \in \mathcal{S}, r \in [0, \infty]\}.$$

Then obviously $\widetilde{\mathfrak{M}} \subseteq \mathfrak{M}$. If we define

$$\mathcal{D}_n = \{B \in \mathcal{B} : \pi_{B \cap A_n} \text{ is } \widetilde{\mathfrak{M}}\text{-measurable}\},$$

then we can easily verify that it is a Dynkin system which contains $\mathcal{S}$. Therefore, we obtain $\mathcal{D}_n = \sigma\mathcal{S} = \mathcal{B}$ by Dynkin's theorem (Theorem 94). For each $B \in \mathcal{B}$, we have $\mu(B \cap A_n) \nearrow \mu(B)$. It means that the

mapping π_B is $\widetilde{\mathfrak{M}}$-measurable (limit of measurable mappings). Since $\mathfrak{M}$ is the smallest σ-algebra such that the π_B are measurable, we get $\mathfrak{M} \subseteq \widetilde{\mathfrak{M}}$.

□

Lemma 20. *It holds that $\mathcal{N} \in \mathfrak{M}$. Hence, $\mathfrak{N} \subseteq \mathfrak{M}$.*

Proof: Consider a countable π-system $\mathcal{S} \subseteq \mathcal{K}$ that generates $\mathcal{B}$ and contains A_n such that $A_n \nearrow E$. Then

$$\mathcal{N}_0 = \{\mu \in \mathcal{M} : \mu(A) \in \mathbb{N}_0 \ \forall A \in \mathcal{S}\} = \bigcap_{A \in \mathcal{S}} \mathcal{M}_{A,\mathbb{N}_0} \in \mathfrak{M},$$

because it is a countable intersection of measurable sets. Clearly, $\mathcal{N} \subseteq \mathcal{N}_0$. Define $\mathcal{D}_n = \{B \in \mathcal{B} : \mu(B \cap A_n) \in \mathbb{N}_0 \ \forall \mu \in \mathcal{N}_0\}$. It is a Dynkin system that contains $\mathcal{S}$. Hence, $\mathcal{D}_n = \mathcal{B}$ by Dynkin's theorem (Theorem 94). For each $\mu \in \mathcal{N}_0$ and $B \in \mathcal{B}$, we have $\mu(B) = \lim_{n\to\infty} \mu(B \cap A_n) \in \mathbb{N} \cup \{0, \infty\}$, and so $\mu \in \mathcal{N}$.

□

For $B_1, \ldots, B_n \in \mathcal{B}$, we denote the σ-algebra generated by the mappings $\pi_{B_1}, \ldots, \pi_{B_n}$:

$$\mathfrak{M}_{B_1,\ldots,B_n} = \sigma\{\pi_{B_i} \text{ measurable}, i \in \{1, \ldots, n\}\}.$$

Lemma 21. *Let $\mathcal{S} \subseteq \mathcal{B}_0$ be a ring such that $\sigma\mathcal{S} = \mathcal{B}$. Then*

$$\mathfrak{M}_0 = \bigcup\{\mathfrak{M}_{A_1,\ldots,A_n} : n \in \mathbb{N}, A_1, \ldots, A_n \in \mathcal{S} \text{ pairwise disjoint}\}$$

is an algebra and $\sigma\mathfrak{M}_0 = \mathfrak{M}$.

Proof: Since $\mathfrak{M}_{A_1,\ldots,A_n}$ is a σ-algebra, the system $\mathfrak{M}_0$ is closed under complements and contains $\emptyset$ and $\mathcal{M}$. We have to show that it is closed under finite unions (or intersections). First consider the sets $\mathcal{U} = \mathcal{M}_{A,I_1} \in \mathfrak{M}_A$ and $\mathcal{V} = \mathcal{M}_{B,I_2} \in \mathfrak{M}_B$ with $A, B \in \mathcal{S}$ and $I_1, I_2 \in \mathcal{B}([0,\infty])$. Then we have

$$\mathcal{U} \cap \mathcal{V} = \{\mu : \mu(A \setminus B) + \mu(A \cap B) \in I_1, \mu(B \setminus A) + \mu(A \cap B) \in I_2\} \in \mathfrak{M}_{A\setminus B, A\cap B, B\setminus A} \subseteq \mathfrak{M}_0$$

and similarly $\mathcal{U} \cup \mathcal{V} \in \mathfrak{M}_{A\setminus B, A\cap B, B\setminus A} \subseteq \mathfrak{M}_0$. Analogously, we can prove $\mathcal{U} \cap \mathcal{V}, \mathcal{U} \cup \mathcal{V} \in \mathfrak{M}_0$ for $\mathcal{U} \in \mathfrak{M}_{A_1,\ldots,A_m}$ and $\mathcal{V} \in \mathfrak{M}_{B_1,\ldots,B_n}$.

Denote $\mathcal{D} = \{B \in \mathcal{B} : \pi_B \text{ is } \sigma\mathfrak{M}_0\text{-measurable}\}$. Obviously, it is a monotone system that contains $\mathcal{S}$. Thus, by Theorem 95, we get $\sigma\mathcal{S} = \mathcal{B} \subseteq \mathcal{D}$, which gives $\mathcal{D} = \mathcal{B}$. According to the definition of $\mathfrak{M}$, we have $\mathfrak{M} \subseteq \sigma\mathfrak{M}_0$. Hence, $\mathfrak{M} = \sigma\mathfrak{M}_0$.

□

Definition 32. We say that a sequence of finite measures $\mu_n \in \mathcal{M}_\mathrm{f}$ converges *weakly* to $\mu \in \mathcal{M}_\mathrm{f}$ (we write $\mu_n \xrightarrow[n\to\infty]{w} \mu$), if

$$\int_E f(x)\,\mu_n(\mathrm{d}x) \underset{n\to\infty}{\longrightarrow} \int_E f(x)\,\mu(\mathrm{d}x)$$

for any continuous and bounded function f on E. The sequence $\mu_n \in \mathcal{M}$ converges *vaguely* to $\mu \in \mathcal{M}$ (we write $\mu_n \xrightarrow[n\to\infty]{v} \mu$), if

$$\int_E f(x)\,\mu_n(\mathrm{d}x) \underset{n\to\infty}{\longrightarrow} \int_E f(x)\,\mu(\mathrm{d}x)$$

for any continuous and bounded function $f : E \to \mathbb{R}$ with compact support.

Example: For $E = \mathbb{R}$, consider $\mu_n = \delta_n$ as the Dirac measure at $n \in \mathbb{N}$. Then μ_n converges vaguely to zero measure but it does not converge weakly.

Definition 33. The *Prokhorov distance* between two finite measures $\mu, \nu \in \mathcal{M}_\mathrm{f}$ is defined as

$$\varrho_P(\mu,\nu) = \inf\{\varepsilon > 0 : \mu(F) \le \nu(F^\varepsilon) + \varepsilon, \nu(F) \le \mu(F^\varepsilon) + \varepsilon \text{ for every } F \in \mathcal{F}\},$$

where $F^\varepsilon = \{x \in E : \exists y \in F, \varrho(x,y) < \varepsilon\}$ is the open ε-neighbourhood of the closed set F. For $\mu, \nu \in \mathcal{M}$ we put

$$\widehat{\varrho}_P(\mu,\nu) = \int_0^\infty \mathrm{e}^{-r} \frac{\varrho_P(\mu^{(r)}, \nu^{(r)})}{1 + \varrho_P(\mu^{(r)}, \nu^{(r)})}\,\mathrm{d}r,$$

where $\mu^{(r)}$ is the restriction of μ to the ball $b(x_0, r)$ for an arbitrarily chosen fixed point $x_0 \in E$.

Remark 12. It is not difficult to see that ϱ_P defines a metric on $\mathcal{M}_{\mathrm{f}}$ and $\widehat{\varrho}_P$ defines a metric on $\mathcal{M}$ (see Exercise class).

Proposition 22.

a) *The space $(\mathcal{M}_{\mathrm{f}}, \varrho_P)$ is a complete separable metric space. Its Borel σ-algebra coincides with the trace $\mathfrak{M}_{\mathrm{f}}$ of $\mathfrak{M}$ to $\mathcal{M}_{\mathrm{f}}$. The metric ϱ_P generates the weak convergence of measures:*

$$\varrho_P(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0 \Longleftrightarrow \mu_n \xrightarrow[n\to\infty]{w} \mu.$$

b) *The space $(\mathcal{M}, \widehat{\varrho}_P)$ is a complete separable metric space. Its Borel σ-algebra coincides with $\mathfrak{M}$ and the convergence in $(\mathcal{M}, \widehat{\varrho}_P)$ coincides with the vague convergence of measures:*

$$\widehat{\varrho}_P(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0 \Longleftrightarrow \mu_n \xrightarrow[n\to\infty]{v} \mu.$$

Proof: [14], Theorem 2.2 or [5], Sections A2.5 and A2.6.

3.2 Random measures

Definition 34. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A *random measure* Ψ is a measurable mapping $\Psi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{M}, \mathfrak{M})$. A *point process* Φ is a measurable mapping $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{N}, \mathfrak{N})$. The *distribution* of a random measure Ψ is the probability measure Q on $(\mathcal{M}, \mathfrak{M})$ given by $Q(\mathcal{U}) = \mathbb{P}(\{\omega \in \Omega : \Psi(\omega) \in \mathcal{U}\}), \mathcal{U} \in \mathfrak{M}$.

Remark 13. A point process is a special case of a random measure. In the case $E = \mathbb{R}^d$, the term "process" has nothing to do with a dynamic evolution in time, so a more appropriate term would probably be a "point field".

Remark 14. A random measure Ψ is the mapping from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathcal{M}, \mathfrak{M})$. It means that for $\omega \in \Omega$, $\Psi(\omega)$ is a locally finite measure. The value of this measure for the set $B \in \mathcal{B}$ would be denoted by the symbol $\Psi(\omega)(B)$. Often we omit the argument ω and write only $\Psi(B)$. Then $\Psi(B)$: $\Omega \to [0, \infty]$ defines a random variable.

Lemma 23. *The mapping $\Psi : \Omega \to \mathcal{M}$ is a random measure if and only if $\Psi(B)$ is a random variable for all $B \in \mathcal{B}$.*

Proof: Exercise class.

Remark 15. The statement of Lemma 23 remains true if we consider only Borel sets B that form a system $\mathcal{S}$ from Lemma 19.

Definition 35. For a random measure Ψ, we define its *intensity measure* by the relation $\Lambda(B) = \mathbb{E}\Psi(B)$, $B \in \mathcal{B}$.

Remark 16. An intensity measure is a Borel measure, σ-additivity follows from Levi's monotone convergence theorem ([11], Theorem 8.5). An intensity measure does not have to be locally finite. There could exist $K \in \mathcal{K}$ such that $\Psi(K) < \infty$ but $\Lambda(K) = \mathbb{E}\Psi(K) = \infty$.

Definition 36. The *finite-dimensional distributions* of a random measure Ψ are the distributions of the random vector $(\Psi(B_1), \dots, \Psi(B_n))^{\mathrm{T}}$ for $B_1, \dots, B_n \in \mathcal{B}$. They are determined by the distribution functions

$$F_{B_1,\dots,B_n}(t_1, \dots, t_n) = \mathbb{P}(\Psi(B_1) \le t_1, \dots, \Psi(B_n) \le t_n), \quad t_1, \dots, t_n \in \mathbb{R}^+.$$

The distribution of a random measure is uniquely determined by its finite-dimensional distributions.

Theorem 24. *Let $\mathcal{S} \subseteq \mathcal{B}_0$ be a ring such that $\sigma\mathcal{S} = \mathcal{B}$. Let Ψ_1 and Ψ_2 be two random measures defined on the same probability space. If the finite-dimensional distributions of Ψ_1 and Ψ_2 coincide for any $n \in \mathbb{N}$ and $B_1, \dots, B_n \in \mathcal{S}$ pairwise disjoint, then Ψ_1 and Ψ_2 are equally distributed.*

Proof: By the assumption, the distributions of Ψ_1 and Ψ_2 coincide on the algebra $\mathfrak{M}_0$, which is the π-system and thus Theorem 96 states that Ψ_1 and Ψ_2 have the same distribution on $\sigma\mathfrak{M}_0$, which is equal to $\mathfrak{M}$ by Lemma 21. □

Proposition 25. *Let T be a complete separable metric space, $\mathcal{T}$ its Borel σ-algebra and $\mathcal{T}_0 \subseteq \mathcal{T}$ the system of bounded Borel sets. Let $\{\xi_B, B \in \mathcal{T}\}$ be a family of non-negative random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying*

1. *$B \in \mathcal{T}_0 \Rightarrow \xi_B < \infty$ a.s.,*
2. *$A, B \in \mathcal{T}_0$, $A \cap B = \emptyset \Rightarrow \xi_{A\cup B} \overset{a.s.}{=} \xi_A + \xi_B$,*
3. *$B_n \in \mathcal{T}_0$, $B_n \searrow \emptyset \Rightarrow \xi_{B_n} \xrightarrow[n\to\infty]{a.s.} 0$.*

Then there exists a random measure Ψ such that $\Psi(B) \overset{a.s.}{=} \xi_B$ for all $B \in \mathcal{T}$.

Proof: It is a slight generalization of the theorem stating the existence of a regular version of the conditional expectation ([19], Theorem VI.1.21). The complete proof can be found in [6], Theorem 9.1.XV. □

Remark 17. We assume that $\mathbb{P}(\{\omega : \xi_{A\cup B}(\omega) \neq \xi_A(\omega) + \xi_B(\omega)\}) = 0$ for each A and B disjoint. However, we need $\mathbb{P}(\{\omega : \xi_{A\cup B}(\omega) \neq \xi_A(\omega) + \xi_B(\omega)$ for each A and B disjoint$\}) = 0$.

Theorem 26. *(existence of a random measure with given finite-dimensional distributions)*
Let $\{F_{B_1,\dots,B_n} : n \in \mathbb{N}, B_1, \dots, B_n \in \mathcal{B}\}$ be a family of distribution functions satisfying

1. *(non-negativity) $F_B(t) = 0$ for any $t < 0$ and $B \in \mathcal{B}$,*
2. *(local finiteness) $\lim_{t\to\infty} F_B(t) = 1$ for $B \in \mathcal{B}_0$,*
3. *(symmetry)*
$$F_{B_1,\dots,B_n}(t_1, \dots, t_n) = F_{B_{i_1},\dots,B_{i_n}}(t_{i_1}, \dots, t_{i_n})$$
for any $n \in \mathbb{N}$, $B_1, \dots, B_n \in \mathcal{B}$, $t_1, \dots, t_n \in \mathbb{R}^+$, and any permutation $i_1, \dots, i_n$ of the numbers $1, \dots, n$,
4. *(projectivity)*
$$\lim_{t_{n+1}\to\infty} F_{B_1,\dots,B_n,B_{n+1}}(t_1, \dots, t_n, t_{n+1}) = F_{B_1,\dots,B_n}(t_1, \dots, t_n)$$
for any $n \in \mathbb{N}$, $B_1, \dots, B_{n+1} \in \mathcal{B}$, and $t_1, \dots, t_n \in \mathbb{R}^+$,
5. *(additivity) $F_{B_1,B_2,B_1\cup B_2}(t_1, t_2, t_3)$ is concentrated on the diagonal $t_1 + t_2 = t_3$ for any disjoint $B_1, B_2 \in \mathcal{B}$,*
6. *(continuity in $\emptyset$) $\lim_{n\to\infty} F_{B_n}(t) = 1$ for any $t > 0$ and any sequence $B_n \in \mathcal{B}_0$ such that $B_n \searrow \emptyset$.*

Then there exists a random measure Ψ such that $(\Psi(B_1), \dots, \Psi(B_n))^{\mathrm{T}}$ has the distribution function $F_{B_1,\dots,B_n}$ for any $n \in \mathbb{N}$ and $B_1, \dots, B_n \in \mathcal{B}$. Moreover, the distribution of Ψ is uniquely determined.

Proof: Conditions 3. and 4. mean that $\{F_{B_1,\dots,B_n} : n \in \mathbb{N}, B_1, \dots, B_n \in \mathcal{B}\}$ forms a projective system of distribution functions. Therefore, we can apply the Daniell–Kolmogorov existence theorem. It assures the existence of a family $\{\xi_B, B \in \mathcal{B}\}$ satisfying

$$\mathbb{P}(\xi_{B_1} \leq t_1, \dots, \xi_{B_n} \leq t_n) = F_{B_1,\dots,B_n}(t_1, \dots, t_n), \quad t_1, \dots, t_n \in \mathbb{R}^+$$

for any $n \in \mathbb{N}$ and $B_1, \dots, B_n \in \mathcal{B}$. Condition 1. implies that the random variables ξ_B are non-negative a.s. while condition 2. means that ξ_B is a.s. finite-valued for a bounded Borel set B. The conditions 5. and 6. guarantee that the assumptions of Proposition 25 are satisfied. The sixth condition means that $\xi_{B_n} \xrightarrow[n\to\infty]{P} 0$, and hence $\xi_{B_{n_k}} \xrightarrow[k\to\infty]{a.s.} 0$ for some subsequence n_k. This already implies $\xi_{B_n} \xrightarrow[n\to\infty]{a.s.} 0$ because the sequence is monotone a.s. Therefore, by Proposition 25, there exists a random measure Ψ such that $\Psi(B) \overset{a.s.}{=} \xi_B$ for every $B \in \mathcal{B}_0$. This implies that Ψ and $\{\xi_B, B \in \mathcal{B}_0\}$ have the same finite-dimensional distributions. The distribution of Ψ is unique by Theorem 24. This completes the proof.

□

3.3 Simple point processes

Definition 37. Let $\mathcal{N}^* = \{\nu \in \mathcal{N} : \nu(\{x\}) \leq 1$ for every $x \in E\}$. We say that a locally finite counting measure $\nu \in \mathcal{N}$ is *simple* if $\nu \in \mathcal{N}^*$.

To show the measurability of $\mathcal{N}^*$, we will need to introduce the sequence of nested countable decompositions of the metric space E.

Definition 38. We say that the sequence $\{\mathcal{S}_n, n \in \mathbb{N}\}$ is a *DC-system (dissecting-covering system)* for E if

1. $\mathcal{S}_n = \{A_1^n, A_2^n, \ldots\} \subseteq \mathcal{B}_0(E)$ is a disjoint countable decomposition of E for each $n \in \mathbb{N}$, i.e. $A_i^n \cap A_j^n = \emptyset$ for $i \neq j$ and $E = \cup_j A_j^n$,
2. for any $n \in \mathbb{N}$ and $A \in \mathcal{S}_n$, there exist $A_1, \ldots, A_k \in \mathcal{S}_{n+1}$ such that $A = A_1 \cup \cdots \cup A_k$,
3. $\lim_{n\to\infty} \sup_{A\in\mathcal{S}_n} \operatorname{diam} A = 0$, where $\operatorname{diam} A = \sup\{\varrho(x,y) : x, y \in A\}$.

Lemma 27. *There exists a DC-system on E. If Λ is an arbitrary locally finite and diffuse (i.e. $\Lambda(\{x\}) = 0$ for any $x \in E$) Borel measure on E, then there exists a DC-system on E that furthermore satisfies*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{S}_n} \Lambda(A) = 0. \tag{15}$$

Proof: First we construct a disjoint decomposition $\mathcal{S}_1$ of E into Borel sets with diameters smaller or equal to one. Let S be a countable dense subset of E. Then $E = \cup_{x\in S} b(x, 1/2) = \cup_{i=1}^{\infty} \tilde{A}_i$. The sets $\tilde{A}_i$ are not necessarily disjoint so we define $A_1^1 = \tilde{A}_1$, $A_2^1 = \tilde{A}_2 \setminus \tilde{A}_1$, $A_3^1 = \tilde{A}_3 \setminus (\tilde{A}_1 \cup \tilde{A}_2), \ldots$ Successively we construct decompositions $\mathcal{S}_2, \mathcal{S}_3, \ldots$ by induction in the following way. If we already have $\mathcal{S}_{n-1}$ and if $A \in \mathcal{S}_{n-1}$, then from the open cover of the compact set $\bar{A}$ by open balls,

$$\bar{A} \subseteq \cup_{x\in\bar{A}} \operatorname{int} b\left(x, \frac{1}{2n}\right),$$

we can select a finite subcover from which we obtain a disjoint decomposition of $A = A_1 \cup \ldots \cup A_k$ into Borel sets of diameter at most $1/n$. The decomposition $\mathcal{S}_n$ is created by the union of dissections of all sets from $\mathcal{S}_{n-1}$.

If a diffuse measure $\Lambda \in \mathcal{M}$ is given, then

$$\lim_{\varepsilon\to 0} \Lambda(b(x,\varepsilon)) = 0$$

for arbitrary $x \in E$. We can modify the above construction of the DC-system by considering balls $\operatorname{int} b(x, \varepsilon(x,n))$ with $\varepsilon(x,n) \leq 1/2n$ and $\Lambda(\operatorname{int} b(x, \varepsilon(x,n))) < 1/n$ in the construction of $\mathcal{S}_n$. In such a way, we ensure that the condition (15) is satisfied.

□

Lemma 28. *The set of all simple locally finite counting measures is measurable, i.e. $\mathcal{N}^* \in \mathfrak{N}$.*

Proof: Consider a DC-system $\{\mathcal{S}_n, n \in \mathbb{N}\}$ in E and put $\mathcal{S} = \cup_{n=1}^{\infty} \mathcal{S}_n$. Then

$$\begin{aligned}\mathcal{N}^* &= \{\nu \in \mathcal{N} : \forall A \in \mathcal{S}\ \exists n \in \mathbb{N}\ \forall B \in \mathcal{S}_n : B \subseteq A \Rightarrow \nu(B) \leq 1\} \\ &= \bigcap_{A\in\mathcal{S}} \bigcup_{n\in\mathbb{N}} \bigcap_{B\in\mathcal{S}_n : B\subseteq A} \{\nu \in \mathcal{N} : \nu(B) \leq 1\},\end{aligned}$$

and so $\mathcal{N}^* \in \mathfrak{N}$.

□

Definition 39. We say that a point $x \in E$ is an *atom* of $\nu \in \mathcal{N}$ if $\nu(\{x\}) > 0$. A pair $(x, m) \in E \times \mathbb{N}$ is called an *atomic pair* if x is an atom and $m = \nu(\{x\})$.

Since $\nu \in \mathcal{N}$ is locally finite, it has at most countably many atoms (Lemma 18). The following lemma states that the atoms may be enumerated in a measurable way.

Lemma 29. *There exist measurable mappings $\zeta_i : \mathcal{N} \to E$ such that*

$$\nu = \sum_{i=1}^{\nu(E)} \delta_{\zeta_i(\nu)}$$

for any $\nu \in \mathcal{N}$.

Proof: Consider a DC-system $\{\mathcal{S}_n\}$, where $\mathcal{S}_n = \{A_1^n, A_2^n, \ldots\} \subseteq \mathcal{B}_0(E)$. Let $B \in \mathcal{B}$ be a Borel set with $\nu(B) > 0$. We can inductively define a sequence $i_1, i_2, \ldots$ such that $A_{i_1}^1 \supseteq A_{i_2}^2 \supseteq \cdots$ and $\nu(A_{i_r}^r \cap B) \in \mathbb{N}$ for $r \in \mathbb{N}$. It follows that $\nu(\cap_{k\in\mathbb{N}} A_{i_k}^k \cap B) \in \mathbb{N}$. From the properties of the DC-system, $\cap_{k\in\mathbb{N}} A_{i_k}^k$ has diameter zero. Therefore, B contains an atom. It means that ν can be written as

$$\nu = \sum_{(x,m)} m\delta_x,$$

where the sum runs over the atomic pairs (x,m) of ν.

For $x \in E$, the relations $x \in A^k_{j_k(x)}$, $k \in \mathbb{N}$, define uniquely a sequence $(j_1(x), j_2(x), \ldots)$ of integer numbers. This allows us to define a linear order $\prec$ on the space E:

$$x \prec y \Longleftrightarrow (j_1(x), j_2(x), \ldots) \leq_{\text{lex}} (j_1(y), j_2(y), \ldots),$$

where $\leq_{\text{lex}}$ denotes the lexicographical order. For each $p \in \mathbb{N}$, we construct a measurable map $\zeta_p : \mathcal{N} \to E$. It will associate with every counting measure ν its p-th atom (counted w.r.t. $\prec$ with multiplicities). Let (x,m) be an atomic pair of ν, then all atoms y of ν with $y \prec x$ and $y \neq x$, lie in the bounded set $\cup_{i=1}^{j_1(x)} A^1_i$. Hence, their number is finite and the sum of their multiplicities gives a finite number, say n. We define $\zeta_{n+j}(\nu) = x$ for $j = 1, \ldots, m$. If this is done for all atomic pairs (x,m) of ν, then $\zeta_p(\nu)$ is defined for all $p \in \mathbb{N}$ if $\nu(E) = \infty$, and it is defined for $p = 1, \ldots, q$ if $\nu(E) = q < \infty$. In the latter case, we put $\zeta_p(\nu) = a$ for $p > q$, where $a \in E$ is an arbitrary given point. For $p \in \mathbb{N}$ and $B \in \mathcal{B}$, the set $\{\nu \in \mathcal{N} : \nu(E) < p, \zeta_p(\nu) \in B\}$ is either empty (if $a \notin B$) or equal to $\{\nu \in \mathcal{N} : \nu(E) < p\}$. Thus, it is measurable in both cases. Furthermore, we have

$$\begin{aligned}\{\nu \in \mathcal{N} : \nu(E) \geq p, \zeta_p(\nu) \in B\} = \bigcup_{j=1}^{\infty} \bigcup_{i_1,\ldots,i_j=1}^{\infty} \Big\{ & \nu \in \mathcal{N} : \nu(B \cap A^1_{i_1} \cap \cdots \cap A^j_{i_j}) = \nu(A^1_{i_1} \cap \cdots \cap A^j_{i_j}) \in \mathbb{N}, \\ & \nu\left(\cup_{(r_1,\ldots,r_j) <_{\text{lex}} (i_1,\ldots,i_j)} A^1_{r_1} \cap \cdots \cap A^j_{r_j}\right) \leq p-1, \\ & \nu\left(\cup_{(r_1,\ldots,r_j) \leq_{\text{lex}} (i_1,\ldots,i_j)} A^1_{r_1} \cap \cdots \cap A^j_{r_j}\right) \geq p \Big\} \in \mathfrak{N},\end{aligned}$$

which shows the measurability of ζ_p.

□

Now we are ready to define a simple point process.

Definition 40. A point process Φ is called *simple* if $\mathbb{P}(\Phi \in \mathcal{N}^*) = 1$.

Remark 18. A simple point process can be also understood as a measurable mapping $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{N}^*, \mathfrak{N}^*)$, where $\mathfrak{N}^* = \{\mathcal{U} \cap \mathcal{N}^* : \mathcal{U} \in \mathfrak{N}\}$ is the trace of the σ-algebra $\mathfrak{N}$ on $\mathcal{N}^*$.

If ν in Lemma 29 is simple, we have $\zeta_i(\nu) \neq \zeta_j(\nu)$ for each $i \neq j$. A simple point process Φ is then a random measure given as the sum of Dirac measures: $\Phi = \sum_{i=1}^{\Phi(E)} \delta_{X_i}$, where the X_i are random elements in E.

Lemma 30. *Define $\mathcal{N}^*_{B,r} = \{\nu \in \mathcal{N}^* : \nu(B) = r\}$ for $B \in \mathcal{B}$ and $r \in \mathbb{N} \cup \{0, \infty\}$. Then $\mathfrak{N}^* = \sigma\{\mathcal{N}^*_{K,0} : K \in \mathcal{K}\}$.*

Proof: In Lemma 19, we can take $\mathcal{S} = \mathcal{K}$ and obtain

$$\mathfrak{N}^* = \sigma\{\{\nu \in \mathcal{N}^* : \nu(K) < r\} : K \in \mathcal{K}, r \in [0, \infty]\} = \sigma\{\mathcal{N}^*_{K,r} : K \in \mathcal{K}, r \in \mathbb{N}_0\}.$$

Using the DC-system $\{\{A^n_1, A^n_2, \ldots\}, n \in \mathbb{N}\}$, we can write for $r \in \mathbb{N}$,

$$\mathcal{N}^*_{K,r} = \bigcup_{n_0 \in \mathbb{N}} \bigcap_{n \geq n_0} \bigcup_{i_1,\ldots,i_r} \left[(\mathcal{N}^*_{K \cap \bar{A}^n_{i_1}, 0})^c \cap \cdots \cap (\mathcal{N}^*_{K \cap \bar{A}^n_{i_r}, 0})^c \cap \mathcal{N}^*_{K \setminus (\text{int } A^n_{i_1} \cup \cdots \cup \text{int } A^n_{i_r}), 0} \right],$$

which is the element of $\sigma\{\mathcal{N}^*_{K,0} : K \in \mathcal{K}\}$. Since $\mathfrak{N}^*$ is the smallest σ-algebra containing $\mathcal{N}^*_{K,r}$, we see that $\mathfrak{N}^* \subseteq \sigma\{\mathcal{N}^*_{K,0} : K \in \mathcal{K}\}$. The reverse inclusion is obvious.

□

Corollary 31. *The mapping $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{N}^*, \mathfrak{N}^*)$ is a simple point process if and only if the events $\{\Phi(K) = 0\} \in \mathcal{A}$ for any $K \in \mathcal{K}$.*

Definition 41. Let Φ be a point process. *Void probabilities* are $\mathbb{P}(\Phi(K) = 0)$, $K \in \mathcal{K}$.

The distribution of a simple point process is uniquely determined by its void probabilities.

Theorem 32. *Let Φ_1 and Φ_2 be two simple point processes on E such that $\mathbb{P}(\Phi_1(K) = 0) = \mathbb{P}(\Phi_2(K) = 0)$ for any $K \in \mathcal{K}$. Then Φ_1 and Φ_2 have the same distribution.*

Proof: Let Q_1 and Q_2 be the distributions of Φ_1 and Φ_2, respectively. Then $Q_1(\mathcal{N}^*_{K,0}) = \mathbb{P}(\Phi_1(K) = 0) = \mathbb{P}(\Phi_2(K) = 0) = Q_2(\mathcal{N}^*_{K,0})$ for any $K \in \mathcal{K}$. It means that the distributions coincide on the family $\{\mathcal{N}^*_{K,0} : K \in \mathcal{K}\}$ which is a π-system that generates $\mathfrak{N}^*$ (Lemma 30). Now it suffices to apply Theorem 96.

□

Remark 19. For $E = \mathbb{R}$, it is impossible to replace the collection of compact sets by a smaller class of test sets such as the closed intervals. One can construct two simple point processes with different distributions satisfying $\mathbb{P}(\Phi_1(I) = 0) = \mathbb{P}(\Phi_2(I) = 0)$ for any interval $I \subseteq \mathbb{R}$ (see Exercise class).

Simple locally finite counting measures are uniquely related to locally finite sets.

Definition 42. The *support* of a locally finite measure $\mu \in \mathcal{M}(E)$ is defined as the smallest closed subset A of E such that $\mu(E \setminus A) = 0$. It is denoted by $\operatorname{supp}\mu$ and it can be written as

$$\operatorname{supp}\mu = \bigcap\{F \in \mathcal{F} : \mu(E \setminus F) = 0\}.$$

Remark 20. The support of $\nu \in \mathcal{N}$ is a locally finite set,

$$\operatorname{supp}\nu = \{x \in E : \nu(\{x\}) \geq 1\} \in \mathcal{F}_{lf},$$

see the proof of Lemma 18.

For $A, A_1, \ldots, A_k \subseteq E$, we define the following subsets of the system $\mathcal{F}$ of closed sets:

$$\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \emptyset\}, \quad \mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \emptyset\},$$

and

$$\mathcal{F}^A_{A_1,\ldots,A_k} = \mathcal{F}^A \cap \mathcal{F}_{A_1} \cap \cdots \cap \mathcal{F}_{A_k} = \{F \in \mathcal{F} : F \cap A = \emptyset, F \cap A_1 \neq \emptyset, \ldots, F \cap A_k \neq \emptyset\}.$$

For $k = 0$ we put $\mathcal{F}^A_{A_1,\ldots,A_k} = \mathcal{F}^A$.

Definition 43. We define the σ-algebra $\mathfrak{F}$ on $\mathcal{F}$ as $\mathfrak{F} = \sigma\{\mathcal{F}^K : K \in \mathcal{K}\}$. A *random closed set* in E is a measurable mapping $\Xi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{F}, \mathfrak{F})$.

Remark 21. The space $\mathcal{F}$ of closed sets could be endowed with the topology $\mathcal{T}$ generated by the set system

$$\{\mathcal{F}^K_{G_1,\ldots,G_k} : K \in \mathcal{K}, G_1, \ldots, G_k \in \mathcal{G}, k \in \mathbb{N}_0\}.$$

This set system contains $\mathcal{F} = \mathcal{F}^\emptyset$ and is closed under finite intersections because

$$\mathcal{F}^K_{G_1,\ldots,G_k} \cap \mathcal{F}^{K'}_{G'_1,\ldots,G'_m} = \mathcal{F}^{K\cup K'}_{G_1,\ldots,G_k,G'_1,\ldots,G'_m}.$$

The topology $\mathcal{T}$ is known as the *Fell topology* or also the *hit-or-miss topology*.

Lemma 33. *The system of all locally finite sets is measurable:* $\mathcal{F}_{\mathrm{lf}} \in \mathfrak{F}$.

Proof: Let $\{\mathcal{S}_n, n \in \mathbb{N}\}$ be a DC-system in E and define $\mathcal{S} = \cup_{n=1}^\infty \mathcal{S}_n$. Then

$$\mathcal{F}_{\mathrm{lf}} = \{F \in \mathcal{F} : \forall A \in \mathcal{S} \exists k \in \mathbb{N} \forall n \in \mathbb{N} : \operatorname{card}\{B \in \mathcal{S}_n : B \subseteq A, F \cap \bar{B} \neq \emptyset\} \leq k\}.$$

□

Definition 44. A *random locally finite set* in E is a measurable mapping $\Xi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{F}_{\mathrm{lf}}, \mathfrak{F}_{\mathrm{lf}})$, where $\mathfrak{F}_{\mathrm{lf}} = \{\mathcal{U} \cap \mathcal{F}_{\mathrm{lf}} : \mathcal{U} \in \mathfrak{F}\}$ is the trace of the σ-algebra $\mathfrak{F}$ on $\mathcal{F}_{\mathrm{lf}}$.

Lemma 34. *The map $i : \nu \mapsto \operatorname{supp}\nu$ is a bijection between $(\mathcal{N}^*, \mathfrak{N}^*)$ and $(\mathcal{F}_{\mathrm{lf}}, \mathfrak{F}_{\mathrm{lf}})$. It is measurable and its inverse $i^{-1} : F \mapsto \sum_{x\in F} \delta_x$ is measurable as well.*

Proof: It is enough to realize that $i^{-1}(\mathcal{F}^K \cap \mathcal{F}_{\mathrm{lf}}) = \{\nu \in \mathcal{N}^* : \operatorname{supp}\nu \cap K = \emptyset\} = \{\nu \in \mathcal{N}^* : \nu(K) = 0\} \in \mathfrak{N}^*$.

Corollary 35. *If Φ is a simple point process on E, then $\operatorname{supp}\Phi$ is a random locally finite set in E. Conversely, if Ξ is a random locally finite set in E, then $\sum_{X\in\Xi} \delta_X$ is a simple point process on E.*

Theorem 36. *(Choquet–Matheron) The distribution of a random closed set Ξ is uniquely determined by the probabilities $\mathbb{P}(\Xi \cap K = \emptyset)$, $K \in \mathcal{K}$.*

Proof: The family $\mathfrak{F}_0 = \{\mathcal{F}^K : K \in \mathcal{K}\}$ forms a π-system that generates the σ-algebra $\mathfrak{F}$. Hence, if two distributions coincide on $\mathfrak{F}_0$, they also coincide on $\mathfrak{F}$ by Theorem 96. The distribution Q of Ξ is thus uniquely determined by $Q(\mathcal{F}^K) = \mathbb{P}(\Xi \cap K = \emptyset)$, $K \in \mathcal{K}$.

□

Theorem 32 is a consequence of Theorem 36 and Lemma 34.

3.4 Poisson point process

The most elementary example of a point process is δ_X, where X is a random element in E. Obviously, $\delta_X(B)$ is a Bernoulli random variable with parameter $\mathbb{P}(X \in B)$. By Lemma 23, δ_X is indeed a point process.

Definition 45. Let ν be a measure on E. Consider $B \in \mathcal{B}(E)$ such that $0 < \nu(B) < \infty$. For $n \in \mathbb{N}$, let $X_1, \ldots, X_n$ be independent and ν-uniformly distributed random elements in B, i.e.

$$\mathbb{P}(X_i \in A) = \frac{\nu(A)}{\nu(B)}, \quad A \subseteq B,\ A \in \mathcal{B}.$$

Then $\Phi^{(n)} = \sum_{i=1}^{n} \delta_{X_i}$ is the *binomial point process* of n points in B according to ν.

Remark 22. The measurability of $\Phi^{(n)}$ is clear (it is a sum of measurable). We may notice that $\Phi^{(n)}(A)$, $A \in \mathcal{B}$, has a binomial distribution with parameters n and $\nu(A \cap B)/\nu(B)$. The intensity measure of $\Phi^{(n)}$ is

$$\Lambda(A) = \mathbb{E}\Phi^{(n)}(A) = n\frac{\nu(A \cap B)}{\nu(B)}.$$

If ν is a diffuse measure, $\Phi^{(n)}$ is a simple point process.

Figure 10 provides examples of three realizations of a binomial point process of 10 uniformly distributed points.

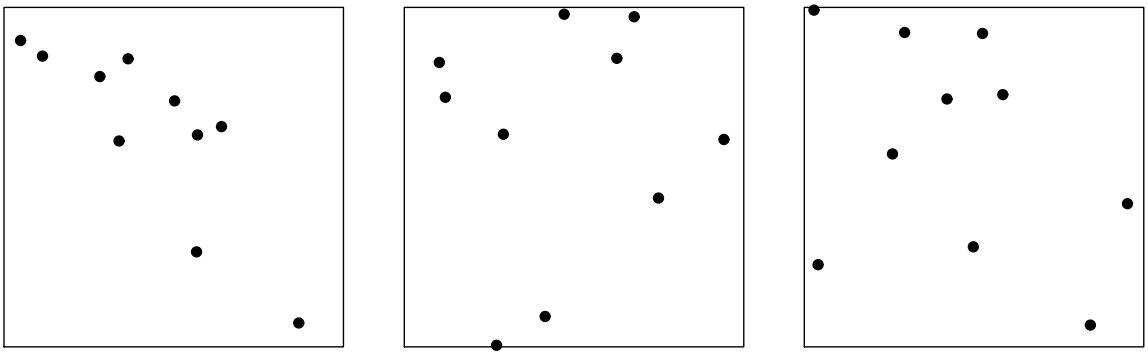

Figure 10. Three different realizations of a binomial point process of 10 uniformly distributed points in the unit square window.

More generally, we can consider a binomial point process with a random number of points in B.

Definition 46. Let ν be a measure and consider $B \in \mathcal{B}$ such that $0 < \nu(B) < \infty$. Let N be a non-negative integer-valued random variable. Furthermore, let $X_1, X_2, \ldots$ be independent ν-uniformly distributed random elements in B that are independent of N. Then $\Phi = \sum_{i=1}^{N} \delta_{X_i}$ is a *mixed binomial point process*. If $N = 0$, we put $\Phi(B) = 0$.

The Poisson point process is a canonical model in point process theory. It serves as a reference model when studying the summary characteristics. It is used as a cornerstone for the construction of more complex models.

On the real line ($E = \mathbb{R}$), the homogeneous Poisson point process is used to model events occurring completely at random in time. It is defined by imposing that the increments are independent and exponentially distributed. In the spatial case, the Poisson point process represents the locations (points) of randomly scattered objects. There are no interactions among the points (we speak about *complete spatial randomness*). Instead of the time increments, we have to work with the numbers of points in disjoint regions.

Definition 47. Let Λ be a locally finite measure on E. A point process Φ satisfying
(i) $\Phi(B)$ has a Poisson distribution with parameter $\Lambda(B)$ for each $B \in \mathcal{B}_0$,
(ii) $\Phi(B_1), \ldots, \Phi(B_n)$ are independent for each $n \in \mathbb{N}$ and $B_1, \ldots, B_n \in \mathcal{B}_0$ pairwise disjoint,

is called a *Poisson point process* with intensity measure Λ.

Remark 23. If $\Lambda(B) = 0$, then $\Phi(B) \overset{a.s.}{=} 0$. For $A \in \mathcal{B}$ unbounded, either $\Phi(A)$ has a Poisson distribution with parameter $\Lambda(A)$ if $\Lambda(A) < \infty$ or $\Phi(A) \overset{a.s.}{=} \infty$ if $\Lambda(A) = \infty$.

Three realizations of a Poisson point process are shown in Figure 11.

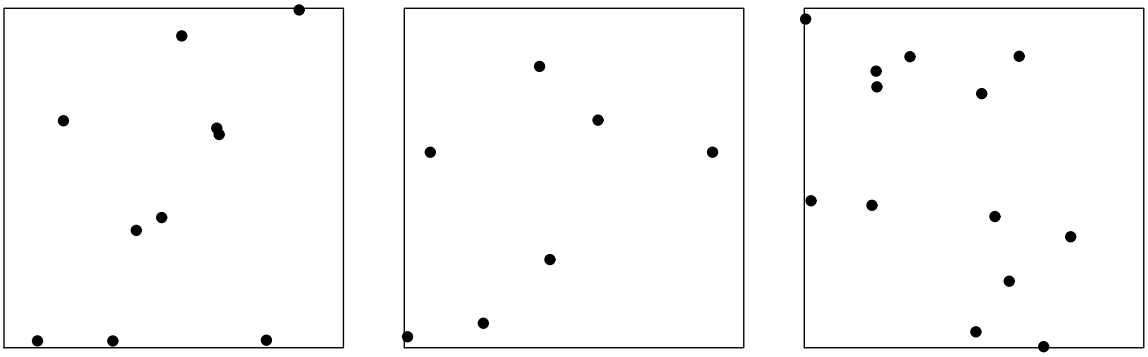

Figure 11. Three different realizations of a Poisson point process in the unit square window. The intensity measure is a multiple of the Lebesgue measure. The expected number of points in the window is 10. The actual observed numbers are 10, 7 and 13.

The uniqueness and existence is ensured by Theorem 26.

Corollary 37. *Let Λ be a locally finite measure on E. Then there exists a Poisson point process with intensity measure Λ and its distribution is uniquely determined.*

Proof: Let

$$P_B(j) = \mathrm{e}^{-\Lambda(B)} \frac{\Lambda(B)^j}{j!}, \quad j \in \mathbb{N}_0,\ B \in \mathcal{B},$$

be the probability mass function of a Poisson distribution with parameter $\Lambda(B)$. We allow $\Lambda(B) = \infty$ and in this case we put $P_B(j) = 0$, $j \in \mathbb{N}_0$. For $B_1, \ldots, B_n \in \mathcal{B}_0$ pairwise disjoint, define

$$F_{B_1,\ldots,B_n}(t_1, \ldots, t_n) = \prod_{i=1}^{n} \sum_{j_i \le t_i} P_{B_i}(j_i), \quad t_1, \ldots, t_n \in \mathbb{R}.$$

For $A_1, \ldots, A_n \in \mathcal{B}_0$ (not necessarily pairwise disjoint), we define $F_{A_1,\ldots,A_n}$ as follows. Let $B_1, \ldots, B_k \in \mathcal{B}_0$ be pairwise disjoint such that, for $i = 1, \ldots, n$, we can write $A_i = \cup_{j \in I_i} B_j$, where $I_i \subseteq \{1, \ldots, k\}$. The collection of sets $B_1, \ldots, B_k$ is called a disjoint decomposition of the sets $A_1, \ldots, A_n$. We put

$$F_{A_1,\ldots,A_n}(t_1, \ldots, t_n) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_k=0}^{\infty} \mathbf{1}_{[\sum_{i \in I_1} j_i \le t_1, \ldots, \sum_{i \in I_n} j_i \le t_n]} P_{B_1}(j_1) \cdots P_{B_k}(j_k).$$

This definition could be extended to $A_1, \ldots, A_n \in \mathcal{B}$ with $\Lambda(A_i) < \infty$. If $\Lambda(A_i) = \infty$ for some $i \in \{1, \ldots, n\}$, we define $F_{A_1,\ldots,A_n}(t_1, \ldots, t_n) = 0$ for $t_1, \ldots, t_n \in \mathbb{R}$. It is not difficult to show that the family $\{F_{A_1,\ldots,A_n} : n \in \mathbb{N}, A_1, \ldots, A_n \in \mathcal{B}\}$ satisfies the assumptions of Theorem 26.

□

Lemma 38. *The Poisson point process Φ with intensity measure $\Lambda \in \mathcal{M}$ is simple if and only if Λ is diffuse.*

Proof: If Λ is not diffuse, then there exists $x \in E$ such that $\Lambda(\{x\}) > 0$. It follows that

$$\mathbb{P}(\Phi(\{x\}) = k) = \mathrm{e}^{-\Lambda(\{x\})} \frac{\Lambda(\{x\})^k}{k!} > 0, \quad k \in \mathbb{N},$$

hence the point process Φ is not simple.

For the converse implication, suppose by contradiction that Λ is diffuse and Φ is not simple, that is $\mathbb{P}(\Phi \in \mathcal{N}^*) < 1$. Then there is a compact set $K \in \mathcal{K}$ with $\alpha = \mathbb{P}(\Phi|_K \notin \mathcal{N}^*) > 0$, where $\Phi|_K$ denotes the restriction of Φ to the set K. Denote $\varepsilon = \Lambda(K) > 0$. For arbitrary $k \in \mathbb{N}$ there exist pairwise disjoint Borel sets $B_1^{(k)}, \dots, B_k^{(k)} \in \mathcal{B}$ such that $K = \cup_{i=1}^k B_i^{(k)}$ and $\Lambda(B_i^{(k)}) = \varepsilon/k$, $i = 1, \dots, k$. There must be a number $i \in \{1, \dots, k\}$ with

$$\mathbb{P}(\Phi(B_i^{(k)}) > 1) \geq \frac{\alpha}{k}.$$

Therefore,

$$1 - \mathrm{e}^{-\Lambda(B_i^{(k)})}(1 + \Lambda(B_i^{(k)})) \geq \frac{\alpha}{k},$$

which after multiplying by k yields

$$k - \mathrm{e}^{-\varepsilon/k}(k + \varepsilon) \geq \alpha > 0.$$

The left-hand side converges to 0 for $k \to \infty$, which is a desired contradiction.

□

From Lemma 38 and Theorem 32, it follows that a point process Φ is Poisson with diffuse intensity measure Λ if and only if it is simple and $\mathbb{P}(\Phi(K) = 0) = \mathrm{e}^{-\Lambda(K)}$, $K \in \mathcal{K}$.

Theorem 39. *Let Φ be a Poisson point process with diffuse intensity measure $\Lambda \in \mathcal{M}$. Choose a Borel set $B \in \mathcal{B}$ such that $0 < \Lambda(B) < \infty$. Then conditionally on $\Phi(B) = n$, the restriction of Φ to B has the same distribution as a binomial point process of n points in B according to Λ.*

Proof: For arbitrary compact set $K \subseteq B$, we have

$$\begin{aligned}\mathbb{P}(\Phi|_B(K) = 0 \mid \Phi(B) = n) &= \frac{\mathbb{P}(\Phi(K) = 0, \Phi(B \setminus K) = n)}{\mathbb{P}(\Phi(B) = n)} = \frac{\mathbb{P}(\Phi(K) = 0)\mathbb{P}(\Phi(B \setminus K) = n)}{\mathbb{P}(\Phi(B) = n)} \\ &= \frac{\mathrm{e}^{-\Lambda(K)}\frac{\Lambda(B\setminus K)^n}{n!}\mathrm{e}^{-\Lambda(B\setminus K)}}{\frac{\Lambda(B)^n}{n!}\mathrm{e}^{-\Lambda(B)}} = \left(\frac{\Lambda(B \setminus K)}{\Lambda(B)}\right)^n,\end{aligned}$$

which are the void probabilities of a binomial point process. Since Λ is diffuse, we work with simple point processes and it suffices to apply Theorem 32.

□

Theorem 39 says that from a single realization, we are unable to distinguish the difference between a binomial point process and a Poisson point process. The number of points in the window is deterministic for a binomial point process while it is random (and has the Poisson distribution) for a Poisson point process, see Figure 10 and Figure 11.

A mixed binomial point process with N having the Poisson distribution is a Poisson point process (see Exercise class).

Denote by Π_Λ the distribution of a Poisson point process on E with intensity measure $\Lambda \in \mathcal{M}$.

Lemma 40. *For any $\mathcal{U} \in \mathfrak{N}$, $\Lambda \mapsto \Pi_\Lambda(\mathcal{U})$ is a measurable mapping from $\mathcal{M}$ to $[0, 1]$.*

Proof: The statement is obvious for $\mathcal{U} = \mathcal{N}_{B,r}$, $B \in \mathcal{B}_0$, $r \in \mathbb{N}_0$ because the probabilities $\Pi_\Lambda(\mathcal{U}) = \mathrm{e}^{-\Lambda(B)}\frac{\Lambda(B)^r}{r!}$ in the Poisson process are measurable functions of Λ. Similarly it holds for $\mathcal{U} = \mathcal{N}_{B_1,I_1} \cap \mathcal{N}_{B_2,I_2} \cap \dots \cap \mathcal{N}_{B_n,I_n}$, where $B_1, \dots, B_n \in \mathcal{B}_0$ are pairwise disjoint, $I_1, \dots, I_n \subseteq \mathbb{N}_0$ and $n \in \mathbb{N}$. In this case, $\Pi_\Lambda(\mathcal{U}) = \prod_{i=1}^n \mathrm{e}^{-\Lambda(B_i)} \sum_{r_i \in I_i} \frac{\Lambda(B_i)^{r_i}}{r_i!}$. Let $\mathcal{D} = \{\mathcal{U} \in \mathfrak{N} : \Lambda \mapsto \Pi_\Lambda(\mathcal{U}) \text{ is measurable}\}$ and let $\mathfrak{N}_0$ be the class of all finite unions of the sets in the form $\mathcal{N}_{B_1,I_1} \cap \dots \cap \mathcal{N}_{B_n,I_n}$, where $n \in \mathbb{N}$, $B_1, \dots, B_n \in \mathcal{B}_0$ are pairwise disjoint, $I_1, \dots, I_n \subseteq \mathbb{N}_0$. Then $\mathcal{D}$ is a monotone system containing an algebra $\mathfrak{N}_0$. It follows from Theorem 95 that $\mathcal{D}$ contains $\sigma\mathfrak{N}_0 = \mathfrak{N}$. Hence, $\mathcal{D} = \mathfrak{N}$.

□

Now we mention a generalization of a Poisson process where the intensity measure is random.

Definition 48. Let Ψ be a random diffuse measure on E with distribution Q_Ψ. The *Cox point process* Φ with *driving random measure* Ψ has distribution given as the mixture $Q(\cdot) = \int_{\mathcal{M}} \Pi_\Lambda(\cdot)\, Q_\Psi(\mathrm{d}\Lambda)$.

Remark 24. In other words, a Cox point process Φ is a Poisson point process with random intensity measure Ψ. The previous definition means that conditionally on $\Psi = \Lambda$, Φ is a Poisson point process

with intensity measure Λ. Therefore, Φ is sometimes also called a *doubly stochastic Poisson process*. We have assumed that Ψ is diffuse to ensure that Φ is simple.

Lemma 41. *The intensity measure of a Cox point process with driving measure Ψ is equal to the intensity measure of the random measure Ψ.*

Proof: For arbitrary $B \in \mathcal{B}$, we have

$$\begin{aligned}\mathbb{E}\Phi(B) &= \int_{\mathcal{N}} \mu(B)\, Q(\mathrm{d}\mu) = \int_{\mathcal{M}} \int_{\mathcal{N}} \mu(B)\, \Pi_\Lambda(\mathrm{d}\mu)\, Q_\Psi(\mathrm{d}\Lambda) \\ &= \int_{\mathcal{M}} \mathbb{E}\Phi_\Lambda(B)\, Q_\Psi(\mathrm{d}\Lambda) = \int_{\mathcal{M}} \Lambda(B)\, Q_\Psi(\mathrm{d}\Lambda) = \mathbb{E}\Psi(B).\end{aligned}$$

Using the conditioning, one can get an equivalent and more transparent argument:

$$\mathbb{E}\Phi(B) = \mathbb{E}[\mathbb{E}(\Phi(B) \mid \Psi)] = \mathbb{E}\Psi(B).$$

□

Remark 25. Similarly, we can show that the void probabilities of a Cox point process Φ with driving measure Ψ are $\mathbb{P}(\Phi(K) = 0) = \mathbb{E}\mathrm{e}^{-\Psi(K)}$, $K \in \mathcal{K}$.

A Cox point process is a natural generalization of a Poisson point process obtained by supposing that the intensity measure is not deterministic but random. The Poisson point process is a special case of the Cox point process ($\Psi = \Lambda$ is deterministic). The simplest non-trivial example of a Cox point process is provided by taking Ψ to be a random multiple of some deterministic measure Λ.

Definition 49. Consider a deterministic measure $\Lambda \in \mathcal{M}$. Let $\Pi_{t\Lambda}$ be the distribution of a Poisson point process with intensity measure $t\Lambda$. Let Y be a non-negative random variable with distribution P_Y. A point process with the distribution $Q = \int_0^\infty \Pi_{t\Lambda}\, P_Y(\mathrm{d}t)$ is called the *mixed Poisson point process*. It is an example of a Cox point process where the driving random measure Ψ is $Y\Lambda$.

According to Lemma 41, the intensity measure of the mixed Poisson point process is $(\mathbb{E}Y) \cdot \Lambda(\cdot)$.

3.5 Moment measures

We have already defined the intensity measure $\Lambda(\cdot) = \mathbb{E}\Psi(\cdot)$ of a random measure Ψ.

Definition 50. The *n-th order moment measure* of a random measure Ψ is defined as

$$M^{(n)}(A) = \mathbb{E}\Psi^n(A), \quad A \in \mathcal{B}(E^n),$$

where Ψ^n denotes the n-th power of Ψ in the usual sense of product of measures. In particular,

$$M^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}[\Psi(B_1) \cdots \Psi(B_n)], \quad B_1, \ldots, B_n \in \mathcal{B}(E).$$

Remark 26. The measure $M^{(n)}$ is in fact the intensity measure of a random measure Ψ^n on E^n. For a point process Φ on E, Φ^n is a point process on E^n with atoms being ordered n-tuples of the atoms of Φ.

Denote by $E^{[n]} = \{(x_1, \ldots, x_n) \in E^n : x_i \neq x_j \text{ for } i \neq j\}$ the set of n-tuples of pairwise distinct points from E. It is an open subset of E^n. The trace of $\mathcal{B}^n$ on $E^{[n]}$ will be denoted by $\mathcal{B}^{[n]}$. For $\mu \in \mathcal{M}$ let $\mu^{[n]} = \mu^n|_{E^{[n]}}$.

Definition 51. The *n-th order factorial moment measure* of a random measure Ψ is defined as

$$\alpha^{(n)}(A) = \mathbb{E}\Psi^{[n]}(A), \quad A \in \mathcal{B}(E^n).$$

Remark 27. First-order moment measures coincide with the intensity measure: $M^{(1)} = \alpha^{(1)} = \Lambda$.

Remark 28. Let Φ be a simple point process. Then

$$M^{(n)}(A) = \mathbb{E} \sum_{X_1, \ldots, X_n \in \operatorname{supp}\Phi} \mathbf{1}_{[(X_1, \ldots, X_n) \in A]}, \quad A \in \mathcal{B}(E^n),$$

and

$$\alpha^{(n)}(A) = \mathbb{E} \sum_{X_1,\dots,X_n \in \operatorname{supp}\Phi}^{\neq} \mathbf{1}_{[(X_1,\dots,X_n)\in A]}, \quad A \in \mathcal{B}(E^n),$$

where $\sum_{X_1,\dots,X_n \in \operatorname{supp}\Phi}^{\neq}$ means that the summation is only over the n-tuples of pairwise distinct points $X_1, \dots, X_n$.

The relation between the n-th order moment measure and the moments of the numbers of points is the following:

$$M^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}[\Phi(B_1) \cdots \Phi(B_n)], \quad B_1, \dots B_n \in \mathcal{B}.$$

In particular, $M^{(n)}(B \times \cdots \times B) = \mathbb{E}\Phi(B)^n$ for $B \in \mathcal{B}$. Similarly, the n-th order factorial moment measure yields the n-th factorial moment of the number of points in B:

$$\alpha^{(n)}(B \times \cdots \times B) = \mathbb{E}[\Phi(B)(\Phi(B)-1)\cdots(\Phi(B)-n+1)].$$

The verification of these relations is left to Exercise class.

The factorial moment measure has a simple form for the Poisson point process.

Theorem 42. *Let Φ be a Poisson point process with diffuse intensity measure Λ. For arbitrary $n \in \mathbb{N}$, its n-th order factorial moment measure is $\alpha^{(n)} = \Lambda^n$.*

Proof: For $B_1, \dots, B_n \in \mathcal{B}$ pairwise disjoint, we get

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}\Phi(B_1) \cdots \Phi(B_n).$$

From the property (ii) of the Poisson process (Definition 47), it follows that

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}\Phi(B_1) \cdots \mathbb{E}\Phi(B_n) = \Lambda(B_1) \cdots \Lambda(B_n).$$

The family $\{B_1 \times \cdots \times B_n : B_i \in \mathcal{B} \text{ and } B_i \cap B_j = \emptyset \text{ for } i \neq j\}$ is the π-system that generates the Borel σ-algebra $\mathcal{B}^{[n]}$ on $E^{[n]}$. Since the measures $\alpha^{(n)}$ and Λ^n coincide on this system, by Theorem 96 they also coincide on $\mathcal{B}^{[n]}$. This finishes the proof as both $\alpha^{(n)}$ and Λ^n have zero measure on $E^n \setminus E^{[n]}$. □

Remark 29. The n-th order factorial moment measure of a Cox point process with driving measure Ψ is $\alpha^{(n)} = \mathbb{E}\Psi^n$.

The following theorem will often be very useful.

Theorem 43. *(Campbell's theorem)*

(i) Let Ψ be a random measure and let h be an arbitrary non-negative measurable function on E^n. Then

$$\mathbb{E} \int_{E^n} h(x_1, \dots, x_n)\, \Psi^n(\mathrm{d}(x_1, \dots, x_n)) = \int_{E^n} h(x_1, \dots, x_n)\, M^{(n)}(\mathrm{d}(x_1, \dots, x_n))$$

and

$$\mathbb{E} \int_{E^{[n]}} h(x_1, \dots, x_n)\, \Psi^{[n]}(\mathrm{d}(x_1, \dots, x_n)) = \int_{E^{[n]}} h(x_1, \dots, x_n)\, \alpha^{(n)}(\mathrm{d}(x_1, \dots, x_n)).$$

(ii) For a simple point process Φ and an arbitrary non-negative measurable function h on E^n, we have

$$\mathbb{E} \sum_{X_1,\dots,X_n \in \operatorname{supp}\Phi} h(X_1, \dots, X_n) = \int_E \cdots \int_E h(x_1, \dots, x_n)\, M^{(n)}(\mathrm{d}x_1, \dots, \mathrm{d}x_n)$$

and

$$\mathbb{E} \sum_{X_1,\dots,X_n \in \operatorname{supp}\Phi}^{\neq} h(X_1, \dots, X_n) = \int_E \cdots \int_E h(x_1, \dots, x_n)\, \alpha^{(n)}(\mathrm{d}x_1, \dots, \mathrm{d}x_n).$$

Proof: For the indicators both relations follow directly from Definition 50. The rest of the proof proceeds by a standard argument of measure theory.

□

Definition 52. The *Laplace transform* of a random measure Ψ is the functional L_Ψ defined by the relation

$$L_\Psi(f) = \mathbb{E}\exp\left\{-\int_E f(x)\,\Psi(\mathrm{d}x)\right\},$$

where f is a non-negative measurable function on E.

Remark 30. In particular, if we choose $f = \sum_{i=1}^n t_i \mathbf{1}_{B_i}$ to be a simple function, where $B_1, \ldots, B_n \in \mathcal{B}$ are pairwise disjoint, then

$$L_\Psi(f) = \mathbb{E}\exp\left\{-\sum_{i=1}^n t_i \Psi(B_i)\right\}$$

is the Laplace transform of the random vector $(\Psi(B_1), \ldots, \Psi(B_n))$ in $(t_1, \ldots, t_n) \in (\mathbb{R}^+)^n$.

Corollary 44. *The Laplace transform uniquely determines the distribution of a random measure.*

Proof: Since the Laplace transform of a non-negative random vector uniquely determines its distribution, the assertion follows from Theorem 24.

□

Lemma 45.

(i) The Laplace transform of a Poisson point process Φ with intensity measure Λ is

$$L_\Phi(f) = \exp\left\{-\int_E \left(1 - \mathrm{e}^{-f(x)}\right)\,\Lambda(\mathrm{d}x)\right\}.$$

(ii) The Laplace transform of a mixed Poisson point process Φ with driving measure $Y \cdot \Lambda$ is

$$L_\Phi(f) = L_Y\left(\int_E (1 - \mathrm{e}^{-f(x)})\,\Lambda(\mathrm{d}x)\right),$$

where $L_Y(t) = \mathbb{E}\mathrm{e}^{-tY}$, $t \geq 0$, is the Laplace transform of the non-negative random variable Y.

Proof:

(i) For a simple function $f = \sum_{i=1}^n t_i \mathbf{1}_{B_i}$, we get

$$\begin{aligned} L_\Phi(f) &= \mathbb{E}\exp\left\{-\sum_{i=1}^n t_i \Phi(B_i)\right\} = \prod_{i=1}^n \mathbb{E}\mathrm{e}^{-t_i\Phi(B_i)} \\ &= \prod_{i=1}^n \mathrm{e}^{-\Lambda(B_i)(1-\mathrm{e}^{-t_i})} = \exp\left\{-\sum_{i=1}^n \Lambda(B_i)(1-\mathrm{e}^{-t_i})\right\}. \end{aligned}$$

We have used knowledge of the Laplace transform of the Poisson distributed random variables $\Phi(B_i)$. A standard argument of measure theory yields the result for arbitrary non-negative measurable function f.

(ii) Successively, we can write

$$\begin{aligned} L_\Phi(f) &= \int_{\mathcal{N}} \mathrm{e}^{-\int_E f(x)\,\mu(\mathrm{d}x)}\,Q(\mathrm{d}\mu) = \int_0^\infty \int_{\mathcal{N}} \mathrm{e}^{-\int_E f(x)\,\mu(\mathrm{d}x)}\,\Pi_{t\Lambda}(\mathrm{d}\mu)\,P_Y(\mathrm{d}t) \\ &= \int_0^\infty \exp\left\{-t\int_E (1-\mathrm{e}^{-f(x)})\,\Lambda(\mathrm{d}x)\right\}\,P_Y(\mathrm{d}t) = L_Y\left(\int_E (1-\mathrm{e}^{-f(x)})\,\Lambda(\mathrm{d}x)\right), \end{aligned}$$

where Q is the distribution of Φ, $\Pi_{t\Lambda}$ is the distribution of a Poisson point process with intensity measure $t\Lambda$ and P_Y is the distribution of Y.

□

3.6 Palm distribution

Definition 53. Let $(S,\mathcal{S})$ and $(T,\mathcal{T})$ be two measurable spaces. A map $K : S\times\mathcal{T}\to[0,\infty]$ is called *kernel* from $(S,\mathcal{S})$ to $(T,\mathcal{T})$ if it satisfies the following properties:
(i) the map $s\mapsto K(s,B)$ is a non-negative measurable function on S for every $B\in\mathcal{T}$,
(ii) $K(s,\cdot)$ is a measure on $(T,\mathcal{T})$ for every $s\in S$.

We say that K is a *Markov kernel* or *probability kernel* if $K(s,\cdot)$ is a probability measure for every $s\in S$.

Example: A random measure on E is a kernel from $(\Omega,\mathcal{A})$ to $(E,\mathcal{B})$.

Example: Let X be a random variable defined on $(\Omega,\mathcal{A},\mathbb{P})$ that takes values in $(S,\mathcal{S})$, and let $\mathcal{F}\subseteq\mathcal{A}$ be a σ-algebra. A regular conditional distribution of X given $\mathcal{F}$ is a Markov kernel K from $(\Omega,\mathcal{A})$ to $(S,\mathcal{S})$ such that $K(\omega,A)=\mathbb{P}(X\in A\mid\mathcal{F})(\omega)$ for almost all $\omega\in\Omega$ and for each $A\in\mathcal{S}$. It is known that the regular conditional distribution exists if S is a complete separable metric space with Borel σ-algebra $\mathcal{S}$.

Example: Let X and Y be random variables defined on $(\Omega,\mathcal{A},\mathbb{P})$ that take values in $(T,\mathcal{T})$ and $(S,\mathcal{S})$, respectively. Then the conditional distribution $K(y,A)=P_{X|Y}(A\mid y)$ is a Markov kernel from $(S,\mathcal{S})$ to $(T,\mathcal{T})$ that satisfies

$$\mathbb{P}(X\in A,Y\in B)=\int_B P_{X|Y}(A\mid y)\,P_Y(\mathrm{d}y),\quad A\in\mathcal{T},B\in\mathcal{S},$$

where P_Y is the distribution of Y.

Theorem 46. *(desintegration theorem) Let $(S,\mathcal{S})$ be a measurable space and let $(T,\mathcal{T})$ be a complete separable metric space with Borel σ-algebra. Consider a measure μ on $(S\times T,\mathcal{S}\otimes\mathcal{T})$ and assume that its projection $\nu(\cdot)=\mu(\cdot\times T)$ is a σ-finite measure on $(S,\mathcal{S})$. Then there is a Markov kernel K from $(S,\mathcal{S})$ to $(T,\mathcal{T})$ so that the relation*

$$\int_{S\times T} f(s,t)\,\mu(\mathrm{d}(s,t))=\int_S\int_T f(s,t)\,K(s,\mathrm{d}t)\,\nu(\mathrm{d}s) \tag{16}$$

holds for any non-negative measurable function f on $S\times T$. If K' is another kernel from $(S,\mathcal{S})$ to $(T,\mathcal{T})$ with this property, then $\nu(\{s\in S: K(s,B)\neq K'(s,B)\})=0$ for any $B\in\mathcal{T}$.

Proof: For every $B\in\mathcal{T}$, the measure $\mu(\cdot\times B)$ on $(S,\mathcal{S})$ is absolutely continuous w.r.t. ν. It is so because $\mu(A\times B)\le\nu(A)$. Hence, there is its Radon–Nikodym derivative $\xi_B=\frac{\mathrm{d}\mu(\cdot\times B)}{\mathrm{d}\nu(\cdot)}$, which can be chosen to satisfy $\xi_B(s)\in[0,1]$ for every $s\in S$ (again because $\mu(A\times B)\le\nu(A)$). In particular, $\xi_T=1$ a.s.

First let us assume that ν is a probability measure (i.e. $\nu(S)=1$). Then $\{\xi_B,B\in\mathcal{T}\}$ is a family of non-negative random variable on a probability space $(S,\mathcal{S},\nu)$. This family satisfies the assumptions of Proposition 25. The additivity $\xi_{B\cup B'}=\xi_B+\xi_{B'}$ a.s. for disjoint $B,B'\in\mathcal{T}$ follows from $\mu(\cdot\times(B\cup B'))=\mu(\cdot\times B)+\mu(\cdot\times B')$. For the sets $B_n\in\mathcal{T}$ satisfying $B_n\searrow\emptyset$, we have $\mu(S\times B_n)=\int_S\xi_{B_n}(s)\,\nu(\mathrm{d}s)=\mathbb{E}\xi_{B_n}\underset{n\to\infty}{\longrightarrow}0$ and so the ξ_{B_n} converge to zero in L_1. Since $\xi_{B_1}\ge\cdots\ge\xi_{B_n}\ge\cdots$, they converge also a.s. Now by Proposition 25, we obtain a random measure Ψ on T such that $\Psi(B)=\xi_B$ a.s. for every $B\in\mathcal{T}$. Moreover, Ψ is a probability measure a.s. (because $\xi_T=1$ a.s.) and we may put $K(s,B)=\Psi(s)(B)$. From the definition of the Radon–Nikodym derivative, we have

$$\mu(A\times B)=\int_A K(s,B)\,\nu(\mathrm{d}s),\quad A\in\mathcal{S},B\in\mathcal{T}.$$

This is a special case of (16) for $f=\mathbf{1}_{A\times B}$. In order to show (16) for arbitrary f on $S\times T$, we proceed by standard measure theory arguments.

The procedure can be easily generalized to the case of finite measure ν ($\nu(S)<\infty$). We just work with the measure $\tilde\mu(\cdot)=\frac{\mu(\cdot)}{\nu(S)}$ and the probability measure $\tilde\nu(\cdot)=\tilde\mu(\cdot\times T)=\frac{\nu(\cdot)}{\nu(S)}$. For σ-finite measure ν, we use the spaces $(S_n,\mathcal{S}_n,\nu_n)$, where $S_n\nearrow S$, $\nu(S_n)<\infty$ and $\nu_n=\nu|_{S_n}$.

The uniqueness follows from the fact that the Radon–Nikodym derivative is unique ν-a.s.

□

Definition 54. We define the *Campbell measure* associated with a random measure Ψ as

$$C(A)=\mathbb{E}\int_E\mathbf{1}_A(x,\Psi)\,\Psi(\mathrm{d}x),\quad A\in\mathcal{B}\times\mathfrak{M}.$$

In particular,

$$C(B \times \mathcal{U}) = \mathbb{E}\mathbf{1}_{\mathcal{U}}(\Psi)\Psi(B), \quad B \in \mathcal{B},\, \mathcal{U} \in \mathfrak{M}.$$

Remark 31. From

$$C(A) = \int_{\mathcal{M}} \int_E \mathbf{1}_A(x,\mu)\, \mu(\mathrm{d}x)\, Q(\mathrm{d}\mu)$$

we get by standard measure theory arguments,

$$\int_{E\times\mathcal{M}} f(x,\mu)\, C(\mathrm{d}(x,\mu)) = \int_{\mathcal{M}} \int_E f(x,\mu)\, \mu(\mathrm{d}x)\, Q(\mathrm{d}\mu) = \mathbb{E}\int_E f(x,\Psi)\, \Psi(\mathrm{d}x),$$

where Q is the distribution of the random measure Ψ and f is an arbitrary non-negative measurable function on $E \times \mathcal{M}$.

Corollary 47. *Let Ψ be a random measure on E with distribution Q and intensity measure $\Lambda \in \mathcal{M}$. Then there exists a Markov kernel P from $(E,\mathcal{B})$ to $(\mathcal{M},\mathfrak{M})$ satisfying*

$$\int_{\mathcal{M}} \int_E f(x,\mu)\, \mu(\mathrm{d}x)\, Q(\mathrm{d}\mu) = \int_E \int_{\mathcal{M}} f(x,\mu)\, P(x,\mathrm{d}\mu)\, \Lambda(\mathrm{d}x) \tag{17}$$

for an arbitrary non-negative measurable function f on $E \times \mathcal{M}$. If P' is another Markov kernel with this property, then $\Lambda(\{x \in E : P(x,\mathcal{U}) \neq P'(x,\mathcal{U})\}) = 0$ for any $\mathcal{U} \in \mathfrak{M}$.

Proof: Since $\mathcal{M}$ forms a complete separable metric space (Proposition 22), we can use Theorem 46 with $S = E$, $T = \mathcal{M}$, $\nu = \Lambda$ and $\mu = C$. The measure Λ is σ-finite because it is assumed to be locally finite (Lemma 17).

□

Definition 55. If P is the Markov kernel from Corollary 47, then the distribution $P_x(\cdot) = P(x,\cdot)$ is called *Palm distribution* of the random measure Ψ at point $x \in E$.

Remark 32. Actually, it does not make sense to speak about the Palm distribution at one particular point because this distribution could be defined arbitrarily. Nevertheless, Corollary 47 assures that the family $\{P_x : x \in E\}$ of Palm distributions is uniquely determined for Λ-a.a. x. If $\{P_x\}$ and $\{\tilde{P}_x\}$ are two Palm distributions of the random measure Ψ, then for any $\mathcal{U} \in \mathfrak{M}$ we have $P_x(\mathcal{U}) = \tilde{P}_x(\mathcal{U})$ for Λ-a.a. $x \in E$.

Lemma 48. *For a point process Φ with intensity measure $\Lambda \in \mathcal{M}$, we have $P_x(\{\mu \in \mathcal{M} : \mu(\{x\}) \geq 1\}) = 1$ for Λ-a.a. $x \in E$.*

Proof: Take $f(x,\mu) = \mathbf{1}_A(x)\mathbf{1}_{[\mu(\{x\})\geq 1]}$ with arbitrary $A \in \mathcal{B}_0$. Then the definition of Palm distribution gives

$$\begin{aligned}\int_A P_x(\{\mu \in \mathcal{M} : \mu(\{x\}) \geq 1\})\, \Lambda(\mathrm{d}x) &= \int_{\mathcal{M}} \int_A \mathbf{1}_{[\mu(\{x\})\geq 1]}\, \mu(\mathrm{d}x)\, Q(\mathrm{d}\mu) \\ &= \int_{\mathcal{M}} \mu(A)\, Q(\mathrm{d}\mu) = \mathbb{E}\Phi(A) = \Lambda(A).\end{aligned}$$

In the second step we used that Φ is a point process with distribution Q. Therefore,

$$\int_A [1 - P_x(\{\mu \in \mathcal{M} : \mu(\{x\}) \geq 1\})]\, \Lambda(\mathrm{d}x) = 0,$$

which implies $P_x(\{\mu \in \mathcal{M} : \mu(\{x\}) \geq 1\}) = 1$ for Λ-a.a. $x \in A$.

□

Remark 33. If Q is the distribution of a point process, then the P_x are distributions of point processes for Λ-a.a. $x \in E$. The desintegration theorem is used for $\mathcal{N}$ instead of $\mathcal{M}$.

Definition 56. For a point process Φ we define the *reduced Palm distribution* at point x as a probability measure $P_x^!$ given by the relation

$$\int_{\mathcal{N}} g(\nu)\, P_x^!(\mathrm{d}\nu) = \int_{\mathcal{N}} g(\nu - \delta_x)\, P_x(\mathrm{d}\nu)$$

for an arbitrary non-negative measurable function g. In particular, $P_x^!(\mathcal{U}) = P_x(\mathcal{U} + \delta_x)$, where $\mathcal{U} + \delta_x = \{\mu + \delta_x : \mu \in \mathcal{U}\}$.

Theorem 49. *Assume that Φ is a simple point process (i.e. $Q(\mathcal{N}^*) = 1$). Then $P_x(\mathcal{N}^*) = 1$ for Λ-a.a. $x \in E$.*

Proof: For arbitrary $A \in \mathcal{B}_0$, we get by the definition of Palm distribution

$$\int_A P_x(\mathcal{M} \setminus \mathcal{N}^*)\,\Lambda(\mathrm{d}x) = \int_{\mathcal{M}} \int_E \mathbf{1}_{\mathcal{M}\setminus\mathcal{N}^*}(\mu)\mathbf{1}_A(x)\,\mu(\mathrm{d}x)\,Q(\mathrm{d}\mu) = \int_{\mathcal{M}\setminus\mathcal{N}^*} \mu(A)\,Q(\mathrm{d}\mu) = 0.$$

The last equation follows from $Q(\mathcal{M} \setminus \mathcal{N}^*) = 0$. Now we see that $P_x(\mathcal{M} \setminus \mathcal{N}^*) = 0$ for Λ-a.a. $x \in A$.

□

Remark 34. The Palm distribution P_x of a simple point process Φ can be interpreted as the conditional distribution of Φ under the condition that x is an atom of the point process. For $\varepsilon > 0$ small, we have

$$\mathbb{P}(\Phi \in \mathcal{U} \mid \Phi(b(x,\varepsilon)) > 0) = \frac{\mathbb{P}(\Phi \in \mathcal{U}, \Phi(b(x,\varepsilon)) > 0)}{\mathbb{P}(\Phi(b(x,\varepsilon)) > 0)} \approx \frac{\mathbb{E}\mathbf{1}_{\mathcal{U}}(\Phi)\Phi(b(x,\varepsilon))}{\mathbb{E}\Phi(b(x,\varepsilon))} = \frac{C(b(x,\varepsilon)\times\mathcal{U})}{\Lambda(b(x,\varepsilon))} \approx P_x(\mathcal{U}),$$

where $b(x,\varepsilon)$ denotes the ball of center x and radius ε. Lemma 7.2 in [14] provides mathematically rigorous proof. Similarly, $P_x^!$ can be interpreted as the conditional distribution of a point process under the condition that x is an atom that is not counted.

Remark 35. In the theory of point processes, the term *typical point* is often used. Its meaning can be interpreted by the Palm distribution. We say that a typical point x has some property if it has this property under the Palm distribution P_x.

Theorem 50. *(Campbell–Mecke theorem) Let Φ be a simple point process. For an arbitrary non-negative measurable function h,*

$$\mathbb{E}\sum_{X\in\operatorname{supp}\Phi} h(X,\Phi) = \int_{E\times\mathcal{N}^*} h(x,\nu)\,C(\mathrm{d}(x,\nu)) = \int_E\int_{\mathcal{N}^*} h(x,\nu)\,P_x(\mathrm{d}\nu)\,\Lambda(\mathrm{d}x)$$

and

$$\mathbb{E}\sum_{X\in\operatorname{supp}\Phi} h(X,\Phi-\delta_X) = \int_E\int_{\mathcal{N}^*} h(x,\nu)\,P_x^!(\mathrm{d}\nu)\,\Lambda(\mathrm{d}x).$$

Proof: The first equation in the first relation can be shown by the standard measure theory arguments. For $h(x,\nu) = \mathbf{1}_A(x)\mathbf{1}_{\mathcal{U}}(\nu)$ we have

$$\mathbb{E}\sum_{X\in\operatorname{supp}\Phi} h(X,\Phi) = \mathbb{E}\Phi(A)\mathbf{1}_{\mathcal{U}}(\Phi) = C(A\times\mathcal{U}).$$

The second equation follows from Corollary 47. The second relation can be deduced from the first one and Definition 56.

□

Theorem 51. *(Slivnyak theorem) Let Φ be a Poisson point process having distribution Π_Λ and intensity measure $\Lambda \in \mathcal{M}$. Then $P_x = \Pi_\Lambda * \delta_{(\delta_x)}$, i.e. $P_x^! = \Pi_\Lambda$, for Λ-a.a. $x \in E$. The symbol $*$ stands for the convolution of measures. In other words, $\Phi + \delta_x$ has the distribution P_x for Λ-a.a. $x \in E$.*

Proof: The general proof can be found in [6], Proposition 13.1.VII or [14], Theorem 7.3. Here we show the proof for diffuse Λ. In this case, Lemma 38 states that $\Phi + \delta_x$ is a simple point process. We show that the distributions of $\Phi + \delta_x$ form a Markov kernel that satisfies (17). Using Lemma 30, it is enough to consider $f(x,\mu) = \mathbf{1}_B(x)\mathbf{1}_{\mathcal{N}^*_{K,0}}(\mu)$, $B \in \mathcal{B}$, $K \in \mathcal{K}$, in (17). It means that we want to verify that

$$C(B\times\mathcal{N}^*_{K,0}) = \int_B \mathbb{P}((\Phi+\delta_x)(K) = 0)\,\Lambda(\mathrm{d}x).$$

The left-hand side is $C(B \times \mathcal{N}^*_{K,0}) = \mathbb{E}\Phi(B)\mathbf{1}_{[\Phi(K)=0]}$. The right-hand side can be rewritten as

$$\begin{aligned}\int_B \mathbb{P}(\Phi(K)=0, x \notin K)\,\Lambda(\mathrm{d}x) &= \int_{B\setminus K} \mathbb{P}(\Phi(K)=0)\,\Lambda(\mathrm{d}x)\\ &= \Lambda(B\setminus K)\mathbb{P}(\Phi(K)=0) = \mathbb{E}\Phi(B\setminus K)\mathbf{1}_{[\Phi(K)=0]},\end{aligned}$$

where in the last step we have used the independence of $\Phi(B\setminus K)$ and $\Phi(K)$. Clearly, $\mathbb{E}\Phi(B)\mathbf{1}_{[\Phi(K)=0]} = \mathbb{E}\Phi(B\setminus K)\mathbf{1}_{[\Phi(K)=0]}$, which completes the proof.

□

4. Spatial point processes

In this section, we consider the Euclidean space $E = \mathbb{R}^d$ with the standard Euclidean metric. The corresponding Borel σ-algebra $\mathcal{B}(E)$ will be denoted by $\mathcal{B}^d$. The family of bounded Borel sets will be denoted by $\mathcal{B}^d_0$ and the family of compact sets by $\mathcal{K}^d$. A point process on $\mathbb{R}^d$ will be referred to as the *spatial point process*. It serves as a useful model for describing a collection of events occurring at random locations in space.

4.1 Stationary spatial point processes

Definition 57. For $z \in \mathbb{R}^d$ we denote by t_z the shift operator on $\mathcal{M}$. It is given by the relation

$$(t_z\mu)(A) = \mu(A-z), \quad \mu \in \mathcal{M}, A \in \mathcal{B}^d.$$

We say that a random measure Ψ on $\mathbb{R}^d$ is *stationary* if $t_z\Psi$ and Ψ have the same distribution for all $z \in \mathbb{R}^d$, i.e. the distribution of the random measure is translation-invariant.

Remark 36. If $\mu = \sum_{i=1}^{\tau} \delta_{x_i}$, then $t_z\mu = \sum_{i=1}^{\tau} \delta_{x_i+z}$.

Definition 58. For a rotation $\mathcal{O}$ around the origin, we denote by $R_{\mathcal{O}}$ the rotation operator on $\mathcal{M}$. It is defined as

$$(R_{\mathcal{O}}\mu)(A) = \mu(\mathcal{O}^{-1}A), \quad \mu \in \mathcal{M}, A \in \mathcal{B}^d.$$

A random measure Ψ on $\mathbb{R}^d$ is called *isotropic* if $R_{\mathcal{O}}\Psi$ and Ψ have the same distribution for any rotation $\mathcal{O}$, i.e. the distribution of the random measure is rotation-invariant.

Definition 59. A random measure is called *motion-invariant* if it is both stationary and isotropic.

Definition 60. We say that a random measure Ψ on $\mathbb{R}^d$ is *first-order stationary* if $t_z\Lambda = \Lambda$ for all $z \in \mathbb{R}^d$, where Λ is the intensity measure of Ψ. We say that Ψ is *n-th order stationary* if its k-th order moment measures $M^{(k)}$, $k = 1, \dots, n$, are invariant under diagonal shifts, i.e. $M^{(k)}((B_1+z)\times\cdots\times(B_k+z)) = M^{(k)}(B_1\times\cdots\times B_k)$ for all $z \in \mathbb{R}^d$, $B_1, \dots, B_n \in \mathcal{B}^d$ and $k = 1, \dots, n$.

Obviously, each stationary random measure is n-th order stationary for any $n \in \mathbb{N}$.

Lemma 52. *If Ψ is a first-order stationary random measure with locally finite intensity measure Λ, then Λ is a non-negative multiple of the Lebesgue measure, i.e. $\Lambda(B) = \lambda|B|$, $B \in \mathcal{B}^d$, for some $\lambda \geq 0$.*

Proof: From first-order stationarity we have $\Lambda(B+z) = \Lambda(B)$ for each $z \in \mathbb{R}^d$. The assertion follows from the fact that up to a multiplicative constant, the Lebesgue measure is the only translation-invariant and locally finite measure on $(\mathbb{R}^d, \mathcal{B}^d)$, see Exercise 26.6 in [11].

□

Definition 61. The scalar multiple $\lambda \geq 0$ from Lemma 52 is called the *intensity* of a first-order stationary random measure.

The intensity gives the mean measure per unit volume. In case of point processes, it is the expected number of points per unit volume.

Definition 62. Let Ψ be a random measure with intensity measure Λ. If there exists a density λ of Λ w.r.t. the Lebesgue measure (i.e. $\Lambda(B) = \int_B \lambda(x)\,\mathrm{d}x$, $B \in \mathcal{B}^d$), then λ is called the *intensity function*.

From Lemma 52 we know that the intensity measure of a stationary random measure is the Lebesgue measure multiplied by the intensity. Hence, the intensity function is constant and is equal to the intensity.

Definition 63. The Poisson point process Φ on $\mathbb{R}^d$, for which the intensity function λ exists and is constant, is called the *homogeneous* Poisson point process with intensity λ. If this intensity is furthermore equal to 1, then we speak about the *standard* Poisson point process.

Remark 37. Every homogeneous Poisson point process is motion-invariant (see Exercise class).

Definition 64. Let Ψ be a random measure on $\mathbb{R}^d$ with the n-th order factorial moment measure $\alpha^{(n)}$. If there exists a density $\lambda^{(n)}$ of $\alpha^{(n)}$ w.r.t. the (nd)-dimensional Lebesgue measure, then it is called the *n-th order product density*.

Remark 38. The first-order product density coincides with the intensity function and we write $\lambda^{(1)} = \lambda$.

Remark 39. A heuristic interpretation of the n-th order product density of a simple point process is the following. Consider n infinitesimally small disjoint balls with centres $x_1, \dots, x_n$ and volumes $\mathrm{d}x_1, \dots, \mathrm{d}x_n$. Then $\lambda^{(n)}(x_1, \dots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$ is the probability that each of these balls contains a point of the process.

Corollary 53. *If there exists the intensity function λ of the Poisson point process Φ, then its n-th order product density satisfies $\lambda^{(n)}(x_1, \dots, x_n) = \prod_{i=1}^n \lambda(x_i)$, $x_1, \dots, x_n \in \mathbb{R}^d$.*

Proof: The assertion follows from Theorem 42. □

From the definition of stationarity (Definition 57), it follows that the moment measures of a stationary random measure are invariant under diagonal shifts. In other words,

$$M^{(n)}(B_1 \times \cdots \times B_n) = M^{(n)}\big((B_1 + y) \times \cdots \times (B_n + y)\big)$$

and

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = \alpha^{(n)}\big((B_1 + y) \times \cdots \times (B_n + y)\big)$$

for arbitrary $n \in \mathbb{N}$, $B_1, \dots, B_n \in \mathcal{B}^d$ and $y \in \mathbb{R}^d$. If the n-th order product density exists, then it satisfies

$$\lambda^{(n)}(x_1, \dots, x_n) = \lambda^{(n)}(x_1 + y, \dots, x_n + y) \tag{18}$$

for almost all $x_1, \dots, x_n \in \mathbb{R}^d$ and $y \in \mathbb{R}^d$.

Theorem 54. *Let Ψ be a stationary random measure on $\mathbb{R}^d$ with intensity $0 < \lambda < \infty$. Choose arbitrary bounded Borel set $A \in \mathcal{B}_0^d$ with positive Lebesgue measure ($|A| > 0$). For $\mathcal{U} \in \mathfrak{M}$ and $x \in \mathbb{R}^d$, let $t_x^{-1}\mathcal{U} = \{\mu : t_x\mu \in \mathcal{U}\}$ be the preimage of $\mathcal{U}$ under t_x. Then*

$$P_o(\mathcal{U}) = \frac{1}{\lambda|A|}\mathbb{E}\int_A \mathbf{1}_{\mathcal{U}}(t_{-x}\Psi)\,\Psi(\mathrm{d}x), \quad \mathcal{U} \in \mathfrak{M},$$
$$P_x(\mathcal{U}) = P_o(t_x^{-1}\mathcal{U}), \quad \mathcal{U} \in \mathfrak{M},$$

define Palm distributions of Ψ.

Proof: We verify that this system of distributions satisfies (17), and so it meets the definition of Palm distribution. It is enough to take $f(x, \mu) = \mathbf{1}_B(x)\mathbf{1}_{\mathcal{U}}(\mu)$ in (17). From stationarity we know that Λ is proportional to the Lebesgue measure (Lemma 52). Hence,

$$\begin{aligned}\int_B P_x(\mathcal{U})\,\Lambda(\mathrm{d}x) &= \lambda\int_B P_o(t_x^{-1}\mathcal{U})\,\mathrm{d}x = \lambda\int_B \frac{1}{\lambda|A|}\mathbb{E}\int_A \mathbf{1}_{t_x^{-1}\mathcal{U}}(t_{-y}\Psi)\,\Psi(\mathrm{d}y)\,\mathrm{d}x \\ &= \frac{1}{|A|}\int_B \mathbb{E}\int_A \mathbf{1}_{\mathcal{U}}(t_{x-y}\Psi)\,\Psi(\mathrm{d}y)\,\mathrm{d}x = \frac{1}{|A|}\mathbb{E}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_A(y)\mathbf{1}_B(y+z)\mathbf{1}_{\mathcal{U}}(t_z\Psi)\,\Psi(\mathrm{d}y)\,\mathrm{d}z,\end{aligned}$$

where in the last step we used Fubini's theorem and the change of variables $z = x - y$. Now we make the substitution $\Psi(\mathrm{d}y) = (t_z\Psi)(\mathrm{d}x)$ and employ stationarity of Ψ. We obtain

$$\begin{aligned}\int_B P_x(\mathcal{U})\,\Lambda(\mathrm{d}x) &= \frac{1}{|A|}\mathbb{E}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_A(x-z)\mathbf{1}_B(x)\mathbf{1}_{\mathcal{U}}(t_z\Psi)\,(t_z\Psi)(\mathrm{d}x)\,\mathrm{d}z \\ &= \frac{1}{|A|}\mathbb{E}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_A(x-z)\mathbf{1}_B(x)\mathbf{1}_{\mathcal{U}}(\Psi)\,\Psi(\mathrm{d}x)\,\mathrm{d}z \\ &= \mathbb{E}\int_B \mathbf{1}_{\mathcal{U}}(\Psi)\,\Psi(\mathrm{d}x) = C(B \times \mathcal{U}).\end{aligned}$$

□

If we speak about the Palm distribution of a stationary random measure, then we mean the system of distributions $\{P_x, x \in \mathbb{R}^d\}$ from Theorem 54. This system is determined by the Palm distribution P_o in the origin and by the relation $P_x(\cdot) = P_o(t_x^{-1}\cdot)$. Similarly, we consider the reduced Palm distribution of a stationary simple point process Φ in the form

$$P_o^!(\mathcal{U}) = \frac{1}{\lambda|A|}\mathbb{E}\sum_{X\in\operatorname{supp}\Phi\cap A}\mathbf{1}_{\mathcal{U}}(t_{-X}(\Phi-\delta_X)), \quad \mathcal{U}\in\mathfrak{N}^*,$$
$$P_x^!(\mathcal{U}) = P_o^!(t_x^{-1}\mathcal{U}), \quad \mathcal{U}\in\mathfrak{N}^*.$$

The Campbell–Mecke theorem then has the following form.

Theorem 55. *For a stationary simple point process Φ with intensity λ and an arbitrary non-negative measurable function h on $\mathbb{R}^d\times\mathcal{N}^*$,*

$$\mathbb{E}\sum_{X\in\operatorname{supp}\Phi}h(X,\Phi) = \lambda\int_{\mathbb{R}^d}\int_{\mathcal{N}^*}h(x,\nu)\,P_x(\mathrm{d}\nu)\,\mathrm{d}x = \lambda\int_{\mathbb{R}^d}\int_{\mathcal{N}^*}h(x,t_x\nu)\,P_o(\mathrm{d}\nu)\,\mathrm{d}x$$

and

$$\mathbb{E}\sum_{X\in\operatorname{supp}\Phi}h(X,\Phi-\delta_X) = \lambda\int_{\mathbb{R}^d}\int_{\mathcal{N}^*}h(x,\nu)\,P_x^!(\mathrm{d}\nu)\,\mathrm{d}x = \lambda\int_{\mathbb{R}^d}\int_{\mathcal{N}^*}h(x,t_x\nu)\,P_o^!(\mathrm{d}\nu)\,\mathrm{d}x.$$

Remark 40. The expectation w.r.t. P_o will be denoted by the symbol $\mathbb{E}_o$. Similarly, $\mathbb{E}_o^!$ stands for the expectation w.r.t. $P_o^!$. It means that

$$\mathbb{E}_o h(\Phi) = \int_{\mathcal{N}^*}h(\nu)\,P_o(\mathrm{d}\nu) \quad\text{and}\quad \mathbb{E}_o^! h(\Phi) = \int_{\mathcal{N}^*}h(\nu)\,P_o^!(\mathrm{d}\nu).$$

4.2 Point process characteristics

In this subsection, we deal with summary characteristics of spatial point processes. We can distinguish numerical and functional characteristics. Numerical characteristics describe specific properties of the point process by a single number. For stationary point processes, the simplest and most important numerical characteristic is the intensity (Definition 61). Modern point process statistics often uses functional summary characteristics. An example of a functional characteristic could be the intensity function (Definition 62). Intensity and intensity function are first-order characteristics (both are derived from the first-order moment measure). The intensity function is also defined for non-stationary point processes. However, we focus mainly on stationary point processes in this subsection.

First, we mention two examples of numerical summary characteristics that are based on mutual distances between the atoms of a point process.

Definition 65. Let Φ be a stationary simple point process with intensity λ. By D_o denote the distance from the origin o to the nearest point of Φ, i.e. $D_o = \inf\{r\ge 0 : \Phi(b(o,r)) > 0\}$. We define *Pielou's index of randomness* as

$$\mathrm{PI} = \lambda\omega_d\mathbb{E}(D_o)^d$$

and the *aggregation index* or also the *Clark–Evans index* by the relation

$$\mathrm{CE} = \frac{d(\lambda\omega_d)^{1/d}}{\Gamma(1/d)}\mathbb{E}_o^! D_o,$$

where $\omega_d = |b(o,1)| = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$ is the volume of d-dimensional unit ball.

These indices are closely related to the following functional summary characteristics.

Definition 66. Let Φ be a stationary simple point process with intensity λ and let D_o be the distance from o to the nearest point of Φ. The *spherical contact distribution function* is defined as

$$F(r) = \mathbb{P}(\Phi(b(o,r)) > 0) = \mathbb{P}(D_o \le r), \quad r > 0.$$

The *nearest-neighbour distance distribution function* is defined as

$$G(r) = P_o^!(\{\nu \in \mathcal{N} : \nu(b(o,r)) > 0\}), \quad r > 0.$$

Furthermore, the *J-function* is

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \quad r > 0 : F(r) < 1.$$

Remark 41. The generalization of F and G functions can be obtained by replacing balls $b(o,r)$ with sets rB, where B is some convex and compact set containing o.

The aggregation index and the G-function are so-called point-related characteristics. We are interested in the distances from a typical point of the process to the nearest point of the process (called neighbour). In contrast, Pielou's index of randomness and the F-function are location-related characteristics. We look at the distance from some fixed location (e.g. origin) to the nearest point of the process.

Theorem 56. *In case of a homogeneous Poisson point process with intensity λ, we get* $\mathrm{PI} = \mathrm{CE} = 1$, $F(r) = G(r) = 1 - \mathrm{e}^{-\lambda\omega_d r^d}$, *and* $J(r) = 1$.

Proof: Left to Exercise class.

□

Remark 42. The deviations of characteristics from the theoretical values for the Poisson point process indicate either clustering of points or a tendency towards regularity. Cluster processes have Pielou's index of randomness larger than 1, the Clark-Evans index smaller than 1, and values of the J-function smaller than 1. For regular processes the situation is reversed: $\mathrm{PI} < 1$, $\mathrm{CE} > 1$, and $J(r) > 1$.

Characteristics defined so far are "short-sighted" as they take into account only distances to the nearest points. The distances to more distant points are disregarded. However, the spatial correlation often exhibits among the points that are close in space. Another class of characteristics is based on the second-order moment measures. Such characteristics can be defined more generally for random measures.

Definition 67. Let Ψ be a stationary random measure on $\mathbb{R}^d$ with intensity $0 < \lambda < \infty$. The *reduced second-order moment measure* $\mathcal{K}$ is defined by the relation

$$\lambda\mathcal{K}(B) = \int_{\mathcal{M}} \mu(B \setminus \{o\})\, P_o(\mathrm{d}\mu), \quad B \in \mathcal{B}^d.$$

Furthermore, we define the *reduced second-order moment function* or shortly the *K-function* as

$$K(r) = \mathcal{K}(b(o,r)), \quad r \geq 0.$$

Remark 43. In particular, if Φ is a stationary simple point process, then $\lambda\mathcal{K}(B) = \mathbb{E}_o^!\Phi(B)$, $B \in \mathcal{B}^d$. It means that $\lambda K(r)$ can be interpreted as the expected number of points (distinct from o) of the process in the centred ball of radius r under the condition that o is the point of the process.

Lemma 57. *For a stationary random measure Ψ with intensity $0 < \lambda < \infty$, we have*

$$\alpha^{(2)}(A \times B) = \lambda^2 \int_A \mathcal{K}(B - x)\, \mathrm{d}x, \quad A, B \in \mathcal{B}^d.$$

Proof: The right-hand side can be rewritten by Definition 67, Lemma 52, the relation between P_x and P_o from Theorem 54, Corollary 47, and Definition 51:

$$\begin{aligned}
\lambda^2 \int_A \mathcal{K}(B - x)\, \mathrm{d}x &= \int_A \int_{\mathcal{M}} \mu((B - x) \setminus \{o\})\, P_o(\mathrm{d}\mu)\, \Lambda(\mathrm{d}x) \\
&= \int_A \int_{\mathcal{M}} (t_{-x}\mu)((B - x) \setminus \{o\})\, P_x(\mathrm{d}\mu)\, \Lambda(\mathrm{d}x) = \int_{\mathcal{M}} \int_A \mu(B \setminus \{x\})\, \mu(\mathrm{d}x)\, Q(\mathrm{d}\mu) \\
&= \mathbb{E} \int_A \Psi(B \setminus \{x\})\, \Psi(\mathrm{d}x) = \mathbb{E} \int_A \int_B \mathbf{1}_{[x \neq y]}\, \Psi(\mathrm{d}y)\, \Psi(\mathrm{d}x) \\
&= \mathbb{E}\Psi^{[2]}(A \times B) = \alpha^{(2)}(A \times B).
\end{aligned}$$

The relation holds for arbitrary $A, B \in \mathcal{B}^d$.

□

Lemma 58. *Let Ψ be a stationary random measure with intensity $0 < \lambda < \infty$ and let $A \in \mathcal{B}_0^d$ be an arbitrary bounded Borel set with finite Lebesgue measure $|A| > 0$. Then*

$$\mathcal{K}(B) = \frac{1}{\lambda^2 |A|} \mathbb{E} \int_A \Psi((B + x) \setminus \{x\})\, \Psi(\mathrm{d}x), \quad B \in \mathcal{B}^d.$$

Proof: Similarly as in the proof of Lemma 57, we get

$$\begin{aligned} \lambda^2 \int_A \mathcal{K}(B)\, \mathrm{d}x &= \int_A \int_{\mathcal{M}} \mu(B \setminus \{o\})\, P_o(\mathrm{d}\mu)\, \Lambda(\mathrm{d}x) \\ &= \int_A \int_{\mathcal{M}} \mu((B + x) \setminus \{x\})\, P_x(\mathrm{d}\mu)\, \Lambda(\mathrm{d}x) \\ &= \mathbb{E} \int_A \Psi((B + x) \setminus \{x\})\, \Psi(\mathrm{d}x). \end{aligned}$$

□

For a stationary simple point process Φ, we can write

$$\begin{aligned} \mathcal{K}(B) &= \frac{1}{\lambda^2 |A|} \mathbb{E} \sum_{X \in \operatorname{supp} \Phi \cap A} \Phi((B + X) \setminus \{X\}) = \frac{1}{\lambda^2 |A|} \mathbb{E} \sum_{X \in \operatorname{supp} \Phi} \sum_{Y \in \operatorname{supp} \Phi} \mathbf{1}_A(X) \mathbf{1}_{(B+X) \setminus \{X\}}(Y) \\ &= \frac{1}{\lambda^2 |A|} \mathbb{E} \sum_{X, Y \in \operatorname{supp} \Phi}^{\neq} \mathbf{1}_A(X) \mathbf{1}_B(Y - X), \quad B \in \mathcal{B}^d. \end{aligned}$$

This relation may be used as an alternative definition of $\mathcal{K}(B)$.

Corollary 59. *For a stationary Poisson point process, $\mathcal{K}(B) = |B|$ for all $B \in \mathcal{B}^d$. In particular, $K(r) = \omega_d r^d$ for all $r \geq 0$.*

Proof: From Definition 67 and Theorem 51, we get $\lambda \mathcal{K}(B) = \mathbb{E}\Phi(B) = \lambda |B|$, $B \in \mathcal{B}^d$. Therefore, $\mathcal{K}(B) = |B|$ and specially $K(r) = \omega_d r^d$.

□

Definition 68. Often instead of the K-function its transformation is used,

$$L(r) = \left(\frac{K(r)}{\omega_d} \right)^{1/d}, \quad r \geq 0,$$

and it is called the *L-function*. It contains the same information as $K(r)$ but its popularity comes from more advantageous graphical interpretation. From Corollary 59 we know that $L(r) = r$ in case of the Poisson point process. The graphical comparison of the L-function with a line ($L(r) = r$) can be better visualised than the comparison of the K-function with a parabolic or higher degree polynomial curve ($K(r) = \omega_d r^d$). Furthermore, there are also some statistical advantages of using the L-function.

Definition 69. Let Ψ be a random measure on $\mathbb{R}^d$ (not necessarily stationary). Under the assumption that the intensity function λ and the second-order product density $\lambda^{(2)}$ exist, we define the *pair correlation function* by the relation

$$g(x, y) = \frac{\lambda^{(2)}(x, y)}{\lambda(x) \lambda(y)}, \quad x, y \in \mathbb{R}^d : \lambda(x) > 0, \lambda(y) > 0.$$

When Ψ is stationary, we know by (18) that $\lambda^{(2)}(x, y) = \lambda^{(2)}(x - y, o) = \lambda^{(2)}(x - y)$ and $g(x, y) = \frac{\lambda^{(2)}(x-y)}{\lambda^2} = g(x - y)$ are the functions of $x - y$. If Ψ is moreover isotropic, then $\lambda^{(2)}(x, y) = \lambda^{(2)}(\|x - y\|)$ and $g(x, y) = g(\|x - y\|)$ are the functions of Euclidean distance between x and y. Note that here and in the sequel, we use slight abuse of notation and we use the same symbol for the function of the difference $x - y$ or the distance $\|x - y\|$ as for the function of the pair (x, y).

Remark 44. The pair correlation function may attain values from the interval $[0, \infty)$. Therefore, the term "correlation function" could be somewhat misleading.

The pair correlation function has a particularly simple form in the case of the Poisson point process.

Corollary 60. *Let Φ be a Poisson point process whose intensity function λ exists. Then $g(x, y) = 1$ for any $x, y \in \mathbb{R}^d$ with $\lambda(x)\lambda(y) > 0$.*

Proof: From Corollary 53 we have $\lambda^{(2)}(x, y) = \lambda(x)\lambda(y)$, which yields $g(x, y) = 1$.

□

Remark 45. For point processes, the pair correlation function provides a valuable tool for graphical demonstration of deviations from the Poisson point process. If $g(x, y) > 1$, then the simultaneous occurrence of points x and y is more likely than for the Poisson process with the same intensity function. For motion-invariant point processes, the values of pair correlation function $g(r)$ greater than 1 correspond to the fact that the distances r among the points are more typical than in the model of complete spatial randomness. In other words, $g(r) > 1$ indicates aggregation of points at distances r, while $g(r) < 1$ means regularity in the corresponding distances r.

Theorem 61. *Let Ψ be a stationary random measure with pair correlation function g. Then $\mathcal{K}$ is absolutely continuous w.r.t. the Lebesgue measure and*

$$\mathcal{K}(B) = \int_B g(u)\,\mathrm{d}u, \quad B \in \mathcal{B}^d.$$

Proof: According to Lemma 58, we have

$$\begin{aligned}
\mathcal{K}(B) &= \frac{1}{\lambda^2|A|}\mathbb{E}\int_A \Psi((B+x)\setminus\{x\})\,\Psi(\mathrm{d}x) = \frac{1}{\lambda^2|A|}\mathbb{E}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{[x\in A, y-x\in B]}\,\Psi^{[2]}(\mathrm{d}x,\mathrm{d}y)\\
&= \frac{1}{\lambda^2|A|}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{[x\in A, y-x\in B]}\,\alpha^{(2)}(\mathrm{d}x,\mathrm{d}y)\\
&= \frac{1}{\lambda^2|A|}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{[x\in A, y-x\in B]}\lambda^{(2)}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{|A|}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{[x\in A, y-x\in B]}g(y-x)\,\mathrm{d}x\,\mathrm{d}y\\
&= \frac{1}{|A|}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{[x\in A, u\in B]}g(u)\,\mathrm{d}x\,\mathrm{d}u = \int_B g(u)\,\mathrm{d}u, \quad B \in \mathcal{B}^d.
\end{aligned}$$

We used Campbell's theorem (Theorem 43), Definition 64, and Definition 69.

□

Corollary 62. *Let Ψ be a stationary random measure with pair correlation function that is invariant under rotations, i.e. $g(x, y) = g(y - x) = g(\|y - x\|)$. Then*

$$g(r) = \frac{K'(r)}{\sigma_d r^{d-1}},$$

where $K'(r)$ is the derivative of K-function and $\sigma_d = d\omega_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the surface area of unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$.

Proof: From Theorem 61 we obtain

$$K(r) = \int_{b(o,r)} g(u)\,\mathrm{d}u.$$

Using the polar decomposition of the Lebesgue measure (spherical coordinates), we get

$$K(r) = \int_{b(o,r)} g(\|u\|)\,\mathrm{d}u = \int_0^r \sigma_d s^{d-1} g(s)\,\mathrm{d}s.$$

□

4.3 Point process models

Cluster point processes

Denote by $\mathcal{N}_{\mathrm{f}}^* = \mathcal{N}_{\mathrm{f}} \cap \mathcal{N}^*$ the space of all simple finite counting measures on $\mathbb{R}^d$.

Definition 70. Let $\tilde{\Phi}$ be a simple point process on $\mathbb{R}^d \times \mathcal{N}_{\mathrm{f}}^*$. Assume that the point process $\Phi_p(\cdot) = \tilde{\Phi}(\cdot \times \mathcal{N}_{\mathrm{f}}^*)$ is a simple point process (so-called *parent point process*). Define

$$\Phi(B) = \sum_{(X,\zeta)\in \operatorname{supp}\tilde{\Phi}} (t_X\zeta)(B), \quad B \in \mathcal{B}^d,$$

and assume that $\Phi(K) < \infty$ for all $K \in \mathcal{K}^d$ with probability 1. Then Φ is a point process a.s. and it is called a *cluster point process*. For $(X, \zeta) \in \operatorname{supp}\tilde{\Phi}$, we refer to $t_X\zeta$ as the *daughter point process* associated with the *parent point* X.

It means that a cluster point process could be constructed in two steps. In the first step, we consider a parent point process. In the second step, each parent point generates a daughter point process. The superposition of all daughter points creates the resulting process.

Definition 71. Let $\Phi_p = \sum_{i=1}^{\Phi_p(\mathbb{R}^d)} \delta_{X_i}$ be a simple point process on $\mathbb{R}^d$. Let $\zeta_1, \zeta_2, \ldots$ be independent simple finite point processes and independent of Φ_p. Then a cluster point process generated by $\tilde{\Phi} = \sum_{i=1}^{\Phi_p(\mathbb{R}^d)} \delta_{(X_i,\zeta_i)}$ is called a *point process with independent clusters*. If moreover Φ_p is a Poisson point process, then we speak about the *Poisson cluster point process*.

The following theorem shows that point processes with independent Poisson clusters are Cox point processes as well.

Theorem 63. *Let Φ be a point process with independent clusters and let the ζ_i be finite Poisson point processes with diffuse intensity measures Λ_i. Assume that $\Psi(K) = \sum_{i=1}^{\Phi_p(\mathbb{R}^d)} (t_{X_i}\Lambda_i)(K) < \infty$ for all $K \in \mathcal{K}^d$ a.s. Then Φ has the same distribution as a Cox point process with driving intensity measure Ψ.*

Proof: For $K \in \mathcal{K}^d$, we get

$$\begin{aligned}
\mathbb{P}(\Phi(K) = 0) &= \mathbb{E}\left[\mathbb{P}(\Phi(K) = 0 \mid \Phi_p)\right] = \mathbb{E}\left[\mathbb{P}\left(\bigcap_{i=1}^{\Phi_p(\mathbb{R}^d)} [(t_{X_i}\zeta_i)(K) = 0] \mid \Phi_p\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\prod_{i=1}^{\Phi_p(\mathbb{R}^d)} \mathbf{1}_{[\zeta_i(K-X_i)=0]} \mid \Phi_p\right)\right] = \mathbb{E}\prod_{i=1}^{\Phi_p(\mathbb{R}^d)} \mathrm{e}^{-\Lambda_i(K-X_i)} \\
&= \mathbb{E}\exp\left\{-\sum_{i=1}^{\Phi_p(\mathbb{R}^d)} (t_{X_i}\Lambda_i)(K)\right\} = \mathbb{E}\mathrm{e}^{-\Psi(K)},
\end{aligned}$$

which is the void probability of a Cox point process (see Remark 25). The assertion follows from Theorem 32.

□

In the following definition, we introduce an important class of Poisson cluster point processes.

Definition 72. Let p be a probability density on $\mathbb{R}^d$ and let Φ be a Poisson cluster point process satisfying

- $\zeta_i(\mathbb{R}^d)$ are independent identically distributed random variables,
- the processes ζ_i are mixed binomial point processes formed by independent identically distributed points according to the density p.

Then Φ is called the *Neyman–Scott point process*. If moreover $\zeta_i(\mathbb{R}^d)$ has Poisson distribution with parameter λ_c, then Φ is called the *Neyman–Scott Poisson point process*. In the latter case, ζ_i is the Poisson point process with intensity function $\lambda_c p(x)$. The proof of this fact is left to Exercise class.

In Figure 12 (left), there is an explanation of the Neyman–Scott process construction. The parent point process Φ_p is a stationary Poisson point process (the points are shown by crosses), the numbers of daughter points are independent identically distributed random variables with binomial distribution with mean $\mathbb{E}\zeta_i(\mathbb{R}^d) = 5$, and the daughter points are independent identically distributed random vectors with uniform distribution in the square centred at the parent point and of size length 0.2, i.e. the density p has the form $p(x) = 1/0.2^2$ for $x \in [-0.1, 0.1]^2$. Since the supports of individual daughter densities are overlapping, we are not able to distinguish which point belongs to which cluster (daughter process). Furthermore, we can notice that the daughter points in the observation window may come from the parent points that lie outside the window. The resulting realization of the Neyman–Scott process is shown in Figure 12 (right). In this realization no parent points appear.

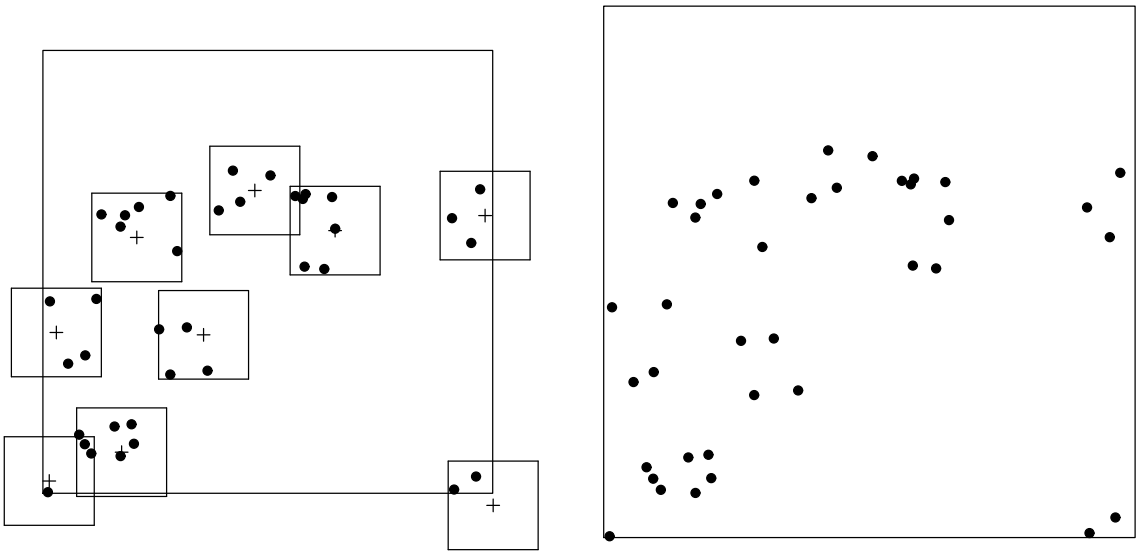

Figure 12. A realization of the Neyman–Scott point process in a unit square window. The intensity of the parent process is 10 (the points are shown by crosses in the left figure), the numbers of daughter points have the binomial distribution with parameters 10 and 1/2, and the daughter points (bullets) are uniformly distributed in the square of size length 0.2 (shown in the left figure). The resulting realization of the point process (only daughter points) is shown on right.

Theorem 64. *Let Φ be a Neyman–Scott Poisson point process such that Φ_p is stationary Poisson point process with intensity λ_p. Then Φ is a stationary point process with intensity $\lambda = \lambda_p\lambda_c$ and pair correlation function*

$$g(x) = 1 + \frac{h(x)}{\lambda_p},$$

where $h(x) = \int p(y)p(y-x)\,\mathrm{d}y$ is the density of the difference $X_1 - X_2$ of two independent daughter points X_1 and X_2. If p is radially symmetric function (that is, $p(x) = p(\|x\|) = p(r)$), then Φ is also isotropic and $g(x) = g(\|x\|) = g(r)$.

Proof: Stationarity of Φ follows from stationarity of Φ_p and the fact that the centred daughter processes ζ_i are independent and identically distributed. By Theorem 63, Φ is the Cox point process with driving measure $\Psi(B) = \sum_{X\in\operatorname{supp}\Phi_p} \Lambda_X(B)$, where $\Lambda_X = t_X\Lambda$ and Λ has the density $\lambda_c p(x)$ w.r.t. the Lebesgue measure. The intensity measure of Φ can be expressed using Theorem 43 as

$$\begin{aligned}\mathbb{E}\,\Psi(B) &= \mathbb{E}\sum_{X\in\operatorname{supp}\Phi_p} \Lambda_X(B) = \lambda_p \int_{\mathbb{R}^d} \Lambda(B-x)\,\mathrm{d}x = \lambda_p \int_{\mathbb{R}^d}\int_{B-x} \lambda_c p(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \lambda_p\lambda_c \int_B \int_{\mathbb{R}^d} p(u-x)\,\mathrm{d}x\,\mathrm{d}u = \lambda_p\lambda_c|B| = \lambda|B|.\end{aligned}$$

Similarly, by the second-order Campbell theorem (Theorem 43) and Theorem 42, we obtain the second-order factorial moment measure of Φ:

$$\begin{aligned}
\mathbb{E}\,\Psi(B_1)\Psi(B_2) &= \mathbb{E}\sum_{X,Y\in\operatorname{supp}\Phi_p}\Lambda(B_1-X)\Lambda(B_2-Y)\\
&= \mathbb{E}\sum_{X,Y\in\operatorname{supp}\Phi_p}^{\neq}\Lambda(B_1-X)\Lambda(B_2-Y) + \mathbb{E}\sum_{X\in\operatorname{supp}\Phi_p}\Lambda(B_1-X)\Lambda(B_2-X)\\
&= \lambda_p^2\lambda_c^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{B_1}\int_{B_2} p(u-x)p(v-y)\,\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}y\,\mathrm{d}x\\
&\quad + \lambda_p\lambda_c^2\int_{\mathbb{R}^d}\int_{B_1}\int_{B_2} p(u-x)p(v-x)\,\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}x\\
&= \lambda^2\cdot|B_1|\cdot|B_2| + \lambda\lambda_c\int_{B_1}\int_{B_2}\int_{\mathbb{R}^d} p(u-x)p(v-x)\,\mathrm{d}x\,\mathrm{d}v\,\mathrm{d}u.
\end{aligned}$$

Hence, the second-order product density exists and has the form

$$\lambda^{(2)}(u,v) = \lambda^2 + \lambda\lambda_c\int_{\mathbb{R}^d} p(u-x)p(v-x)\,\mathrm{d}x.$$

The pair correlation function is then

$$g(u,v) = 1 + \frac{1}{\lambda_p}\int_{\mathbb{R}^d} p(u-x)p(v-x)\,\mathrm{d}x.$$

The change of variables $y = u - x$ yields

$$g(u,v) = 1 + \frac{1}{\lambda_p}\int_{\mathbb{R}^d} p(y)p(v-u+y)\,\mathrm{d}y$$

and we see that $g(u,v)$ is the function of $u-v$: $g(u,v) = g(u-v,o) = g(u-v)$.

□

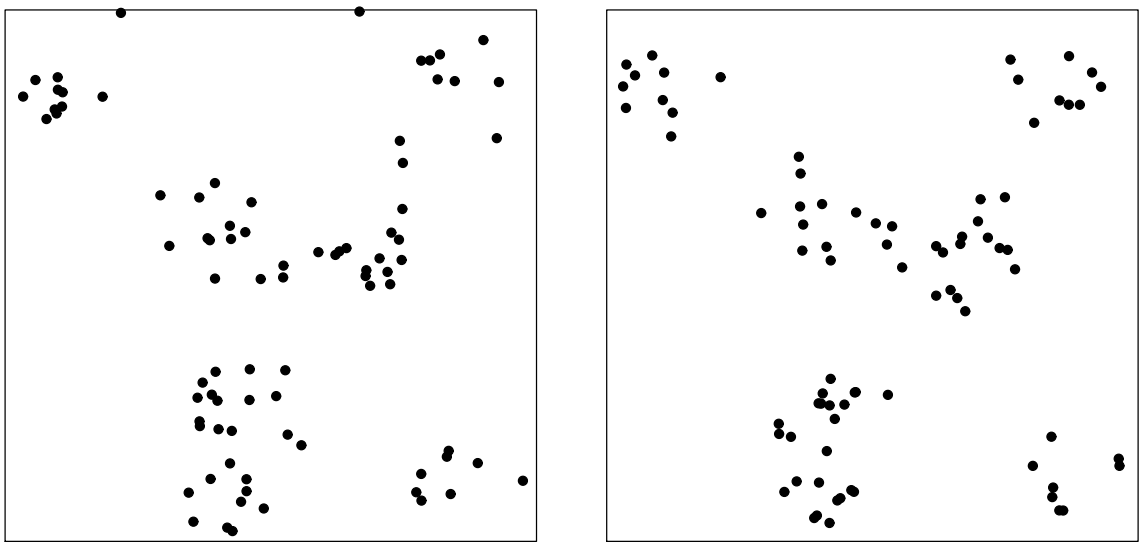

Figure 13. The realizations of Thomas point process (left) and Matérn cluster point process (right) in the unit square. The parent point process is the same in both cases and has an intensity 10. The numbers of points per cluster have Poisson distribution with parameter 5 and in both cases, they are the same for every parent point. The only difference is in the distribution of daughter points. For the Thomas process, it is the normal distribution with a standard deviation $\sigma = 0.05$, while for the Matérn cluster process, the daughter points are uniformly distributed in the ball of radius $R = 0.1$ centred at the parent point.

Notice that the pair correlation function of the Neyman–Scott Poisson point process is always at least 1. The most often used choices of radially symmetric density p give the following point processes:

a) *Thomas point process*: p is the density of d-dimensional normal distribution $N_d(0, \sigma^2 I)$ with zero mean and variance matrix that is a multiple of the identity matrix,

b) *Matérn cluster point process*: p is the density of uniform distribution on the ball $b(o, R)$.

Particular examples of realizations of these models can be found in Figure 13.

Thinned point processes

One of the simplest operations associated with point processes is *thinning*. It is based on deleting some points of the process. We consider the case of so-called *independent thinning*, where the decision about removing or keeping the point is independent of other points.

Definition 73. Let $\Phi_p = \sum_{i=1}^{\Phi_p(\mathbb{R}^d)} \delta_{X_i}$ be a simple point process. Let $U_1, U_2, \ldots$ be independent random variables uniformly distributed on $(0,1)$ and independent of Φ_p. Let $p : \mathbb{R}^d \to [0,1]$ be a measurable function. Then the point process

$$\Phi = \sum_{i=1}^{\Phi_p(\mathbb{R}^d)} \mathbf{1}_{[U_i < p(X_i)]} \delta_{X_i}$$

is called an *independently thinned point process* or a *p-thinning of* Φ_p.

Theorem 65. *A p-thinning of a simple point process Φ_p with intensity measure Λ_p is a point process Φ with intensity measure*

$$\Lambda(B) = \int_B p(x)\, \Lambda_p(\mathrm{d}x), \quad B \in \mathcal{B}^d.$$

Proof: We can write

$$\begin{aligned}\mathbb{E}\Phi(B) &= \mathbb{E}\big[\mathbb{E}[\Phi(B) \mid \Phi_p]\big] = \mathbb{E}\big[\mathbb{E}[\sum_{i:X_i \in B} \mathbf{1}_{[U_i < p(X_i)]} \mid \Phi_p]\big] \\ &= \mathbb{E} \sum_{i:X_i \in B} p(X_i) = \int_B p(x)\, \Lambda_p(\mathrm{d}x).\end{aligned}$$

We have used independence of $\{U_i\}$ and $\Phi_p = \sum_{i=1}^{\Phi_p(\mathbb{R}^d)} \delta_{X_i}$, and Campbell's theorem (Theorem 43). □

Theorem 66. *Let Φ_p be a simple point process with pair correlation function g. Then a p-thinning of Φ_p has the same pair correlation function.*

Proof: Denote by $\lambda_p(x)$ and $\lambda_p^{(2)}(x,y)$ the intensity function and the second-order product density of Φ_p, respectively. Let Φ be a p-thinning of Φ_p. From Theorem 65 we obtain that the intensity function of Φ is $\lambda(x) = p(x)\lambda_p(x)$. The second-order product density $\lambda^{(2)}(x,y)$ of Φ could be determined by similar arguments as in the proof of Theorem 65. Since

$$\begin{aligned}\mathbb{E}\Phi^{[2]}(B_1 \times B_2) &= \mathbb{E}\big[\mathbb{E}[\Phi^{[2]}(B_1 \times B_2) \mid \Phi_p]\big] = \mathbb{E}\big[\mathbb{E}[\sum_{i,j:X_i \in B_1, X_j \in B_2}^{\neq} \mathbf{1}_{[U_i < p(X_i)]} \mathbf{1}_{[U_j < p(X_j)]} \mid \Phi_p]\big] \\ &= \mathbb{E} \sum_{i,j:X_i \in B_1, X_j \in B_2}^{\neq} p(X_i)p(X_j) = \int_{B_1} \int_{B_2} p(x)p(y)\lambda_p^{(2)}(x,y)\, \mathrm{d}y\, \mathrm{d}x,\end{aligned}$$

we see that $\lambda^{(2)}(x,y) = p(x)p(y)\lambda_p^{(2)}(x,y)$. Therefore,

$$\frac{\lambda^{(2)}(x,y)}{\lambda(x)\lambda(y)} = \frac{p(x)p(y)\lambda_p^{(2)}(x,y)}{p(x)\lambda_p(x)p(y)\lambda_p(y)} = \frac{\lambda_p^{(2)}(x,y)}{\lambda_p(x)\lambda_p(y)} = g(x,y).$$

□

By an independent thinning of a Poisson point process, we get again a Poisson point process.

Theorem 67. *Let Φ_p be a Poisson point process with diffuse intensity measure $\Lambda_p \in \mathcal{M}$. A p-thinning of Φ_p is a Poisson point process Φ with intensity measure*

$$\Lambda(B) = \int_B p(x)\,\Lambda_p(\mathrm{d}x), \quad B \in \mathcal{B}^d.$$

Proof: We compute the void probabilities of the thinned point process Φ and we show that they are equal to the void probabilities of a Poisson point process with intensity measure Λ. The assertion then follows from Theorem 32. For $K \in \mathcal{K}^d$ we have

$$\begin{aligned}
\mathbb{P}(\Phi(K)=0) &= \sum_{n=0}^{\infty} \mathbb{P}(\Phi_p(K)=n)\mathbb{P}(\Phi(K)=0 \mid \Phi_p(K)=n) \\
&= \sum_{n=0}^{\infty} \frac{\Lambda_p(K)^n}{n!} \mathrm{e}^{-\Lambda_p(K)} \mathbb{P}\left(\bigcap_{i:X_i\in K} [U_i \geq p(X_i)] \mid \Phi_p(K)=n \right) \\
&= \mathrm{e}^{-\Lambda_p(K)} \sum_{n=0}^{\infty} \frac{\Lambda_p(K)^n}{n!} \left[\int_0^1 \int_K \mathbf{1}_{[u\geq p(x)]}\, \frac{\Lambda_p(\mathrm{d}x)}{\Lambda_p(K)}\,\mathrm{d}u \right]^n \\
&= e^{-\Lambda_p(K)} \sum_{n=0}^{\infty} \frac{1}{n!} \left[\int_K (1-p(x))\,\Lambda_p(\mathrm{d}x) \right]^n \\
&= \mathrm{e}^{-\Lambda_p(K)} \exp\left\{ \int_K (1-p(x))\,\Lambda_p(\mathrm{d}x) \right\} = \mathrm{e}^{-\int_K p(x)\,\Lambda_p(\mathrm{d}x)}.
\end{aligned}$$

In the third line, we expressed the conditional probability by use of Theorem 39, which states that conditionally on $\Phi_p(K)=n$, the atoms $X_1,\dots,X_n$ of the process Φ_p are independent and identically distributed according to the probability measure $\Lambda_p|_K/\Lambda_p(K)$.

□

Hard-core point processes

Many point process models result from thinning of a given point process. In Definition 73 we considered independent thinning.

For hard-core point processes there are no points in distance smaller than a given distance $r>0$. The most natural construction is to use some thinning rule. Since the atoms are removed or kept depending on their location in relation to other atoms of the process, we speak about the *dependent thinning*. Such procedure is used in the definition of the following two hard-core point processes.

Definition 74. Let Φ be a stationary Poisson point process with intensity λ and let $r>0$ (so-called *hard-core distance*) be given. The point process

$$\Phi_I(\cdot) = \int_{\mathbb{R}^d} \delta_x(\cdot)\mathbf{1}_{[\Phi(b(x,r)\setminus\{x\})=0]}\,\Phi(\mathrm{d}x)$$

is called the *Matérn hard-core point process of type I*. Let Φ_{m} be a Poisson point process on $\mathbb{R}^d\times[0,1]$ with intensity measure $\Lambda(B\times I)=\lambda\cdot|B|\cdot|I|$, $B\in\mathcal{B}^d$, $I\in\mathcal{B}([0,1])$. The point process

$$\Phi_{II}(\cdot) = \int_{\mathbb{R}^d\times[0,1]} \delta_x(\cdot)\mathbf{1}_{[\Phi_{\mathrm{m}}((b(x,r)\setminus\{x\})\times[0,u])=0]}\,\Phi_{\mathrm{m}}(\mathrm{d}(x,u))$$

is called the *Matérn hard-core point process of type II*.

For the first type, we remove every point of the process Φ for which there is at least one another point at distance smaller than r. In the second type, each point $X\in\operatorname{supp}\Phi$ has associated some weight $U(X)\in[0,1]$ and we remove only the points for which there exists at least one another point with smaller weight at distance smaller than r. Equivalently, we can imagine that for each pair of distinct points $X,Y\in\operatorname{supp}\Phi$ satisfying $0<\|X-Y\|\leq r$, we remove either both points (type I) or only the point with larger weight (type II).

In Figure 14, a realization of stationary Poisson point process is depicted. By its thinning, Matérn hard-core point processes of types I and II are obtained. Obviously, we have fewer points in the process of type I than in type II due to a stricter thinning condition.

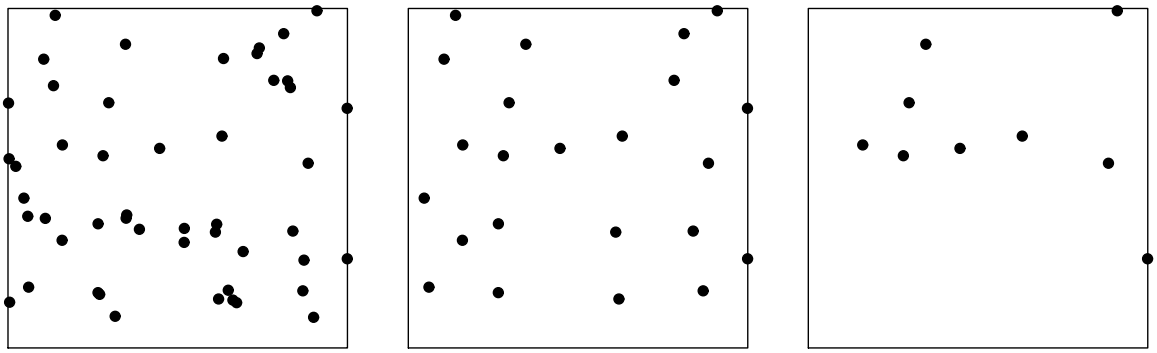

Figure 14. Left: a realization of stationary Poisson point process with intensity 50 in the unit square. From this configuration of points, the realizations of Matérn hard-core point processes of type II (middle) and type I (right) are derived. The hard-core distance is chosen to be $r = 0.1$.

Theorem 68. *Point processes Φ_I and Φ_{II} from Definition 74 are stationary and their intensities are*

$$\lambda_I = \lambda \mathrm{e}^{-\lambda\omega_d r^d} \quad \textit{and} \quad \lambda_{II} = \frac{1-\mathrm{e}^{-\lambda\omega_d r^d}}{\omega_d r^d},$$

respectively.

Proof: Stationarity follows from the construction. We determine the intensities by computing the expected numbers of point in the set $B \in \mathcal{B}_0^d$.

First, consider the type I process. From Campbell–Mecke theorem and Slivnyak theorem, we get

$$\begin{aligned}\mathbb{E}\Phi_I(B) &= \mathbb{E}\int_B \mathbf{1}_{[\Phi(b(x,r)\setminus\{x\})=0]}\,\Phi(\mathrm{d}x) = \lambda\int_B\int_{\mathcal{N}} \mathbf{1}_{[\nu(b(x,r))=0]}\,P_x^!(\mathrm{d}\nu)\,\mathrm{d}x\\ &= \lambda\int_B \mathbb{P}(\Phi(b(x,r))=0)\,\mathrm{d}x = \lambda\mathrm{e}^{-\lambda\omega_d r^d}|B|.\end{aligned}$$

For type II process, we again use Campbell–Mecke theorem and Slivnyak theorem. We get

$$\begin{aligned}\mathbb{E}\Phi_{II}(B) &= \mathbb{E}\int_{B\times[0,1]} \mathbf{1}_{[\Phi_{\mathrm{m}}((b(x,r)\setminus\{x\})\times[0,u])=0]}\,\Phi_{\mathrm{m}}(\mathrm{d}(x,u)) = \lambda\int_B\int_0^1 \mathbb{P}(\Phi_{\mathrm{m}}(b(x,r)\times[0,u])=0)\,\mathrm{d}u\,\mathrm{d}x\\ &= \lambda|B|\int_0^1 \mathrm{e}^{-\lambda u\omega_d r^d}\,\mathrm{d}u = \lambda|B|\frac{1-\mathrm{e}^{-\lambda\omega_d r^d}}{\lambda\omega_d r^d},\end{aligned}$$

which leads to the required relation $\mathbb{E}\Phi_{II}(B) = \lambda_{II}|B|$.

□

Another possibility how to construct a hard-core point process is based on the sequential approach. The points are added one-by-one.

Definition 75. Let $r > 0$ and $B \in \mathcal{B}_0^d$ be given. A *random sequential adsorption (or simple sequential inhibition)* model is the point process inside the set B constructed by the following algorithm:
(i) generate X_1 from the uniform distribution on B,
(ii) if $k-1$ points are generated, then generate X_k from the uniform distribution in $B \setminus \cup_{i=1}^{k-1} b(X_i, r)$,
(iii) the construction ends in n steps provided that $B \subseteq \cup_{i=1}^{n} b(X_i, r)$.
The output is the point process $\Phi = \sum_{i=1}^{n} \delta_{X_i}$.

Remark 46. The sets $B \setminus \cup_{i=1}^{k-1} b(X_i, r)$ may have quite complicated geometrical shapes. Hence, it is common in practice to use the rejection method. The point X_k is generated uniformly in B and it is

rejected whenever it falls closer than r to any of the points $X_1, \dots, X_{k-1}$. In that case, a new candidate X_k is generated and the procedure is repeated until it falls in $B \setminus \cup_{i=1}^{k-1} b(X_i, r)$. The algorithm is usually terminated if the desired number of points is reached or the number of consecutive rejections reached some prescribed value. Three examples of the realizations for different choices of r are shown in Figure 15.

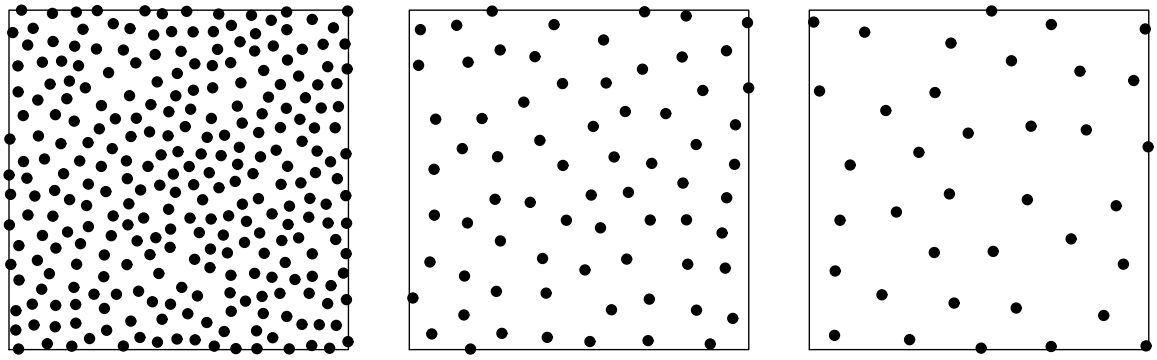

Figure 15. Three realizations of the random sequential adsorption point process in the unit square for $r = 0.05$ (left), $r = 0.1$ (middle), and $r = 0.15$ (right).

Definition 75 gives an example of a finite point process. In the following subsection, we will deal with finite point processes given by the density w.r.t. the distribution of a Poisson point process.

4.4 Finite point processes with density

In this subsection, we will work with finite point processes in $\mathbb{R}^d$. Recall that the system of finite counting measures is denoted by

$$\mathcal{N}_{\mathrm{f}} = \{\nu \in \mathcal{N} : \nu(\mathbb{R}^d) < \infty\}$$

and the system of simple finite counting measures is $\mathcal{N}_{\mathrm{f}}^* = \mathcal{N}_{\mathrm{f}} \cap \mathcal{N}^*$. Let $\mathfrak{N}_{\mathrm{f}}^* = \{\mathcal{U} \cap \mathcal{N}_{\mathrm{f}}^* : \mathcal{U} \in \mathfrak{N}\}$ be the trace of $\mathfrak{N}$ on $\mathcal{N}_{\mathrm{f}}^*$.

Let Φ_P be a Poisson point process with finite diffuse intensity measure Λ. We know that Φ_P is finite (because $\Lambda(\mathbb{R}^d) < \infty$) and simple (because Λ is diffuse). The distribution Π_Λ of Φ_P can be expressed as follows (for $\mathcal{U} \in \mathfrak{N}_{\mathrm{f}}^*$):

$$\begin{aligned}
\Pi_\Lambda(\mathcal{U}) = \mathbb{P}(\Phi_P \in \mathcal{U}) &= \sum_{n=0}^{\infty} \mathbb{P}(\Phi_P(\mathbb{R}^d) = n)\mathbb{P}(\Phi_P \in \mathcal{U} \mid \Phi_P(\mathbb{R}^d) = n) \\
&= \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[\mathbf{1}_{\mathcal{U}}(\varnothing) + \sum_{n=1}^{\infty} \frac{\Lambda(\mathbb{R}^d)^n}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{U}}\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \frac{\Lambda(\mathrm{d}x_1)}{\Lambda(\mathbb{R}^d)} \cdots \frac{\Lambda(\mathrm{d}x_n)}{\Lambda(\mathbb{R}^d)} \right] \\
&= \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[\mathbf{1}_{\mathcal{U}}(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{U}}\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \Lambda(\mathrm{d}x_1) \cdots \Lambda(\mathrm{d}x_n) \right]. \qquad (19)
\end{aligned}$$

We are going to investigate point processes whose distribution is absolutely continuous w.r.t. Π_Λ. They can be specified by their density function.

Definition 76. Let Π_Λ be the distribution of a finite simple Poisson point process Φ_P. Assume that $p : \mathcal{N}_{\mathrm{f}}^* \to \mathbb{R}^+$ is a measurable function such that $\int_{\mathcal{N}_{\mathrm{f}}^*} p(\nu)\, \Pi_\Lambda(\mathrm{d}\nu) = 1$. We say that Φ is a *point process with density* p w.r.t. Π_Λ if its distribution satisfies

$$\mathbb{P}(\Phi \in \mathcal{U}) = \int_{\mathcal{U}} p(\nu)\, \Pi_\Lambda(\mathrm{d}\nu) = \mathbb{E}\mathbf{1}_{\mathcal{U}}(\Phi_P) p(\Phi_P), \quad \mathcal{U} \in \mathfrak{N}_{\mathrm{f}}^*.$$

Thanks to (19), we can rewrite this distribution as

$$\mathbb{P}(\Phi \in \mathcal{U}) = \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[\mathbf{1}_{\mathcal{U}}(\varnothing) p(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{U}}\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) p\Big(\sum_{i=1}^{n} \delta_{x_i}\Big) \Lambda(\mathrm{d}x_1) \cdots \Lambda(\mathrm{d}x_n) \right],$$

where Λ is the intensity measure of Φ_P.

Often we consider Λ to be the restriction of the Lebesgue measure to some bounded set $B \in \mathcal{B}_0^d$. The density p is then w.r.t. the distribution of the standard Poisson point process on B.

Notice that $\mathbb{E}h(\Phi) = \mathbb{E}h(\Phi_P)p(\Phi_P)$ for arbitrary Π_Λ-integrable function h on $\mathcal{N}_{\mathrm{f}}^*$.

Example: Consider the density p in the form

$$p(\sum_{i=1}^{n} \delta_{x_i}) = \alpha \prod_{i=1}^{n} \beta(x_i),$$

where $\alpha > 0$ and $\beta : \mathbb{R}^d \to \mathbb{R}^+$ is a measurable function such that $\int_{\mathbb{R}^d} \beta(x)\,\Lambda(\mathrm{d}x) < \infty$. By an empty product ($n = 0$) we understand one, i.e. $p(\varnothing) = \alpha$. Define a finite measure

$$\Lambda_\beta(B) = \int_B \beta(x)\,\Lambda(\mathrm{d}x), \quad B \in \mathcal{B}^d.$$

We make sure that $\int_{\mathcal{N}_{\mathrm{f}}^*} p(\nu)\,\Pi_\Lambda(\mathrm{d}\nu) = 1$ for a suitable choice of α. The distribution of the point process Φ with density p is

$$\begin{aligned}\mathbb{P}(\Phi \in \mathcal{U}) &= \alpha \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[\mathbf{1}_{\mathcal{U}}(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{U}}(\sum_{i=1}^{n} \delta_{x_i}) \beta(x_1) \cdots \beta(x_n)\,\Lambda(\mathrm{d}x_1) \cdots \Lambda(\mathrm{d}x_n) \right] \\ &= \alpha \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[\mathbf{1}_{\mathcal{U}}(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{U}}(\sum_{i=1}^{n} \delta_{x_i})\,\Lambda_\beta(\mathrm{d}x_1) \cdots \Lambda_\beta(\mathrm{d}x_n) \right].\end{aligned}$$

By comparison with (19), we see that it is again the distribution of a Poisson point process. The intensity measure is now Λ_β and the normalising constant α must be $\alpha = \mathrm{e}^{\Lambda(\mathbb{R}^d) - \Lambda_\beta(\mathbb{R}^d)} = \mathrm{e}^{\int_{\mathbb{R}^d} (1-\beta(x))\,\Lambda(\mathrm{d}x)}$.

For the definition of a point process with density, we can use an arbitrary non-negative measurable function h with $0 < \alpha = \int_{\mathcal{N}_{\mathrm{f}}^*} h(\nu)\,\Pi_\Lambda(\mathrm{d}\nu) < \infty$. If we divide by α, we get the density function $p = h/\alpha$. The equality up to the normalising constant is written as $p(\nu) \propto h(\nu)$. In order to recognize whether h is Π_Λ-integrable several sufficient conditions are useful.

Definition 77. Let Λ be a finite diffuse measure on $\mathbb{R}^d$. We say that a measurable function $h : \mathcal{N}_{\mathrm{f}}^* \to \mathbb{R}^+$ is *locally stable* if there exists a non-negative measurable function $k : \mathbb{R}^d \to \mathbb{R}^+$ with $\int_{\mathbb{R}^d} k(x)\,\Lambda(\mathrm{d}x) < \infty$ so that

$$h(\nu + \delta_x) \le k(x)h(\nu) \quad \forall \nu \in \mathcal{N}_{\mathrm{f}}^*, x \in \mathbb{R}^d \setminus \operatorname{supp} \nu.$$

The function h is *Ruelle stable* if there exist a constant $c \ge 0$ and a non-negative measurable function k with $\int_{\mathbb{R}^d} k(x)\,\Lambda(\mathrm{d}x) < \infty$ so that

$$h(\nu) \le c \prod_{x \in \operatorname{supp} \nu} k(x) \quad \forall \nu \in \mathcal{N}_{\mathrm{f}}^*.$$

If we can obtain $\nu \in \mathcal{N}_{\mathrm{f}}^*$ from $\mu \in \mathcal{N}_{\mathrm{f}}^*$ by omitting some of the atoms, we write $\nu \preceq \mu$. This relation is equivalent to $\operatorname{supp} \nu \subseteq \operatorname{supp} \mu$.

Definition 78. A non-negative measurable function h on $\mathcal{N}_{\mathrm{f}}^*$ is called *hereditary* if the implication

$$h(\mu) > 0 \Rightarrow h(\nu) > 0 \quad \forall \nu \preceq \mu$$

holds for any $\mu \in \mathcal{N}_{\mathrm{f}}^*$.

Theorem 69. *Let $h : \mathcal{N}_{\mathrm{f}}^* \to \mathbb{R}^+$ be a measurable function and let Π_Λ be the distribution of a finite Poisson point process with diffuse intensity measure Λ.*

a) If h is locally stable, then it is Ruelle stable.

b) If h is Ruelle stable, then it is Π_Λ-integrable.

c) If h is locally stable, then it is hereditary.

Proof: a) Put $c = h(\varnothing)$. The assertion is easily shown by induction on $n = \nu(\mathbb{R}^d)$. For $n = 0$ it is obvious that $h(\nu) = c \le c$. If the claim is true for some n, then

$$h(\nu + \delta_x) \le k(x)h(\nu) \le k(x)c \prod_{y \in \operatorname{supp} \nu} k(y) = c \prod_{y \in \operatorname{supp}(\nu + \delta_x)} k(y).$$

b) Denote $K = \int_{\mathbb{R}^d} k(x)\, \Lambda(\mathrm{d}x)$. Then

$$\begin{aligned}\int_{\mathcal{N}_{\mathrm{f}}^*} h(\nu)\, \Pi_\Lambda(\mathrm{d}\nu) &= \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[h(\varnothing) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} h(\sum_{i=1}^{n} \delta_{x_i})\, \Lambda(\mathrm{d}x_1) \cdots \Lambda(\mathrm{d}x_n) \right] \\ &\le \mathrm{e}^{-\Lambda(\mathbb{R}^d)} \left[c + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} ck(x_1) \cdots k(x_n)\, \Lambda(\mathrm{d}x_1) \cdots \Lambda(\mathrm{d}x_n) \right] \\ &= c\mathrm{e}^{-\Lambda(\mathbb{R}^d)} \sum_{n=0}^{\infty} \frac{1}{n!} K^n = c\mathrm{e}^{-\Lambda(\mathbb{R}^d)} \mathrm{e}^K < \infty.\end{aligned}$$

The first equality follows from (19) by the standard measure theory arguments.
c) From the definition of local stability we have $h(\nu + \delta_x) > 0 \Rightarrow h(\nu) > 0$.

□

Definition 79. For a finite simple point process Φ with density p, the *Papangelou conditional intensity* is defined as

$$\lambda^*(x, \nu) = \frac{p(\nu + \delta_x)}{p(\nu)}, \quad x \in \mathbb{R}^d \setminus \operatorname{supp} \nu,\ \nu \in \mathcal{N}_{\mathrm{f}}^* : p(\nu) > 0.$$

Remark 47. The local stability of a hereditary density function p can be equivalently defined by the condition $\lambda^*(x, \nu) \le k(x)$ for every $x \in \mathbb{R}^d$ and $\nu \in \mathcal{N}_{\mathrm{f}}^*$ with $p(\nu) > 0$, where $\int_{\mathbb{R}^d} k(x)\, \Lambda(\mathrm{d}x) < \infty$.

Theorem 70. *(Georgii–Nguyen–Zessin identity) Let Π_Λ be the distribution of a Poisson point process with finite diffuse intensity measure Λ. Let Φ be a finite simple point process with density p w.r.t. Π_Λ and let λ^* be its Papangelou conditional intensity. Then*

$$\mathbb{E} \sum_{X \in \operatorname{supp} \Phi} h(X, \Phi - \delta_X) = \int_{\mathbb{R}^d} \mathbb{E} h(x, \Phi) \lambda^*(x, \Phi)\, \Lambda(\mathrm{d}x) \tag{20}$$

for an arbitrary non-negative measurable function h on $\mathbb{R}^d \times \mathcal{N}_{\mathrm{f}}^*$.

Proof: We can write

$$\mathbb{E} \sum_{X \in \operatorname{supp} \Phi} h(X, \Phi - \delta_X) = \mathbb{E} \sum_{X \in \operatorname{supp} \Phi_P} h(X, \Phi_P - \delta_X) p(\Phi_P),$$

where Φ_P is the Poisson point process with distribution Π_Λ. Now we can apply the Campbell–Mecke theorem (Theorem 50) and the Slivnyak theorem (Theorem 51),

$$\mathbb{E} \sum_{X \in \operatorname{supp} \Phi_P} h(X, \Phi_P - \delta_X) p(\Phi_P) = \int_{\mathbb{R}^d} \mathbb{E} h(x, \Phi_P) p(\Phi_P + \delta_x)\, \Lambda(\mathrm{d}x).$$

The integrand can be rewritten as

$$\mathbb{E} h(x, \Phi_P) p(\Phi_P + \delta_x) = \mathbb{E} h(x, \Phi_P) \lambda^*(x, \Phi_P) p(\Phi_P) = \mathbb{E} h(x, \Phi) \lambda^*(x, \Phi),$$

which completes the proof.

□

Corollary 71. *Let Π be the distribution of the standard Poisson point process on a bounded set $B \in \mathcal{B}_0^d$. Let Φ be a finite point process with density p w.r.t. Π and Papangelou conditional intensity λ^*. Then its intensity function is $\lambda(x) = \mathbb{E}\lambda^*(x, \Phi)$, $x \in B$.*

Proof: In Theorem 70, it suffices to take $h(x,\nu)=\mathbf{1}_A(x)$, $A\subseteq B$. Then $\mathbb{E}\Phi(A)=\int_A \mathbb{E}\lambda^*(x,\Phi)\,\mathrm{d}x$. □

In many applications point processes with pairwise interactions appear.

Definition 80. Let

$$p(\nu)\propto \prod_{x\in\operatorname{supp}\nu} g(\delta_x) \prod_{\{x,y\}\subseteq\operatorname{supp}\nu} g(\delta_x+\delta_y),\quad \nu\in\mathcal{N}_{\mathrm{f}}^*, \tag{21}$$

where g is a non-negative measurable function such that the right-hand side in (21) is Π_Λ-integrable. A point process Φ with density p w.r.t. the distribution of a Poisson point process is then called the *pairwise interaction point process*. The function g is called the *interaction function*. We define the *range of interaction* as

$$R=\inf\{r>0: g(\delta_x+\delta_y)=1 \text{ for all } \|x-y\|>r\}.$$

The point process Φ is called *repulsive* if $g(\delta_x+\delta_y)\le 1$ for all $x\neq y$.

If $g(\delta_x)=\beta>0$ is constant and $g(\delta_x+\delta_y)=\theta(\|x-y\|)$ is invariant under translations and rotations, then we speak about the *homogeneous* pairwise interaction point process. Motivated by the statistical physics, we use the term *pair potential function* for $\theta:(0,\infty)\to\mathbb{R}^+$. It determines the strength of interaction between two points. The values smaller than 1 correspond to repulsive interactions, while the values larger than 1 mean attractive interactions.

The Papangelou conditional intensity of a pairwise interaction point process has the form

$$\lambda^*(x,\nu)=g(\delta_x)\prod_{y\in\operatorname{supp}\nu} g(\delta_x+\delta_y).$$

The function (21) is clearly hereditary and it is also locally stable provided that

$$\int_{\mathbb{R}^d} g(\delta_x)\,\Lambda(\mathrm{d}x)<\infty \tag{22}$$

and $g(\delta_x+\delta_y)\le 1$, because then $\lambda^*(x,\nu)\le g(\delta_x)$ and we can take $k(x)=g(\delta_x)$ in Remark 47. It means that under the assumption (22), the density of a repulsive pairwise interaction point process is always Π_Λ-integrable. The condition (22) holds automatically for homogeneous pairwise interaction point processes.

We have shown earlier that the Poisson point process is obtained in a particular case $g(\delta_x+\delta_y)=1$. The range of interaction is then $R=0$.

Different choices of θ lead to various models of homogeneous pairwise interaction point processes. A trivial choice $\theta(r)=1$ for every $r>0$ corresponds to the Poisson point process with intensity measure $\beta\Lambda$. The simplest and the most well-known non-trivial example of a homogeneous pairwise interaction point process is the Strauss process for which

$$\theta(r)=\gamma^{\mathbf{1}_{[r\le R]}},\quad 0\le\gamma\le 1,\ R>0,$$

where 0^0 is understood as 1.

Definition 81. Let $\beta>0$, $0\le\gamma\le 1$, and $R>0$ be real parameters. The *Strauss point process* Φ is given by the density

$$p(\nu)=\alpha\beta^{\nu(\mathbb{R}^d)}\gamma^{S_R(\nu)},\quad \nu\in\mathcal{N}_{\mathrm{f}}^*,$$

where α is the normalizing constant and $S_R(\nu)=\sum_{\{x,y\}\subseteq\operatorname{supp}\nu}\mathbf{1}_{[\|x-y\|\le R]}$ is the number of pairs of atoms of ν with distance at most R.

The Papangelou conditional intensity of the Strauss point process is $\lambda^*(x,\nu)=\beta\gamma^{t_R(x)}$, where $t_R(x)=\sum_{y\in\operatorname{supp}\nu}\mathbf{1}_{[\|y-x\|\le R]}$. The condition $\gamma\le 1$ entails $\lambda^*(x,\nu)\le\beta$ which in turn implies that p is locally stable. The normalizing constant $\alpha=\left(\int_{\mathcal{N}_{\mathrm{f}}^*}\beta^{\nu(\mathbb{R}^d)}\gamma^{S_R(\nu)}\,\Pi_\Lambda(\mathrm{d}\nu)\right)^{-1}$ is mostly unknown. It can be computed for the case $\gamma=1$ which corresponds to the Poisson point process with intensity measure $\beta\Lambda$. Then $\alpha=\mathrm{e}^{(1-\beta)\Lambda(\mathbb{R}^d)}$. The case $\gamma=0$ leads to the hard-core point process with hard-core distance R because $p(\nu)>0$ is only possible for $S_R(\nu)=0$ (remember that we put $0^0=1$). Then there does

not exist any pair of points in ν with distance smaller than R. The parameter R is also the range of interaction of the process. Figure 16 shows simulated realizations of the Strauss point process for extreme values $\gamma = 1$ (left) and $\gamma = 0$ (right) as well as for intermediate value $\gamma = 0.5$ (middle), where the pairs of points appear at a distance smaller than R but less often than in the Poisson point process ($\gamma = 1$).

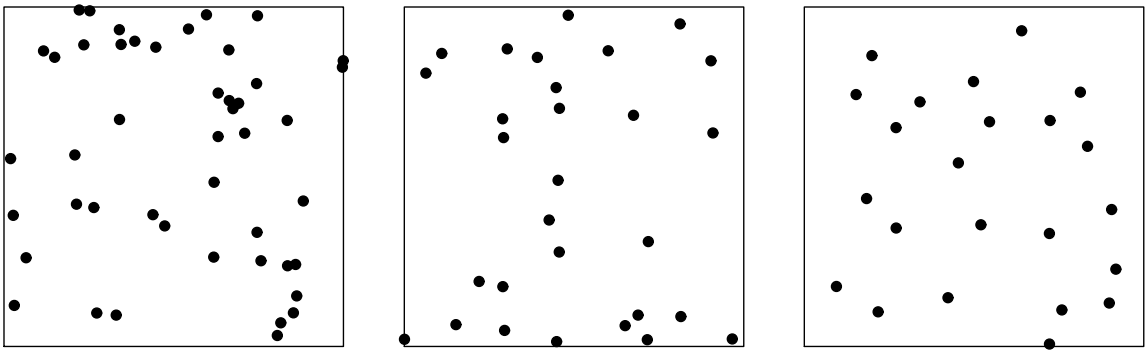

Figure 16. Three realizations of the Strauss point process in the unit square window $[0,1]^2$. The choice of parameters is $\beta = 50$, $R = 0.1$, and $\gamma = 1$ (left), $\gamma = 0.5$ (middle), and $\gamma = 0$ (right). The distribution is considered w.r.t. the standard Poisson point process on $[0,1]^2$.

Strauss proposed this process as a model of clustering [18]. This would correspond to $\gamma > 1$. However, it can be shown that $p(\nu)$ is not Π_Λ-integrable in this case. For $0 \le \gamma \le 1$, Π_Λ-integrability follows from local stability (Theorem 69). Therefore, the Strauss point process only models repulsion between points. It is an example of a repulsive pairwise interaction point process.

Definition 82. Let Π be the distribution of the standard Poisson point process on a bounded set $B \in \mathcal{B}_0^d$. Let $\beta > 0$, $\gamma > 0$, and $R > h > 0$ be real parameters. The *Strauss hard-core point process* is given by the density

$$p(\nu) = \begin{cases} 0, & \text{if there exist } x, y \text{ such that } \nu(\{x,y\}) = 2 \text{ and } 0 < \|x-y\| < h, \\ \alpha\beta^{\nu(B)}\gamma^{S_R(\nu)}, & \text{otherwise,} \end{cases}$$

w.r.t. Π. Hence, it is an example of a homogeneous pairwise interaction point process with pair potential function

$$\theta(r) = \begin{cases} 0 & \text{for } r \le h, \\ \gamma & \text{for } h < r \le R, \\ 1 & \text{for } r > R. \end{cases}$$

Remark 48. Since B is bounded and the pairs at distances smaller than $h > 0$ are excluded, there exists an upper bound for the number of points. This ensures Π-integrability of p also for $\gamma > 1$.

Let us mention an example where higher-order interactions occur.

Definition 83. The *Widom–Rowlinson point process* is given by the density

$$p(\nu) = \alpha\beta^{\nu(\mathbb{R}^d)}\gamma^{-|U_{\nu,r}|}, \quad \nu \in \mathcal{N}_{\mathrm{f}}^*,$$

where $U_{\nu,r} = \bigcup_{x \in \operatorname{supp}\nu} b(x,r)$ and $\beta > 0$, $r > 0$, and $\gamma > 0$. In $d = 2$ it is also known as the *area-interaction point process*.

The Papangelou condition intensity of the Widom–Rowlinson process has the form

$$\lambda^*(x,\nu) = \beta\gamma^{-|b(x,r)\setminus\cup_{y\in\operatorname{supp}\nu} b(y,r)|}.$$

It is easy to verify that λ^* is bounded:

$$\lambda^*(x,\nu) \le \begin{cases} \beta & \text{for } \gamma \ge 1, \\ \beta\gamma^{-\omega_d r^d} & \text{for } \gamma < 1, \end{cases}$$

which means that the density p is locally stable and thus Π-integrable by Theorem 69.

Pairwise interaction point processes, as well as the Widom–Rowlinson point process, are special cases of a more general class of point processes, so-called Markov point processes.

Markov point processes

Definition 84. Let $\sim$ be a reflexive and symmetric relation on $\mathbb{R}^d \times \mathbb{R}^d$. Two points $x, y \in \mathbb{R}^d$ satisfying $x \sim y$ are said to be *neighbours*. For $B \in \mathcal{B}_0^d$, we define the *neighbourhood* of B as $\partial B = \{x \in \mathbb{R}^d \setminus B : \exists y \in B, x \sim y\}$.

The most often used neighbourhood relation is defined by $x \sim y \Leftrightarrow \|x - y\| \le R$ for given $R > 0$ (so-called *R-neighbourhood*).

Definition 85. A measurable function $h : \mathcal{N}_{\mathrm{f}}^* \to \mathbb{R}^+$ satisfies *local Markov property* w.r.t. $\sim$ if for every $\nu \in \mathcal{N}_{\mathrm{f}}^*$ with $h(\nu) > 0$ and every x with $\nu(\{x\}) = 0$, we have

$$\frac{h(\nu + \delta_x)}{h(\nu)} = \frac{h(\nu + \delta_x - \sum_{i=1}^k \delta_{y_i})}{h(\nu - \sum_{i=1}^k \delta_{y_i})}$$

for any $y_1, \dots, y_k \in \operatorname{supp}\nu \setminus \partial\{x\}$, i.e. the y_i are atoms of ν that are not related to x ($y_i \not\sim x$). We say that h is a *Markov function* w.r.t. the relation $\sim$ if it is hereditary and satisfies local Markov property.

Example: The pairwise interaction function

$$h(\nu) = \alpha \prod_{x \in \operatorname{supp}\nu} g(\delta_x) \prod_{\{x,y\} \subseteq \operatorname{supp}\nu : x \sim y} g(\delta_x + \delta_y), \quad \nu \in \mathcal{N}_{\mathrm{f}}^*,$$

is Markov because

$$\frac{h(\nu + \delta_x)}{h(\nu)} = g(\delta_x) \prod_{y \in \operatorname{supp}\nu \cap \partial\{x\}} g(\delta_x + \delta_y)$$

depends only on x and on the neighbours of x in $\operatorname{supp}\nu$.

Definition 86. Let Φ be a finite point process with a density p. We say that Φ is a *Markov point process* if p is a Markov function.

Remark 49. The Papangelou conditional intensity $\lambda^*(x, \nu)$ of a Markov point process depends only on x and $\nu|_{\partial\{x\}}$.

Definition 87. A set $A \subseteq \mathbb{R}^d$ is a *clique* w.r.t. $\sim$ if $x \sim y$ for all $x, y \in A$. It means that all points of A are neighbours. The empty set and all singletons are cliques. We say that a non-negative measurable function g is an *interaction function* w.r.t. $\sim$, if $g(\nu) = 1$ whenever $\operatorname{supp}\nu$ is not a clique.

Theorem 72. *(Hammersley–Clifford–Ripley–Kelly theorem) A measurable function $h : \mathcal{N}_{\mathrm{f}}^* \to \mathbb{R}^+$ is a Markov function if and only if there is an interaction function $g : \mathcal{N}_{\mathrm{f}}^* \to \mathbb{R}^+$ such that*

$$h(\mu) = \prod_{\nu \preceq \mu} g(\nu), \quad \mu \in \mathcal{N}_{\mathrm{f}}^*. \tag{23}$$

Proof: One implication is obvious. The function h given by (23) is hereditary and satisfies local Markov property.

Assume that h is Markov and define g inductively: $g(\varnothing) = h(\varnothing)$; $g(\mu) = 1$ if $\operatorname{supp}\mu$ is not a clique; and

$$g(\mu) = \frac{h(\mu)}{\prod_{\nu \prec \mu} g(\nu)}$$

if $\operatorname{supp}\mu \neq \emptyset$ is a clique, where $\nu \prec \mu$ means both $\nu \preceq \mu$ and $\nu \neq \mu$. When $\prod_{\nu \prec \mu} g(\nu) = 0$, there exists $\nu \prec \mu$ such that $g(\nu) = 0$. Therefore, $h(\nu) = 0$, which in turn implies $h(\mu) = 0$ from hereditarity of h. In this case, we define $g(\mu) = 1$ (i.e. we put $0/0 = 1$). The function g is correctly defined and it is an interaction function. We distinguish three cases in order to show that it also fulfills (23).

1. $\operatorname{supp}\mu$ is a clique: $h(\varnothing) = g(\varnothing)$ and $h(\mu) = \prod_{\nu \preceq \mu} g(\nu)$ for $\mu \neq \varnothing$.

2. $h(\mu) = 0$ and $\operatorname{supp}\mu$ is not a clique: there exist $x, y \in \operatorname{supp}\mu$ such that $x \not\sim y$. If $\prod_{\nu \prec \mu} g(\nu) = 0$, then $\prod_{\nu \preceq \mu} g(\nu) = 0 = h(\mu)$. We will show that the complementary case is impossible. For contradiction, assume that $\prod_{\nu \prec \mu} g(\nu) > 0$. From local Markov property we know that

$$\frac{h(\mu)}{h(\mu - \delta_x)} = \frac{h(\mu - \delta_y)}{h(\mu - \delta_x - \delta_y)}.$$

Since $h(\mu)=0$, it could not happen that both $h(\mu-\delta_x)$ and $h(\mu-\delta_y)$ are positive. Therefore, we found ν which has one atom less than μ and at the same time $h(\nu)=0$. Furthermore, $\operatorname{supp}\nu$ could not be a clique because then $h(\nu)=0$ would imply $g(\nu)=0$. Now repeat the arguments and find $\tilde\nu \prec \nu$ such that $\operatorname{supp}\tilde\nu$ is not a clique and contains one point less than $\operatorname{supp}\nu$. We can continue with removing the atoms, and at some point, we have to end with a clique (all singletons are cliques). In this way, we get the contradiction with $\prod_{\nu\prec\mu} g(\nu)>0$.

3. $h(\mu)>0$: we will show (23) by induction on $n=\mu(\mathbb{R}^d)$. From the construction of g, we have $h(\varnothing)=g(\varnothing)$ and $h(\delta_z)=g(\varnothing)g(\delta_z)$. Hence, the relation (23) holds for $n=0$ and $n=1$. Assume that it is satisfied for $0,1,\dots,n-1$ and consider μ with $\mu(\mathbb{R}^d)=n\geq 2$ such that $\operatorname{supp}\mu$ is not a clique (for cliques we already know that (23) holds). There exist $x,y\in\operatorname{supp}\mu$ such that $x\not\sim y$. Denote $\nu=\mu-\delta_x-\delta_y$. Since h is hereditary, we get $h(\nu)>0$. Moreover, h satisfies local Markov property which yields

$$h(\mu)=\frac{h(\nu+\delta_y+\delta_x)}{h(\nu+\delta_x)}h(\nu+\delta_x)=\frac{h(\nu+\delta_y)}{h(\nu)}h(\nu+\delta_x).$$

Now we utilize the induction hypothesis,

$$h(\mu)=\frac{\prod_{\varphi\preceq\nu+\delta_y} g(\varphi)\prod_{\varphi\preceq\nu+\delta_x} g(\varphi)}{\prod_{\varphi\preceq\nu} g(\varphi)}=\prod_{\varphi\preceq\nu+\delta_x+\delta_y} g(\varphi).$$

We have used that $x\not\sim y$ implies $g(\varphi)=1$ for φ with $\varphi(\{x,y\})=2$.

□

According to Theorem 72, every Markov point process possesses a density in the form

$$p(\mu)=\prod_{\nu\preceq\mu} g(\nu)=g(\varnothing)\exp\left\{\sum_{\nu\preceq\mu,\nu\neq\varnothing}\log g(\nu)\right\}=\frac{\exp\{-U(\mu)\}}{\mathcal{Z}}.$$

In statistical physics, the term *Gibbs point process* is used and

$$U(\mu)=-\sum_{\nu\preceq\mu,\nu\neq\varnothing}\log g(\nu)$$

is interpreted as the *potential energy* of the configuration μ. The terms $\log g(\nu)$ are the *potentials* and the constant $\mathcal{Z}=1/g(\varnothing)$ is called the *partition function*. The Papangelou conditional intensity is then $\lambda^*(x,\mu)=\mathrm{e}^{-E(x,\mu)}$, where $E(x,\mu)=U(\mu+\delta_x)-U(\mu)$ is the energy needed for adding the point x to the configuration μ.

An example of a Markov point process w.r.t. the R-neighbourhood is the Strauss point process from Definition 81. Its Papangelou conditional intensity $\lambda^*(x,\nu)=\beta\gamma^{\nu(\partial\{x\})}$ depends only on x and $\nu|_{\partial\{x\}}$. The interaction function from Theorem 72 is

$$g(\nu)=\begin{cases}\alpha, & \nu(\mathbb{R}^d)=0,\\ \beta, & \nu(\mathbb{R}^d)=1,\\ \gamma, & \nu(\mathbb{R}^d)=2,\text{ where } 0<\|x-y\|\leq R \text{ for } x,y\in\operatorname{supp}\nu,\\ 1, & \text{otherwise.}\end{cases}$$

The Strauss point process contains only pairwise interactions.

A Markov point process with higher-order interactions is given in Definition 83. The Papangelou conditional intensity $\lambda^*(x,\nu)=\beta\gamma^{-|b(x,r)\setminus\bigcup_{y\in\operatorname{supp}\nu}b(y,r)|}$ depends on y satisfying $\|x-y\|\leq 2r$. Therefore, the Widom–Rowlinson point process satisfies the local Markov property w.r.t. the $(2r)$-neighbourhood. The interaction function g from Theorem 72 is obtained from the inclusion-exclusion formula,

$$g(\nu)=\begin{cases}\alpha, & \nu(\mathbb{R}^d)=0,\\ \beta\gamma^{-\omega_d r^d}, & \nu(\mathbb{R}^d)=1,\\ \gamma^{(-1)^k|b(x_1,r)\cap\dots\cap b(x_k,r)|}, & \nu=\sum_{i=1}^k\delta_{x_i},\ k\geq 2.\end{cases}$$

We see that $g(\nu)$ could be distinct from 1 also for configurations with more than two points.

4.5 Inhomogeneous point processes

We have already seen models and constructions that could lead to spatial point processes with non-constant intensity functions (then we speak about inhomogeneous point processes). The simplest example is the Poisson point process with a non-constant intensity function. Inhomogeneity could be easily introduced in finite point processes possessing a density. For example, we can consider

$$p(\nu) = \alpha\gamma^{S_R(\nu)} \prod_{x\in\operatorname{supp}\nu} \beta(x), \quad \nu \in \mathcal{N}_{\mathrm{f}}^*,$$

where $\gamma \in [0,1]$ and $R > 0$ are parameters and β is a non-negative bounded measurable function. This leads to the Strauss point process with non-constant first-order interactions. For $\gamma = 1$ it reduces to the inhomogeneous Poisson point process.

In this subsection, we consider some further examples of inhomogeneous spatial point processes.

Second-order intensity reweighted stationary point processes

We define a special class of spatial point processes where we allow non-constant intensity function, but the second-order characteristics are translation-invariant.

Definition 88. Let Φ be a simple point process with intensity function λ. Assume that

$$\mathcal{K}_{\mathrm{inhom}}(B) = \int_{\mathcal{N}} \int_{B+x} \frac{1}{\lambda(y)}\, \nu(\mathrm{d}y)\, P_x^!(\mathrm{d}\nu), \quad B \in \mathcal{B}^d,$$

does not depend on x for a.a. $x \in \mathbb{R}^d$. Then Φ is called a *second-order intensity reweighted stationary* point process. For this process, we can introduce the *inhomogeneous K-function*

$$K_{\mathrm{inhom}}(r) = \mathcal{K}_{\mathrm{inhom}}(b(o,r)), \quad r > 0.$$

Lemma 73. *Let Φ be a second-order intensity reweighted stationary point process. Then for arbitrary $A \in \mathcal{B}^d$ with positive and finite Lebesgue measure ($0 < |A| < \infty$), we have*

$$\mathcal{K}_{\mathrm{inhom}}(B) = \frac{1}{|A|}\mathbb{E} \sum_{X,Y\in\operatorname{supp}\Phi}^{\neq} \frac{\mathbf{1}_{[X\in A, Y-X\in B]}}{\lambda(X)\lambda(Y)}, \quad B \in \mathcal{B}^d. \tag{24}$$

Proof: We apply the Campbell–Mecke theorem (Theorem 50) for the function

$$h(x,\nu) = \frac{\mathbf{1}_A(x)}{\lambda(x)|A|} \int_{\mathbb{R}^d} \frac{1_{[y-x\in B]}}{\lambda(y)}\, \nu(\mathrm{d}y).$$

It yields that

$$\mathbb{E} \sum_{X\in\operatorname{supp}\Phi} h(X, \Phi - \delta_X) = \frac{1}{|A|}\mathbb{E} \sum_{X\in\operatorname{supp}\Phi} \frac{\mathbf{1}_A(X)}{\lambda(X)} \sum_{Y\in\operatorname{supp}\Phi, Y\neq X} \frac{\mathbf{1}_B(Y-X)}{\lambda(Y)}$$

is equal to

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathcal{N}} h(x,\nu)\, P_x^!(\mathrm{d}\nu)\, \Lambda(\mathrm{d}x) &= \frac{1}{|A|}\int_A \int_{\mathcal{N}} \int_{B+x} \frac{1}{\lambda(x)\lambda(y)} \nu(\mathrm{d}y)\, P_x^!(\mathrm{d}\nu)\, \lambda(x)\,\mathrm{d}x \\ &= \frac{1}{|A|}\int_A \mathcal{K}_{\mathrm{inhom}}(B)\,\mathrm{d}x = \mathcal{K}_{\mathrm{inhom}}(B).\end{aligned}$$

□

Remark 50. Note that the right-hand side in (24) is well-defined because $\lambda(X) > 0$ a.s. for all $X \in \operatorname{supp}\Phi$.

Lemma 74. *Every stationary point process Φ is second-order intensity reweighted stationary. Moreover, $\mathcal{K}_{\mathrm{inhom}}(B) = \mathcal{K}(B)$ for $B \in \mathcal{B}^d$, where $\mathcal{K}$ is the reduced second-order moment measure of Φ.*

Proof: The reduced Palm distributions for stationary point processes are $P_x^!(\cdot) = P_o^!(t_x^{-1}\cdot)$. Hence,

$$\begin{aligned}\int_{\mathcal{N}}\int_{B+x}\frac{1}{\lambda(y)}\,\nu(\mathrm{d}y)\,P_x^!(\mathrm{d}\nu) &= \int_{\mathcal{N}}\int_{B+x}\frac{1}{\lambda(y)}\,(t_x\nu)(\mathrm{d}y)\,P_o^!(\mathrm{d}\nu)\\ &= \int_{\mathcal{N}}\int_{B}\frac{1}{\lambda(y)}\,\nu(\mathrm{d}y)\,P_o^!(\mathrm{d}\nu)\\ &= \mathbb{E}_o^!\sum_{Y\in\operatorname{supp}\Phi\cap B}\frac{1}{\lambda(Y)} = \frac{1}{\lambda}\mathbb{E}_o^!\Phi(B) = \frac{1}{\lambda}\lambda\mathcal{K}(B) = \mathcal{K}(B).\end{aligned}$$

□

Lemma 75. *Every Poisson point process with intensity function λ is second-order intensity reweighted stationary and $\mathcal{K}_{\mathrm{inhom}}(B) = |B|$, $B \in \mathcal{B}^d$.*

Proof: Using Slivnyak's theorem (Theorem 51), we get

$$\begin{aligned}\mathcal{K}_{\mathrm{inhom}}(B) &= \int_{\mathcal{N}}\int_{B+x}\frac{1}{\lambda(y)}\,\nu(\mathrm{d}y)\,P_x^!(\mathrm{d}\nu)\\ &= \mathbb{E}\int_{B+x}\frac{1}{\lambda(y)}\,\Phi(\mathrm{d}y) = \mathbb{E}\sum_{Y\in\operatorname{supp}\Phi\cap(B+x)}\frac{1}{\lambda(Y)}\\ &= \int_{\mathbb{R}^d}\frac{\mathbf{1}_{B+x}(y)}{\lambda(y)}\lambda(y)\,\mathrm{d}y = |B+x| = |B|.\end{aligned}$$

□

Theorem 61 could be extended into the following statement.

Theorem 76. *Let Φ be a simple point process on $\mathbb{R}^d$ such that its pair correlation function g exists and is translation-invariant, i.e. $g(x,y) = g(y-x)$. Then Φ is second-order intensity reweighted stationary and*

$$\mathcal{K}_{\mathrm{inhom}}(B) = \int_B g(u)\,\mathrm{d}u, \quad B \in \mathcal{B}^d.$$

Proof: If we denote

$$\mathcal{K}_{\mathrm{inhom}}^x(B) = \int_{\mathcal{N}}\int_{B+x}\frac{1}{\lambda(y)}\,\nu(\mathrm{d}y)\,P_x^!(\mathrm{d}\nu), \quad B \in \mathcal{B}^d,$$

then for any $A \in \mathcal{B}^d$,

$$\begin{aligned}\int_A \mathcal{K}_{\mathrm{inhom}}^x(B)\,\mathrm{d}x &= \int_A\int_{\mathcal{N}}\int_{B+x}\frac{1}{\lambda(x)\lambda(y)}\,\nu(\mathrm{d}y)\,P_x^!(\mathrm{d}\nu)\,\Lambda(\mathrm{d}x)\\ &= \mathbb{E}\sum_{X\in\operatorname{supp}\Phi\cap A}\sum_{Y\in\operatorname{supp}\Phi, Y\neq X}\frac{\mathbf{1}_{B+X}(Y)}{\lambda(X)\lambda(Y)} = \mathbb{E}\sum_{X,Y\in\operatorname{supp}\Phi}^{\neq}\frac{\mathbf{1}_A(X)\mathbf{1}_{B+X}(Y)}{\lambda(X)\lambda(Y)}\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{\mathbf{1}_A(x)\mathbf{1}_B(y-x)}{\lambda(x)\lambda(y)}\,\alpha^{(2)}(\mathrm{d}x,\mathrm{d}y) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{1}_A(x)\mathbf{1}_B(y-x)g(x,y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{1}_A(x)\mathbf{1}_B(u)g(u)\,\mathrm{d}x\,\mathrm{d}u = |A|\int_B g(u)\,\mathrm{d}u.\end{aligned}$$

From this relation, we see that $\mathcal{K}_{\mathrm{inhom}}^x(B) = \int_B g(u)\,\mathrm{d}u$ for a.a. $x \in \mathbb{R}^d$.

□

Remark 51. Under the assumptions of Theorem 76, the reduced second-order moment measure $\mathcal{K}_{\mathrm{inhom}}$ is absolutely continuous w.r.t. the Lebesgue measure and its density is equal to the pair correlation function g. A randomly translated regular grid is an example of a point process with the reduced second-order moment measure that is not absolutely continuous (see Exercise class).

Lemma 74 and Lemma 75 are special consequences of Theorem 76. By Theorem 66 and Theorem 76 we can also notice that point processes obtained by independent thinning of stationary processes are second-order intensity reweighted stationary.

There are further possibilities how to obtain tractable models of inhomogeneous point patterns. They can be based on operations that produce an inhomogeneous point process from a homogeneous one. According to Theorem 65, by independent thinning (Definition 73) of a point process with constant intensity function λ, we obtain a point process with intensity function given by the product of λ and the retention probability function that could be non-constant. Here, we mention two operations (local scaling and transformation) that work on finite point processes defined in a bounded domain.

Local scaling

A class of inhomogeneous spatial point processes obtained by a location-dependent scaling has been introduced in [7].

Denote by $\mathcal{H} = (\mathcal{H}^0, \dots, \mathcal{H}^d)$ the vector of k-dimensional Hausdorff measures in $\mathbb{R}^d$. For $c > 0$ we consider the scaled k-dimensional Hausdorff measure ($k = 0, \dots, d$)

$$\mathcal{H}_c^k(A) = \mathcal{H}^k(c^{-1}A), \quad A \in \mathcal{B}^d.$$

Put $\mathcal{H}_c = (\mathcal{H}_c^0, \dots, \mathcal{H}_c^d)$.

Definition 89. Let Φ be a finite point process with density p w.r.t. the distribution Π of the standard Poisson point process on $B \in \mathcal{B}_0^d$. Assume that $p(\nu) \propto g(\nu, \mathcal{H})$, where g is scale-invariant, i.e. $g(c\nu, \mathcal{H}_c) = g(\nu, \mathcal{H})$ for all $\nu \in \mathcal{N}_{\mathrm{f}}^*$ and $c > 0$. Furthermore, we consider a scaling function $c : \mathbb{R}^d \to \mathbb{R}^+$ and a locally scaled version of $\mathcal{H}^k$, $k = 0, \dots, d$, given by

$$\mathcal{H}_c^k(A) = \int_A c(u)^{-k} \mathcal{H}^k(\mathrm{d}u), \quad A \in \mathcal{B}^d,$$

Again we write $\mathcal{H}_c = (\mathcal{H}_c^0, \dots, \mathcal{H}_c^d)$. The scaling function c is assumed to be bounded from below and from above, i.e. there exist positive constants c_1 and c_2 such that $0 < c_1 < c(u) < c_2$ for all $u \in \mathbb{R}^d$. We assume that $g(\nu, \mathcal{H}_c)$ is integrable w.r.t. the distribution Π_c of the Poisson point process with intensity measure $\mathcal{H}_c^d$ on B. A *locally scaled point process* Φ_c on B with scaling function c and template process Φ is then a finite point process defined by the density

$$p^{(c)}(\nu) \propto g(\nu, \mathcal{H}_c)$$

w.r.t. Π_c.

Remark 52. The density of Φ_c w.r.t. Π is

$$p_c(\nu) = \exp\left\{ - \int_B [c(u)^{-d} - 1] \, \mathrm{d}u \right\} \prod_{x \in \operatorname{supp} \nu} c(x)^{-d} p^{(c)}(\nu).$$

The classical examples of homogeneous point processes have scale-invariant density and thus can be used as a template process Φ. For example, for the Strauss point process we get

$$g(\nu, \mathcal{H}) = \beta^{\mathcal{H}^0(\operatorname{supp} \nu)} \gamma^{\sum_{\{x,y\} \subseteq \operatorname{supp} \nu} \mathbf{1}_{[0 < \mathcal{H}^1([x,y]) \le R]}},$$

where $[x, y]$ is the segment with endpoints x and y. The locally scaled Strauss point process Φ_c has density w.r.t. Π_c of the form

$$p^{(c)}(\nu) \propto \beta^{\nu(B)} \gamma^{S_R^{(c)}(\nu)},$$

where $S_R^{(c)}(\nu) = \sum_{\{x,y\} \subseteq \operatorname{supp} \nu} \mathbf{1}_{[0 < \mathcal{H}_c^1([x,y]) \le R]}$. An example of such process is shown in Figure 17.

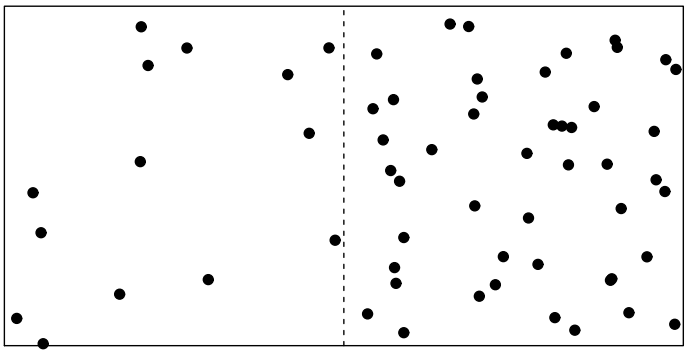

Figure 17. A realization of the locally scaled Strauss point process with parameters $\beta = 100$, $R = 0.1$, and $\gamma = 0.7$, in the window $[0, 2] \times [0, 1]$ composed of two unit squares. The scaling function c is constantly equal to 5 on the left square and constantly equal to 1 on the right square.

Transformation

Theorem 77. *Let Φ be a finite simple point process with density p w.r.t. the distribution of the standard Poisson point process on an open set $B \in \mathcal{B}_0^d$. Let $A \in \mathcal{B}_0^d$ be open and let $h : B \to A$ be a diffeomorphism. Let $J_{h^{-1}}$ be the Jacobian of h^{-1}. Then $h(\Phi) = \sum_{X \in \operatorname{supp} \Phi} \delta_{h(X)}$ is the finite point process with density p_h w.r.t. the distribution of the standard Poisson point process on A. The density p_h has the form*

$$p_h(\nu) = p(h^{-1}(\nu))\, \mathrm{e}^{|A|-|B|} \prod_{y \in \operatorname{supp} \nu} J_{h^{-1}}(y),$$

where $h^{-1}(\nu) = \sum_{y \in \operatorname{supp} \nu} \delta_{h^{-1}(y)}$.

Proof: Using the change of variable theorem ([11], Theorem 34.18) and (19), we obtain

$$\begin{aligned}
\mathbb{P}(h(\Phi) \in \mathcal{U}) &= \mathrm{e}^{-|B|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_B \cdots \int_B \mathbf{1}_{\mathcal{U}}(h(\sum_{i=1}^{n} \delta_{x_i})) p(\sum_{i=1}^{n} \delta_{x_i})\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \mathrm{e}^{-|B|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_A \cdots \int_A \mathbf{1}_{\mathcal{U}}(\sum_{i=1}^{n} \delta_{y_i}) p(\sum_{i=1}^{n} \delta_{h^{-1}(y_i)}) \prod_{i=1}^{n} J_{h^{-1}}(y_i)\, \mathrm{d}y_i \\
&= \mathrm{e}^{-|A|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_A \cdots \int_A \mathbf{1}_{\mathcal{U}}(\sum_{i=1}^{n} \delta_{y_i}) p_h(\sum_{i=1}^{n} \delta_{y_i})\, \mathrm{d}y_1 \cdots \mathrm{d}y_n,
\end{aligned}$$

where the terms for $n = 0$ are $\mathbf{1}_{\mathcal{U}}(\varnothing) p(\varnothing)$ and $\mathbf{1}_{\mathcal{U}}(\varnothing) p_h(\varnothing)$.

□

Corollary 78. *Let Φ be a Markov point process w.r.t. the relation $\sim$. Its density w.r.t. the distribution of the standard Poisson point process on $B \in \mathcal{B}_0^d$ can be expressed as*

$$p(\mu) = \prod_{\nu \preceq \mu} g(\nu),$$

where g is an interaction function. Let $h : B \to A$ be a diffeomorphism. Then $h(\Phi) = \sum_{X \in \operatorname{supp} \Phi} \delta_{h(X)}$ is the Markov point process w.r.t. the induced relation $\sim_h$ that is defined as

$$y_1 \sim_h y_2 \iff h^{-1}(y_1) \sim h^{-1}(y_2), \quad y_1, y_2 \in A.$$

Furthermore, the density of $h(\Phi)$ w.r.t. the distribution of the standard Poisson point process on $A \in \mathcal{B}_0^d$ has the form

$$p_h(\varphi) = \prod_{\psi \preceq \varphi} g_h(\psi),$$

where g_h is an interaction function given as

$$g_h(\nu) = \begin{cases} g(\varnothing)\mathrm{e}^{|A|-|B|}, & \nu = \varnothing, \\ g(h^{-1}(y))J_{h^{-1}}(y), & \nu = \delta_y, \\ g(h^{-1}(\nu)), & \nu(A) \geq 2. \end{cases}$$

Proof: The form of the density follows from Theorem 77. It remains to verify that g_h is an interaction function w.r.t. $\sim_h$. Let ν be such that $\operatorname{supp}\nu$ is not a clique w.r.t. $\sim_h$. Then ν must have at least two atoms and $g_h(\nu) = g(h^{-1}(\nu)) = 1$ because $\operatorname{supp} h^{-1}(\nu)$ is not a clique w.r.t. $\sim$.

□

Notice that the transformation may introduce inhomogeneity in first-order interactions. However, the higher-order interaction structure is not affected.

Example: Consider $A = B = (0,1)^d$ and assume that we want to describe the first-order inhomogeneity by the function $\eta(y)$, $y \in A$. If we can find a diffeomorphism h such that $\eta(y) = J_{h^{-1}}(y)$, then by Corollary 78, $h(\Phi)$ is the Markov point process with density

$$p_h(\mu) = \prod_{y \in \operatorname{supp}\mu} \eta(y) \prod_{\nu \preceq h^{-1}(\mu)} g(\nu).$$

The change of variable theorem implies that

$$\int_A J_{h^{-1}}(y)\,\mathrm{d}y = |B|.$$

Therefore, $\eta(y)$ has to satisfy $\int_A \eta(y)\,\mathrm{d}y = 1$. In particular, consider independent contributions to the inhomogeneity ($\eta(y) = \eta_1(y_1)\cdots\eta_d(y_d)$) having the exponential form

$$\eta_i(u) = \begin{cases} \frac{\theta_i \mathrm{e}^{\theta_i u}}{\mathrm{e}^{\theta_i}-1} & \text{for } \theta_i \neq 0, \\ 1 & \text{for } \theta_i = 0, \end{cases} \qquad u \in (0,1),\ i = 1,\ldots,d.$$

Then $h(x) = (h_1(x_1),\ldots,h_d(x_d))$, where

$$h_i(u) = \begin{cases} \frac{1}{\theta_i}\log(1 + (\mathrm{e}^{\theta_i}-1)u) & \text{for } \theta_i \neq 0, \\ 1 & \text{for } \theta_i = 0, \end{cases} \qquad u \in (0,1),\ i = 1,\ldots,d. \tag{25}$$

If Φ is the Strauss point process with density

$$p(\nu) = \alpha\beta^{\nu(B)} \prod_{\{x,y\} \subseteq \operatorname{supp}\nu} \gamma^{\mathbf{1}_{[0<\|x-y\|\leq R]}},$$

then the transformed point process $h(\Phi)$ has density

$$p_h(\nu) = \alpha\beta^{\nu(B)} \prod_{y \in \operatorname{supp}\nu} \eta(y) \prod_{\{x,y\} \subseteq \operatorname{supp}\nu} \gamma^{\mathbf{1}_{[0<\|h^{-1}(x)-h^{-1}(y)\|\leq R]}}.$$

Figure 18 shows the effect of transformation for different choices of h.

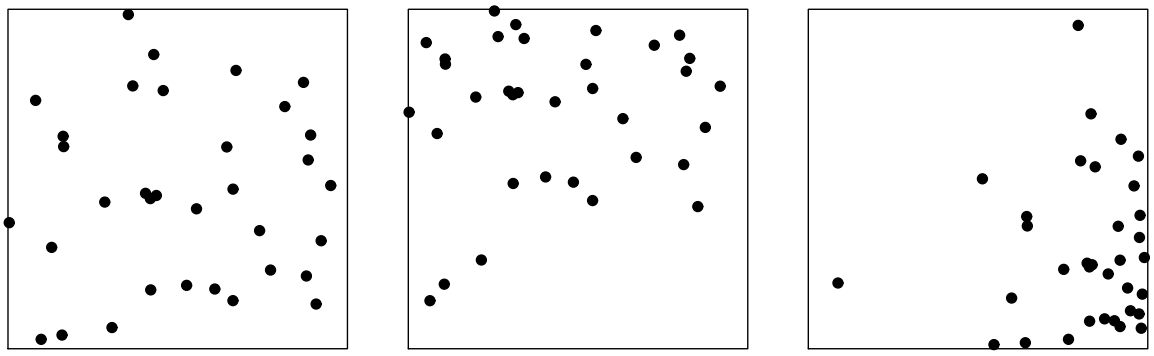

Figure 18. A realization of the Strauss point process (left) and two transformations of this realization given by (25) with $\theta_1 = -1$ and $\theta_2 = 3$ (middle), and $\theta_1 = 5$ and $\theta_2 = -2$ (right). The distribution of the original Strauss point process is w.r.t. the standard Poisson point process on $[0,1]^2$ and the parameters are $\beta = 50$, $R = 0.1$, and $\gamma = 0.5$.

The transformation h could be more generally considered between k-dimensional differentiable manifolds in $\mathbb{R}^d$ (see [20]).

5. Marked point processes

5.1 Basic definitions

A marked point process is obtained from a point process by assigning a certain value to each atom of the process. This additional information is referred to as the mark.

Definition 90. Consider a complete separable locally compact metric space $\mathbb{M}$ which will be called the *mark space*. Its Borel σ-algebra is denoted by $\mathcal{B}(\mathbb{M})$. Let $\mathcal{N}_{\mathrm{m}} = \{\nu \in \mathcal{N}(\mathbb{R}^d \times \mathbb{M}) : \nu(\cdot \times \mathbb{M}) \in \mathcal{N}(\mathbb{R}^d)\}$. By a *marked point process* we understand a point process Φ_{m} on $\mathbb{R}^d \times \mathbb{M}$ such that $\mathbb{P}(\Phi_{\mathrm{m}} \in \mathcal{N}_{\mathrm{m}}) = 1$. For each marked point process Φ_{m}, we can consider the corresponding *unmarked point process* or *ground process* Φ that is given by the projection on the first component of $\mathbb{R}^d \times \mathbb{M}$: $\Phi(B) = \Phi_{\mathrm{m}}(B \times \mathbb{M})$, $B \in B^d$.

It is good to realize that not every point process on $\mathbb{R}^d \times \mathbb{M}$ is a marked point process. We require that the projection on $\mathbb{R}^d$ is a point process. For example, take $\mathbb{M} = \mathbb{R}$ and consider a stationary Poisson point process Φ_{m} on $\mathbb{R}^{d+1}$ (the last component is the mark) with positive intensity. Then the number of points in $K \times \mathbb{M}$ ($K \in \mathcal{K}^d$, $|K| > 0$) is almost surely infinite. Hence, $\Phi_{\mathrm{m}}(\cdot \times \mathbb{M})$ is not a point process.

The unmarked point process Φ is obtained from Φ_{m} if we forget the marks and take into account only the locations of points. We will often add the assumption that Φ is a simple point process.

Definition 91. Let $\mathcal{N}^*_{\mathrm{m}} = \{\nu \in \mathcal{N}(\mathbb{R}^d \times \mathbb{M}) : \nu(\cdot \times \mathbb{M}) \in \mathcal{N}^*(\mathbb{R}^d)\}$. We say that a marked point process Φ_{m} is *simple* if $\mathbb{P}(\Phi_{\mathrm{m}} \in \mathcal{N}^*_{\mathrm{m}}) = 1$. It means that Φ_{m} is an example of a simple point process on $\mathbb{R}^d \times \mathbb{M}$. In particular, its distribution is completely determined by the void probabilities.

Remark 53. Every point process on $\mathbb{R}^d$ can be represented as a simple marked point process with the mark space $\mathbb{N}$. The marks correspond to the multiplicities of the atoms.

If we would like to speak about the mark corresponding to the point $X \in \operatorname{supp}\Phi$, we will denote it $M(X)$. It means that $(X, M(X)) \in \operatorname{supp}\Phi_{\mathrm{m}}$.

It follows from Lemma 29 that a marked point process Φ_{m} can be expressed by a finite or countable sum of Dirac measures and its atoms could be enumerated in a measurable way:

$$\Phi_{\mathrm{m}} = \sum_{i=1}^{\Phi(\mathbb{R}^d)} \delta_{(X_i, M_i)}. \tag{26}$$

If Φ_{m} is simple, then the X_i are mutually distinct.

The mark space could be quite complex. Later we focus mainly on discrete or categorical marks (Subsection 5.3) and real positive marks (Subsection 5.4). The case where $\mathbb{M}$ is the space of non-empty compact sets plays an important role in stochastic geometry (see [17]). Functional marks are studied, e.g. in the paper [4]. The simplest mark space is the finite space, its elements could be without loss of generality labelled by $1, \dots, k$.

Definition 92. A marked point process Φ_{m} with the mark space $\mathbb{M} = \{1, \dots, k\}$ is called the *multivariate point process*. It can be viewed as the k-tuples of point processes in $\mathbb{R}^d$: $\Phi_{\mathrm{m}} = (\Phi_1, \dots, \Phi_k)$, where $\Phi_i(\cdot) = \Phi_{\mathrm{m}}(\cdot \times \{i\})$, $i = 1, \dots, k$. When we want to stress the number of components, we speak about the *k-variate point process* (in particular, bivariate, trivariate, etc.).

An example of a multivariate point process ($k = 3$) is shown in Figure 19.

A stationary marked point process has the distribution invariant under the transformations that shift the points and keep the marks unchanged. Similarly, the distribution of an isotropic marked point process is invariant under rotations of the points around the origin in $\mathbb{R}^d$ and keeping the marks.

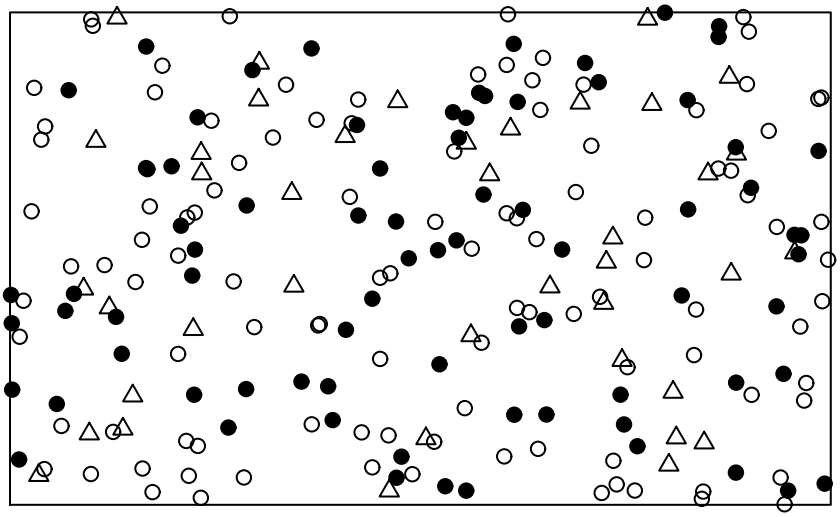

Figure 19. A realization of a stationary Poisson trivariate point process with intensities 3 (triangles), 5 (bullets) and 7 (circles).

Definition 93. We say that a marked point process Φ_{m} is *stationary* if the distribution of $t_z\Phi_{\mathrm{m}}$ is the same as the distribution of Φ_{m} for any $z \in \mathbb{R}^d$, where $(t_z\nu)(B \times L) = \nu((B-z) \times L)$, $B \in \mathcal{B}^d$, $L \in \mathcal{B}(\mathbb{M})$, $\nu \in \mathcal{N}_{\mathrm{m}}$. A marked point process Φ_{m} is *isotropic*, if $R_{\mathcal{O}}\Phi_{\mathrm{m}}$ and Φ_{m} have the same distribution for arbitrary rotation $\mathcal{O}$ around the origin, where $(R_{\mathcal{O}}\nu)(B \times L) = \nu(\mathcal{O}^{-1}B \times L)$. We say that Φ_{m} is *motion-invariant* if it is both stationary and isotropic.

Remark 54. If Φ_{m} is a stationary (or isotropic) marked point process, then also the corresponding ground process Φ is stationary (or isotropic). By the intensity of a stationary marked point process Φ_{m}, we understand the intensity of its ground point process Φ.

Recall that the intensity measure of a marked point process Φ_{m} is defined as $\Lambda_{\mathrm{m}}(B \times L) = \mathbb{E}\Phi_{\mathrm{m}}(B \times L)$, $B \in \mathcal{B}^d$, $L \in \mathcal{B}(\mathbb{M})$.

Theorem 79. *For a stationary marked point process Φ_{m} with finite and positive intensity λ, there exists a uniquely determined probability measure $\mathbb{Q}$ on $\mathbb{M}$ such that the intensity measure of Φ_{m} has the form*

$$\Lambda_{\mathrm{m}}(B \times L) = \lambda\,|B|\,\mathbb{Q}(L), \quad B \in \mathcal{B}^d,\ L \in \mathcal{B}(\mathbb{M}). \tag{27}$$

Proof: For every $L \in \mathcal{B}(\mathbb{M})$, we consider a locally finite measure $\mu_L(B) = \Lambda_{\mathrm{m}}(B \times L)$, $B \in \mathcal{B}^d$. From the assumption of stationarity, it follows that it is a translation-invariant measure on $\mathbb{R}^d$. Hence, it is a multiple of the Lebesgue measure: $\mu_L(B) = \lambda_L|B|$. If we put $\mathbb{Q}(L) = \frac{\lambda_L}{\lambda}$, then we easily see that $\mathbb{Q}$ is a probability measure that satisfies (27).

□

Definition 94. The probability measure $\mathbb{Q}$ from Theorem 79 is called the *stationary mark distribution*. A random element M_0 in the space $\mathbb{M}$ (a measurable mapping from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathbb{M}, \mathcal{B}(\mathbb{M}))$) with distribution $\mathbb{Q}$ is called the *typical mark*.

Corollary 80. *(First-order Campbell theorem for stationary marked point processes) Let Φ_{m} be a stationary simple marked point process and let h be a non-negative measurable function on $\mathbb{R}^d \times \mathbb{M}$. Then it follows that*

$$\mathbb{E} \sum_{(X,M) \in \operatorname{supp} \Phi_{\mathrm{m}}} h(X, M) = \lambda \int_{\mathbb{R}^d} \int_{\mathbb{M}} h(x, m)\, \mathbb{Q}(\mathrm{d}m)\, \mathrm{d}x.$$

Proof: The assertion follows from Theorem 43 and Theorem 79.

□

The second-order factorial moment measure of a simple marked point process Φ_m is denoted by $\alpha^{(2)}_{\mathrm{m}}$. For fixed mark sets $L_1, L_2 \in \mathcal{B}(\mathbb{M})$, it is obvious that the measure $\alpha^{(2)}_{\mathrm{m}}(\cdot \times L_1 \times \cdot \times L_2)$ is absolutely

continuous w.r.t. the second-order factorial moment measure $\alpha^{(2)}(\cdot\times\cdot)=\alpha^{(2)}_{\mathrm{m}}(\cdot\times\mathbb{M}\times\cdot\times\mathbb{M})$ of the unmarked point process Φ. If we assume that $\alpha^{(2)}$ is σ-finite measure, then (for fixed $L_1,L_2\in\mathcal{B}(\mathbb{M})$) there exists the Radon–Nikodym derivative $P_{x_1x_2}(L_1\times L_2)$. Similarly as in the definition of Palm distribution, we can consider the regular version, i.e. such that $P_{x_1x_2}(\cdot\times\cdot)$ is a probability measure for given $x_1,x_2\in\mathbb{R}^d$.

Theorem 81. *Let Φ_{m} be a simple marked point process such that its ground process Φ has σ-finite second-order factorial moment measure $\alpha^{(2)}$. Then there exists a Markov kernel $(x_1,x_2)\mapsto P_{x_1x_2}$ from $(\mathbb{R}^d\times\mathbb{R}^d,\mathcal{B}^d\times\mathcal{B}^d)$ to $(\mathbb{M}\times\mathbb{M},\mathcal{B}(\mathbb{M})\times\mathcal{B}(\mathbb{M}))$ such that*

$$\alpha^{(2)}_{\mathrm{m}}(B_1\times L_1\times B_2\times L_2)=\int_{B_1\times B_2}P_{x_1x_2}(L_1\times L_2)\,\alpha^{(2)}(\mathrm{d}(x_1,x_2)),$$

$B_1,B_2\in\mathcal{B}^d$, $L_1,L_2\in\mathcal{B}(\mathbb{M})$.

Proof: The assertion follows from Theorem 46. □

Definition 95. The distribution $P_{x_1x_2}$ from Theorem 81 is called the *two-point mark distribution* in points $x_1,x_2\in\mathbb{R}^d$, $x_1\neq x_2$.

Remark 55. The two-point mark distribution can be interpreted as the joint distribution of marks in x_1 and x_2 under the condition that x_1 and x_2 are atoms of the process. If Φ_{m} is motion-invariant, then $P_{x_1x_2}$ depends only on the distance $r=\|x_1-x_2\|$. In this case, we write P_{or} instead of $P_{x_1x_2}$. The expectation w.r.t. P_{or} will be denoted by $\mathbb{E}_{or}$.

The Palm distribution (or the reduced Palm distribution) of a simple marked point process Φ_{m} at (x,m) is denoted by $P_{(x,m)}$ (or $P^!_{(x,m)}$). For stationary Φ_{m}, we have $P_{(x,m)}(\cdot)=P_{(o,m)}(t_x^{-1}\cdot)$ (or $P^!_{(x,m)}(\cdot)=P^!_{(o,m)}(t_x^{-1}\cdot)$). Then the Campbell–Mecke theorem (Theorem 55) gives

$$\mathbb{E}\sum_{(X,M)\in\operatorname{supp}\Phi_{\mathrm{m}}}h(X,M,\Phi_{\mathrm{m}})=\lambda\int_{\mathbb{R}^d}\int_{\mathbb{M}}\int_{\mathcal{N}^*_{\mathrm{m}}}h(x,m,t_x\nu)\,P_{(o,m)}(\mathrm{d}\nu)\,\mathbb{Q}(\mathrm{d}m)\,\mathrm{d}x$$

and

$$\mathbb{E}\sum_{(X,M)\in\operatorname{supp}\Phi_{\mathrm{m}}}h(X,M,\Phi_{\mathrm{m}}-\delta_{(X,M)})=\lambda\int_{\mathbb{R}^d}\int_{\mathbb{M}}\int_{\mathcal{N}^*_{\mathrm{m}}}h(x,m,t_x\nu)\,P^!_{(o,m)}(\mathrm{d}\nu)\,\mathbb{Q}(\mathrm{d}m)\,\mathrm{d}x \tag{28}$$

for any non-negative measurable function h on $\mathbb{R}^d\times\mathbb{M}\times\mathcal{N}^*_{\mathrm{m}}$.

Similarly, we can define the Palm distribution w.r.t. any fixed mark set $L\in\mathcal{B}(\mathbb{M})$. Such distribution is obtained by the desintegration of the measure

$$C_L(B\times\mathcal{U})=\mathbb{E}\Phi_{\mathrm{m}}(B\times L)\mathbf{1}_{\mathcal{U}}(\Phi_{\mathrm{m}}),\quad B\in\mathcal{B}^d,\ L\in\mathcal{B}(\mathbb{M}),\ \mathcal{U}\in\mathfrak{N}_{\mathrm{m}},$$

w.r.t. to its projection $\Lambda_L(B)=C_L(B\times\mathcal{N}_{\mathrm{m}})$, which is equal to $\lambda|B|\mathbb{Q}(L)$ in stationary case.

Definition 96. Let Φ_{m} be a stationary simple marked point process with intensity $0<\lambda<\infty$ and stationary mark distribution $\mathbb{Q}$. Suppose that $L\in\mathcal{B}(\mathbb{M})$ satisfies $\mathbb{Q}(L)>0$. The *Palm distribution of Φ_{m} w.r.t. the mark set L* is a probability measure P_o^L on $\mathcal{N}^*_{\mathrm{m}}$ satisfying

$$\mathbb{E}\sum_{(X,M)\in\operatorname{supp}\Phi_{\mathrm{m}}}h(X,\Phi_{\mathrm{m}})\mathbf{1}_L(M)=\lambda\mathbb{Q}(L)\int_{\mathbb{R}^d}\int_{\mathcal{N}^*_{\mathrm{m}}}h(x,t_x\nu)\,P_o^L(\mathrm{d}\nu)\,\mathrm{d}x$$

for arbitrary non-negative measurable h on $\mathbb{R}^d\times\mathcal{N}^*_{\mathrm{m}}$. The *reduced Palm distribution of Φ_{m} w.r.t. the mark set L* is a probability measure $P_o^{!L}$ on $\mathcal{N}^*_{\mathrm{m}}$ satisfying

$$\mathbb{E}\sum_{(X,M)\in\operatorname{supp}\Phi_{\mathrm{m}}}h(X,\Phi_{\mathrm{m}}-\delta_{(X,M)})\mathbf{1}_L(M)=\lambda\mathbb{Q}(L)\int_{\mathbb{R}^d}\int_{\mathcal{N}^*_{\mathrm{m}}}h(x,t_x\nu)\,P_o^{!L}(\mathrm{d}\nu)\,\mathrm{d}x$$

for arbitrary non-negative measurable h on $\mathbb{R}^d\times\mathcal{N}^*_{\mathrm{m}}$.

One can define these Palm distributions as

$$P_o^L(\mathcal{U}) = \int_L P_{(o,m)}(\mathcal{U})\,\frac{\mathbb{Q}(\mathrm{d}m)}{\mathbb{Q}(L)} \qquad \text{and} \qquad P_o^{!L}(\mathcal{U}) = \int_L P_{(o,m)}^!(\mathcal{U})\,\frac{\mathbb{Q}(\mathrm{d}m)}{\mathbb{Q}(L)}.$$

We can interpret P_o^L as the conditional distribution of Φ_{m} given that o is an atom of Φ and its mark lies in L.

For $L = \mathbb{M}$ we will write $P_o = P_o^{\mathbb{M}}$ and $P_o^! = P_o^{!\mathbb{M}}$. In particular, we have

$$P_o^!(\mathcal{U}) = \int_{\mathbb{M}} P_{(o,m)}^!(\mathcal{U})\,\mathbb{Q}(\mathrm{d}m). \tag{29}$$

The expectation w.r.t. $P_{(o,m)}^!$ will be denoted by $\mathbb{E}_{(o,m)}^!$. Similar notation is used for expectations w.r.t. other (reduced) Palm distributions. For example, $\mathbb{E}_o^!$ is the expectation w.r.t. $P_o^!$.

5.2 Marking models

Definition 97. A Poisson point process Φ_{m} on $\mathbb{R}^d \times \mathbb{M}$ with intensity measure Λ_{m} such that $\Lambda_{\mathrm{m}}(K \times \mathbb{M}) < \infty$ for every $K \in \mathcal{K}^d$, is called a *marked Poisson point process* with intensity measure Λ_{m}.

Remark 56. The corresponding unmarked point process Φ is a Poisson point process on $\mathbb{R}^d$ with locally finite intensity measure $\Lambda(\cdot) = \Lambda_{\mathrm{m}}(\cdot \times \mathbb{M})$.

Remark 57. If $\mathbb{M}$ is a finite set, say $\{1, \dots, k\}$, we speak about a multivariate Poisson point process $\Phi_{\mathrm{m}} = (\Phi_1, \dots, \Phi_k)$. In this case, it follows directly from the definition that $\Phi_1, \dots, \Phi_k$ are independent Poisson point processes on $\mathbb{R}^d$ (see Lemma 82 below).

Definition 98. A multivariate point process $\Phi_{\mathrm{m}} = (\Phi_1, \dots, \Phi_k)$ fulfills the *random superposition model* if the subprocesses $\Phi_1, \dots, \Phi_k$ are independent.

Lemma 82. *A multivariate Poisson point process $\Phi_{\mathrm{m}} = (\Phi_1, \dots, \Phi_k)$ fulfills the random superposition model. Moreover, Φ_i is the Poisson point process with intensity measure $\Lambda_i(\cdot) = \Lambda_{\mathrm{m}}(\cdot \times \{i\})$, $i = 1, \dots, k$.*

Proof: Consider the sets $\mathcal{N}^*_{K_1,r_1}, \dots, \mathcal{N}^*_{K_k,r_k}$, where $K_1, \dots, K_k \in \mathcal{K}^d$ and $r_1, \dots, r_k \in \mathbb{N}_0$. Then

$$\begin{aligned}\mathbb{P}(\Phi_1 \in \mathcal{N}^*_{K_1,r_1}, \dots, \Phi_k \in \mathcal{N}^*_{K_k,r_k}) &= \mathbb{P}(\Phi_{\mathrm{m}}(K_1 \times \{1\}) = r_1, \dots, \Phi_{\mathrm{m}}(K_k \times \{k\}) = r_k) \\ &= \prod_{i=1}^k \mathbb{P}(\Phi_i(K_i) = r_i) = \prod_{i=1}^k \mathbb{P}(\Phi_i \in \mathcal{N}^*_{K_i,r_i}).\end{aligned}$$

We also have

$$\mathbb{P}(\Phi_i \in \mathcal{N}^*_{K_i,r_i}) = \mathbb{P}(\Phi_{\mathrm{m}}(K_i \times \{i\}) = r_i) = \mathrm{e}^{-\Lambda_{\mathrm{m}}(K_i \times \{i\})}\frac{\Lambda_{\mathrm{m}}(K_i \times \{i\})^{r_i}}{r_i!} = \mathrm{e}^{-\Lambda_i(K_i)}\frac{\Lambda_i(K_i)^{r_i}}{r_i!}.$$

□

Therefore, we can construct a multivariate Poisson point process by a superposition of independent Poisson point processes. Another way how to obtain a multivariate Poisson point process comes from so-called *independent marking*. This construction can be used for any mark space. At this moment, it will be helpful to represent Φ_{m} by means of (26).

Definition 99. A marked point process $\Phi_{\mathrm{m}} = \sum_i \delta_{(X_i,M_i)}$ is called *independently marked* if the random marks $\{M_i\}$ are i.i.d. and independent of the unmarked point process $\Phi = \sum_i \delta_{X_i}$. The common distribution $\mathbb{Q}$ of the marks $\{M_i\}$ is called the *mark distribution* of Φ_{m}.

Theorem 83. *Let Φ_{m} be an independently marked point process with mark distribution $\mathbb{Q}$. Then its intensity measure is*

$$\Lambda_{\mathrm{m}}(B \times L) = \Lambda(B)\mathbb{Q}(L), \quad B \in \mathcal{B}^d,\ L \in \mathcal{B}(\mathbb{M}), \tag{30}$$

where Λ is the intensity measure of the unmarked point process Φ. If Φ_{m} is stationary, the mark distribution coincides with the stationary mark distribution from Definition 94.

Proof: Denote $\tau = \Phi_{\mathrm{m}}(\mathbb{R}^d \times \mathbb{M}) = \Phi(\mathbb{R}^d)$. For $B \in \mathcal{B}^d$ and $L \in \mathcal{B}(\mathbb{M})$, we obtain

$$\begin{aligned}\Lambda_{\mathrm{m}}(B \times L) &= \mathbb{E}\sum_{i=1}^{\tau} \delta_{(X_i,M_i)}(B \times L) = \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \mathbb{E}\left[\mathbf{1}_{[\tau=k]} \sum_{i=1}^{k} \delta_{(X_i,M_i)}(B \times L)\right] \\ &= \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \sum_{i=1}^{k} \mathbb{E}\mathbf{1}_{[\tau=k]}\mathbf{1}_{[X_i\in B, M_i\in L]} = \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \sum_{i=1}^{k} \mathbb{E}\mathbf{1}_{[\tau=k]}\mathbf{1}_{[X_i\in B]}\mathbb{E}\mathbf{1}_{[M_i\in L]} \\ &= \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \mathbb{E}\left[\mathbf{1}_{[\tau=k]} \sum_{i=1}^{k} \delta_{X_i}(B)\right] \mathbb{Q}(L) = \mathbb{E}\Phi(B) \cdot \mathbb{Q}(L) = \Lambda(B)\mathbb{Q}(L).\end{aligned}$$

For a stationary marked point process, we have $\Lambda(B) = \lambda|B|$. From (27) it follows that the stationary mark distribution is $\mathbb{Q}$.

□

We have already used the principle of independent marking several times. It appears in the definition of independent thinning (Definition 73) and Matérn hard-core point process of type II (Definition 74). In the proof of Theorem 68, we actually utilized the fact that the marked Poisson point process is obtained by independent marking of the Poisson point process.

Theorem 84. *Consider an independently marked point process Φ_{m} such that the corresponding unmarked point process Φ is a Poisson point process in $\mathbb{R}^d$ with diffuse intensity measure Λ. Then Φ_{m} is a simple marked Poisson point process.*

Proof: The void probabilities of a marked Poisson point process with intensity measure Λ_{m} are

$$\exp\{-\Lambda_{\mathrm{m}}(A)\}, \quad A \in \mathcal{K}(\mathbb{R}^d \times \mathbb{M}).$$

Lemma 38 ensures that the marked point process Φ_{m} is simple. We calculate its void probabilities. If we again denote $\tau = \Phi_{\mathrm{m}}(\mathbb{R}^d \times \mathbb{M}) = \Phi(\mathbb{R}^d)$, then

$$\begin{aligned}\mathbb{P}(\Phi_{\mathrm{m}}(A) = 0) &= \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \mathbb{P}(\tau = k, \cap_{i=1}^{k}[(X_i, M_i) \notin A]) \\ &= \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \mathbb{E}\mathbf{1}_{[\tau=k]} \prod_{i=1}^{k} \mathbf{1}_{[(X_i,M_i)\notin A]} = \sum_{k\in\mathbb{N}_0\cup\{\infty\}} \mathbb{E}\mathbf{1}_{[\tau=k]} \prod_{i=1}^{k} \int_{\mathbb{M}} \mathbf{1}_{[(X_i,m)\notin A]}\, \mathbb{Q}(\mathrm{d}m) \\ &= \mathbb{E}\prod_{i=1}^{\tau} \int_{\mathbb{M}} \mathbf{1}_{[(X_i,m)\notin A]}\, \mathbb{Q}(\mathrm{d}m) = \mathbb{E}\prod_{i=1}^{\tau} \left(1 - \int_{\mathbb{M}} \mathbf{1}_{[(X_i,m)\in A]}\, \mathbb{Q}(\mathrm{d}m)\right)\end{aligned}$$

for $A \in \mathcal{K}(\mathbb{R}^d \times \mathbb{M})$. Now it suffices to exploit the identity

$$\mathbb{E}\prod_{X\in\operatorname{supp}\Phi} f(X) = \exp\left\{-\int_{\mathbb{R}^d} (1 - f(x))\, \Lambda(\mathrm{d}x)\right\}$$

for an arbitrary measurable function $f : \mathbb{R}^d \to [0, 1]$ and a Poisson point process Φ on $\mathbb{R}^d$ with intensity measure Λ (see Lemma 45). We obtain

$$\mathbb{P}(\Phi_{\mathrm{m}}(A) = 0) = \exp\left\{-\int_{\mathbb{R}^d}\int_{\mathbb{M}} \mathbf{1}_{[(x,m)\in A]}\, \mathbb{Q}(\mathrm{d}m)\, \Lambda(\mathrm{d}x)\right\},$$

which by Theorem 83 equals to $\exp\{-\Lambda_{\mathrm{m}}(A)\}$.

Now from Theorem 32, it follows that Φ_{m} has the same distribution as a marked Poisson point process with intensity measure Λ_{m}.

□

The previous theorem tells us that by applying independent marking to the Poisson point process, we obtain the marked Poisson point process. However, not every marked Poisson point process can be

obtained by this procedure. According to Theorem 83, the resulting independently marked process has an intensity measure of the product form. Therefore, the marked Poisson point process with a general intensity measure could not be obtained by independent marking. In this situation, the corresponding unmarked point process is still Poisson but the marks could depend on the locations. Nevertheless, this does not happen in the stationary case.

Theorem 85. *A stationary marked Poisson point process is independently marked.*

Proof: We may assume that the intensity measure Λ_{m} is non-trivial (for $\Lambda_{\mathrm{m}} = \varnothing$ there is nothing to prove). Then the representation (26) becomes $\Phi_{\mathrm{m}} = \sum_{i=1}^{\infty} \delta_{(X_i,M_i)}$, because stationarity implies $\tau = \Phi_{\mathrm{m}}(\mathbb{R}^d \times \mathbb{M}) = \infty$ a.s. By Theorem 79, we have $\Lambda_{\mathrm{m}}(B \times L) = \lambda|B|\mathbb{Q}(L)$. Consider a sequence $\{\tilde{M}_i\}$ of i.i.d. random elements with values in $\mathbb{M}$ and distribution $\mathbb{Q}$, independent of the unmarked point process $\Phi = \sum_{i=1}^{\infty} \delta_{X_i}$. Then $\tilde{\Phi}_{\mathrm{m}} = \sum_{i=1}^{\infty} \delta_{(X_i,\tilde{M}_i)}$ is an independently marked point process with intensity measure which is by Theorem 83 equal to $\Lambda_{\mathrm{m}}(B \times L) = \lambda|B|\mathbb{Q}(L)$. Therefore, both marked Poisson point processes Φ_{m} and $\tilde{\Phi}_{\mathrm{m}}$ have the same distribution.

□

Remark 58. The assumption of stationarity was used only to get the appropriate form of intensity measure. In the same way, we get that every marked Poisson point process with intensity measure in the product form $\Lambda_{\mathrm{m}}(B \times L) = \Lambda(B)\mathbb{Q}(L)$ is independently marked.

Independent marking gives the simplest model for marked point processes: the marks are mutually independent and also independent of the locations. For multivariate point processes, the random superposition provides another model that exploits the independence. In general, independent marking and random superposition are different models. However, they coincide for a stationary multivariate Poisson point process. Independent marking follows from Theorem 85, while random superposition is the consequence of Lemma 82.

One possible approach that allows dependent marks is so-called geostatistical marking.

Definition 100. Let Φ be a simple point process on $\mathbb{R}^d$ and let $\{M(x) : x \in \mathbb{R}^d\}$ be a random field, independent of Φ, taking values in the mark space $\mathbb{M}$. Assume that the mapping $(x, \omega) \mapsto M(x, \omega)$ is measurable. Then $\Phi_{\mathrm{m}} = \sum_{X \in \operatorname{supp}\Phi} \delta_{(X,M(X))}$ is called *geostatistically marked point process* or *externally marked point process*.

Spatial correlations present in the random field $\{M(x) : x \in \mathbb{R}^d\}$ cause that the marks of Φ_{m} are correlated. Even if geostatistical marking is suitable in many applications, the assumption of independence of marks on the locations may be restrictive in other applications.

As an example of a situation where the marks depend on the locations, we can mention so-called *constructed marks*. These are generated by a certain mechanism from given unmarked points. The construction usually reflects the geometric arrangement of the points. A simple example may be the distance to the nearest neighbour $M(X) = d(X, \operatorname{supp}(\Phi - \delta_X))$, or the number of further points in the distance smaller than $r > 0$, $M(X) = \Phi(b(X, r)) - 1$.

5.3 Multivariate point processes

In this subsection, we will deal with marked point processes with qualitative marks. We assume that the mark space is $\mathbb{M} = \{1, \ldots, k\}$. A multivariate point process was introduced in Definition 92. We restrict to the stationary case and define basic numerical and functional summary characteristics that capture spatial aspects of marks.

Thus, we consider a stationary simple multivariate point process $\Phi_{\mathrm{m}} = (\Phi_1, \ldots, \Phi_k)$. The point processes Φ_i are stationary as well and we denote their intensities by λ_i, $i = 1, \ldots, k$. The unmarked point process is simple and thus can be written as $\Phi = \sum_{i=1}^{k} \Phi_i$. It has the intensity $\lambda = \sum_{i=1}^{k} \lambda_i$. The stationary mark distribution $\mathbb{Q}$ is an atomic measure on $\mathbb{M}$. We denote by $p_i = \mathbb{Q}(\{i\})$ the probability of the mark i. By (27) we have

$$\mathbb{E}\Phi_i(B) = \Lambda_{\mathrm{m}}(B \times \{i\}) = \lambda p_i |B|.$$

Therefore, the intensity of Φ_i satisfies $\lambda_i = \lambda p_i$. Hence, p_i is given by the ratio of the intensity λ_i and the overall intensity λ.

First we introduce two numerical characteristics.

Definition 101. Let D_j be the distance from the origin to the nearest point of Φ_j. The *bivariate aggregation index* or also the *bivariate Clark–Evans index* is defined as

$$\mathrm{CE}_{ij} = \frac{d(\lambda_j \omega_d)^{1/d}}{\Gamma(1/d)} \mathbb{E}^!_{(o,i)} D_j, \quad i,j = 1,\dots,k.$$

Remark 59. The index reports the expected distance from the point with mark i to the nearest neighbour with mark j in the case of a given process normalized by the same quantity for the case of a multivariate Poisson point process with identical intensities. It means that $\mathrm{CE}_{ij} = 1$ for a multivariate Poisson point process. Values $\mathrm{CE}_{ij} > 1$ indicate the repulsion among the points with marks i and j. On the other hand, values $\mathrm{CE}_{ij} < 1$ suggest the attraction of these points. For $i = j$, we get the aggregation index of the point process Φ_i (see Definition 65).

Definition 102. Let Z_i be the i-th nearest point of Φ from the origin. The *mingling index* takes into account p nearest neighbours of the typical point and gives the expected ratio of those of different type. It is defined as

$$\bar{M}_p = \frac{1}{p} \mathbb{E}^!_o \sum_{i=1}^{p} \mathbf{1}_{[M(o) \neq M(Z_i)]},$$

where we use the convention that $M(o)$ denotes the mark of the origin and $M(Z_i)$ denotes the mark of Z_i.

Remark 60. The mingling index takes larger values if the typical point is surrounded rather by points with different marks. On the contrary, it takes smaller values when the points with distinct marks tend to separate from each other.

Now we define basic functional characteristics for stationary simple multivariate point processes.

Definition 103. The *cross nearest-neighbour distribution function* or also the *cross G-function* is given by

$$G_{ij}(r) = P^!_{(o,i)}(\{\nu \in \mathcal{N}^*_{\mathrm{m}} : \nu(b(o,r) \times \{j\}) > 0\}), \quad r \geq 0.$$

The *condensed G-function* is defined by

$$G_{i\cdot}(r) = P^!_{(o,i)}(\{\nu \in \mathcal{N}^*_{\mathrm{m}} : \nu(b(o,r) \times \mathbb{M}) > 0\}), \quad r \geq 0.$$

Remark 61. The cross G-function is the distribution function of the distance from the typical point with mark i to the nearest neighbour with mark j. The condensed G-function is the distribution function of the distance from the typical point with mark i to the nearest neighbour with arbitrary mark.

Definition 104. We define the *cross J-function* by the relation

$$J_{ij}(r) = \frac{1 - G_{ij}(r)}{1 - F_j(r)}, \quad r \geq 0 : F_j(r) < 1,$$

where $F_j(r)$ is the spherical contact distribution function of Φ_j (see Definition 66). Next we define the *condensed J-function* by the relation

$$J_{i\cdot}(r) = \frac{1 - G_{i\cdot}(r)}{1 - F(r)}, \quad r \geq 0 : F(r) < 1,$$

where $F(r)$ is the spherical contact distribution function of Φ.

Lemma 86. *The J-function $J(r)$ of the point process Φ satisfies*

$$J(r) = \sum_{i=1}^{k} p_i J_{i\cdot}(r).$$

Proof: The relation (29) for stationary multivariate point processes becomes

$$P^!_o(\mathcal{U}) = \sum_{i=1}^{k} p_i P^!_{(o,i)}(\mathcal{U}).$$

Taking $\mathcal{U} = \{\nu \in \mathcal{N}_{\mathrm{m}}^* : \nu(b(o,r) \times \mathbb{M}) = 0\}$, we get

$$1 - G(r) = \sum_{i=1}^{k} p_i(1 - G_{i\cdot}(r)),$$

and now it suffices to divide both sides by $1 - F(r)$.

□

For a stationary multivariate Poisson point process, we have $G_{ij}(r) = 1 - \mathrm{e}^{-\lambda_j \omega_d r^d}$ and $G_{i\cdot}(r) = 1 - \mathrm{e}^{-\lambda \omega_d r^d}$. Consequently, $J_{ij}(r) = 1$ and $J_{i\cdot}(r) = 1$. The cross J-function is identical to 1 for any random superposition model. In that case, Φ_i and Φ_j are independent which leads to $G_{ij}(r) = F_j(r)$ and $J_{ij}(r) = 1$. Note that $J_{ij}(r)$ is different from $J_{ji}(r)$. Values $J_{ij}(r) > 1$ indicate the inhibition of points with mark j by points with mark i. On the other hand, values $J_{ij}(r) < 1$ correspond to the positive association.

Definition 105. Let $J(r)$ be the J-function of the point process Φ. We define the *I-function* by

$$I(r) = \sum_{i=1}^{k} p_i J_{ii}(r) - J(r), \quad r \geq 0 : F(r) < 1.$$

Remark 62. Using Lemma 86, we can also write $I(r) = \sum_{i=1}^{k} p_i(J_{ii}(r) - J_{i\cdot}(r))$. Under the assumption of random superposition, $J_{i\cdot}(r) = J_{ii}(r)$ and the I-function is identically equal to zero. The deviations from zero function may indicate either attractive or repulsive interactions among the points with distinct marks.

Definition 106. The *cross reduced second-order moment measure* $\mathcal{K}_{ij}$ is defined by the relation

$$\lambda_j \mathcal{K}_{ij}(B) = \mathbb{E}^!_{(o,i)} \Phi_j(B), \quad B \in \mathcal{B}^d,$$

where λ_j is the intensity of Φ_j. Further, we define the *cross K-function* as

$$K_{ij}(r) = \mathcal{K}_{ij}(b(o,r)), \quad r \geq 0.$$

Similarly, we define the *condensed reduced second-order moment measure* by

$$\lambda \mathcal{K}_{i\cdot}(B) = \mathbb{E}^!_{(o,i)} \Phi(B), \quad B \in \mathcal{B}^d$$

and the *condensed K-function* by

$$K_{i\cdot}(r) = \mathcal{K}_{i\cdot}(b(o,r)), \quad r \geq 0.$$

Remark 63. It means that $\lambda_j K_{ij}(r)$ represents the mean number of points with mark j in the ball of radius r and centre in the typical point with mark i. For $i = j$ we get the K-function of the subprocess Φ_i. Similarly, $\lambda K_{i\cdot}(r)$ denotes the mean number of points (with arbitrary mark) in the ball of radius r and centre in the typical point with mark i, the centre is not counted. The condensed K-function can be obtained from the cross K-functions using the relation

$$\lambda K_{i\cdot}(r) = \sum_{j=1}^{k} \lambda_j K_{ij}(r).$$

If g is the pair correlation function of a multivariate point process Φ_m, we write shortly $g_{ij}(x,y) = g((x,i),(y,j))$, where $x, y \in \mathbb{R}^d$ and $i, j \in \mathbb{M}$.

Theorem 87. *Let Φ_{m} be a stationary simple multivariate point process and assume that its second-order product density $\lambda_{\mathrm{m}}^{(2)}$ exists. The cross reduced second-order moment measure is absolutely continuous w.r.t. the Lebesgue measure and it can be expressed as*

$$\mathcal{K}_{ij}(B) = \int_B g_{ij}(u)\,\mathrm{d}u, \quad B \in B^d.$$

Furthermore,

$$\mathcal{K}_{ij}(B) = \mathcal{K}_{ji}(-B), \quad B \in \mathcal{B}^d.$$

In particular, the cross K-function is symmetric,

$$K_{ij}(r) = K_{ji}(r), \quad r \geq 0.$$

Proof: In analogy with the proof of Theorem 61, we obtain

$$\mathcal{K}_{ij}(B) = \frac{1}{\lambda_i \lambda_j |A|} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{[x \in A, y-x \in B]} \, \alpha_{ij}^{(2)}(\mathrm{d}x, \mathrm{d}y),$$

where $\alpha_{ij}^{(2)}(B_1 \times B_2) = \alpha_{\mathrm{m}}^{(2)}(B_1 \times \{i\} \times B_2 \times \{j\})$. For the second-order product density $\lambda_{\mathrm{m}}^{(2)}((x,i),(y,j))$, we use shorter notation $\lambda_{ij}^{(2)}(x,y)$. From stationarity we have that $\lambda_{ij}^{(2)}(x,y) = \lambda_{ij}^{(2)}(y-x)$ is the function of $y-x$. The definition of second-order factorial moment measure yields $\alpha_{ij}^{(2)}(B_1 \times B_2) = \alpha_{ji}^{(2)}(B_2 \times B_1)$. Therefore, $\lambda_{ij}^{(2)}(x,y) = \lambda_{ji}^{(2)}(y,x)$, which can be written as $\lambda_{ij}^{(2)}(u) = \lambda_{ji}^{(2)}(-u)$ for $u \in \mathbb{R}^d$. Using the last relation, we have

$$\begin{aligned}
\mathcal{K}_{ij}(B) &= \frac{1}{\lambda_i \lambda_j |A|} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{[x \in A, y-x \in B]} \lambda_{ij}^{(2)}(y-x) \, \mathrm{d}x \, \mathrm{d}y \\
&= \frac{1}{\lambda_i \lambda_j |A|} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{[x \in A, u \in B]} \lambda_{ij}^{(2)}(u) \, \mathrm{d}x \, \mathrm{d}u \\
&= \frac{1}{\lambda_i \lambda_j} \int_B \lambda_{ij}^{(2)}(u) \, \mathrm{d}u = \int_B g_{ij}(u) \, \mathrm{d}u.
\end{aligned}$$

By similar arguments, we get

$$\begin{aligned}
\mathcal{K}_{ji}(-B) &= \frac{1}{\lambda_i \lambda_j |A|} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{[x \in A, y-x \in -B]} \lambda_{ji}^{(2)}(y-x) \, \mathrm{d}x \, \mathrm{d}y \\
&= \frac{1}{\lambda_i \lambda_j |A|} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{[x \in A, u \in -B]} \lambda_{ji}^{(2)}(u) \, \mathrm{d}x \, \mathrm{d}u \\
&= \frac{1}{\lambda_i \lambda_j |A|} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{[x \in A, u \in B]} \lambda_{ij}^{(2)}(u) \, \mathrm{d}x \, \mathrm{d}u = \frac{1}{\lambda_i \lambda_j} \int_B \lambda_{ij}^{(2)}(u) \, \mathrm{d}u = \int_B g_{ij}(u) \, \mathrm{d}u.
\end{aligned}$$

□

Corollary 88. *The pair correlation function of a motion-invariant multivariate point process is symmetric: $g_{ij}(r) = g_{ji}(r)$. Furthermore, the following relation with the cross K-function holds,*

$$g_{ij}(r) = \frac{K_{ij}'(r)}{\sigma_d r^{d-1}}, \quad r > 0.$$

Proof: The proof is analogous to the proof of the relation between the pair correlation function and the K-function for point processes (Corollary 62). Theorem 87 states that the pair correlation function is the density of the cross second-order reduced moment measure,

$$\mathcal{K}_{ij}(B) = \int_B g_{ij}(u) \, \mathrm{d}u.$$

Now isotropy and polar decomposition of the Lebesgue measure lead to

$$K_{ij}(r) = \int_{b(o,r)} g_{ij}(\|u\|) \, \mathrm{d}u = \int_0^r \sigma_d s^{d-1} g_{ij}(s) \, \mathrm{d}s.$$

□

5.4 Processes with quantitative marks

Now we move from qualitative marks to quantitative ones. We will assume that the mark space is $\mathbb{M} = \mathbb{R}^+$. In this subsection, Φ_{m} will be a stationary simple marked point process with intensity λ and stationary mark distribution $\mathbb{Q}$. As usual, the corresponding unmarked point process is denoted by Φ. Again, we define basic numerical and functional summary characteristics. Naturally, we can apply commonly used characteristics to describe stationary mark distribution. If M_0 is the typical mark of Φ_{m}, then $F(t) = \mathbb{Q}([0,t])$ is its distribution function, $\mathbb{E}M_0 = \int_0^\infty m\,\mathbb{Q}(\mathrm{d}m)$ is the mean typical mark and $\operatorname{var} M_0$ is the variance of typical mark.

Definition 107. The *mark-sum measure* is defined as

$$S(B) = \int_{B\times\mathbb{M}} m\,\Phi_{\mathrm{m}}(\mathrm{d}(x,m)), \quad B \in \mathcal{B}^d.$$

Remark 64. It is a stationary random measure on $\mathbb{R}^d$. Its intensity is

$$\lambda_S = \mathbb{E}S([0,1]^d) = \lambda\mathbb{E}M_0.$$

This follows from Theorem 80 with $h(x,m) = m\mathbf{1}_{[x\in[0,1]^d]}$.

Definition 108. We define the *index of mark-sum dispersion* as

$$\mathrm{IMD} = \frac{\operatorname{var} S(B)}{\lambda|B|\mathbb{E}M_0},$$

where B is some test set, e.g. a ball of radius r. This index gives the ratio of the variance of the mark-sum in B and the corresponding mean.

Definition 109. Let Z_1 be a point of Φ that is the nearest neighbour of o. For a measurable function $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, we define the *non-normalised nearest-neighbour correlation index* as

$$\bar{\nu}_f = \mathbb{E}^!_o f(M(o), M(Z_1)),$$

where $M(o)$ denotes the mark of the origin and $M(Z_i)$ denotes the mark of Z_i. If we put

$$c_f = \int_0^\infty \int_0^\infty f(m_1,m_2)\,\mathbb{Q}(\mathrm{d}m_1)\,\mathbb{Q}(\mathrm{d}m_2)$$

and if $c_f > 0$, then we get the *nearest-neighbour correlation index*

$$\bar{n}_f = \frac{\bar{\nu}_f}{c_f}.$$

Remark 65. A value of $\bar{n}_f$ larger than 1 indicates that the mean function f of the marks at the typical point and its nearest neighbour is greater than the mean function of the typical mark and its independent copy. We have $\bar{n}_f = 1$ when the marks of the typical point and its nearest neighbour are independent. The most common choice of f is $f(m_1,m_2) = m_1m_2$. Then we speak about the *nearest-neighbour mark product index*. The values larger than 1 happen when the mean mark product of the typical point and its nearest neighbour is above the average, i.e. it indicates some mutual stimulation. Other common choices for f are $f(m_1,m_2) = m_2$ or $f(m_1,m_2) = \frac{1}{2}(m_1-m_2)^2$.

In the following, we assume that Φ_{m} is motion-invariant.

Definition 110. For a measurable function $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, we define the *second-order factorial moment measure associated with* f by the relation

$$\alpha^{(2)}_f(B_1\times B_2) = \mathbb{E}\sum^{\neq}_{(X_1,M_1),(X_2,M_2)\in\operatorname{supp}\Phi_{\mathrm{m}}} f(M_1,M_2)\mathbf{1}_{[X_1\in B_1, X_2\in B_2]}, \quad B_1,B_2\in\mathcal{B}^d.$$

Assume that there exists a density $\lambda_f^{(2)}(x,y)$ of this measure w.r.t. the Lebesgue measure. From stationarity and isotropy, it follows that $\lambda_f^{(2)}(x,y) = \lambda_f^{(2)}(\|x-y\|)$ is the function of $\|x-y\|$. If the second-order product density $\lambda^{(2)}$ of Φ exists, we define the *non-normalized f-mark correlation function* as

$$\kappa_f(r) = \frac{\lambda_f^{(2)}(r)}{\lambda^{(2)}(r)}, \quad r > 0 : \lambda^{(2)}(r) > 0.$$

We put

$$c_f = \int_0^\infty \int_0^\infty f(m_1, m_2)\, \mathbb{Q}(\mathrm{d}m_1)\, \mathbb{Q}(\mathrm{d}m_2)$$

and assume that $c_f > 0$. Then the *f-mark correlation function* is

$$k_f(r) = \frac{\kappa_f(r)}{c_f}, \quad r > 0.$$

Remark 66. The (non-normalized) f-mark correlation function can take arbitrary non-negative values. Therefore, the notion "correlation function" could be misleading.

The interpretation of the non-normalized f-mark correlation function is clearer from the following lemma.

Lemma 89. *Let Φ_{m} be a motion-invariant simple marked point process. Denote by $M(o)$ the mark of the origin o and by $M(r)$ the mark of arbitrary point in the distance r from o. The non-normalized f-mark correlation function satisfies*

$$\kappa_f(r) = \mathbb{E}_{or} f(M(o), M(r)), \quad r > 0.$$

Proof: We rewrite $\alpha_f^{(2)}$ by the second-order Campbell theorem (Theorem 43) and apply Theorem 81:

$$\begin{aligned}\alpha_f^{(2)}(B_1 \times B_2) &= \int_{B_2 \times \mathbb{R}^+} \int_{B_1 \times \mathbb{R}^+} f(m_1, m_2)\, \alpha_{\mathrm{m}}^{(2)}(\mathrm{d}(x_1, m_1), \mathrm{d}(x_2, m_2)) \\ &= \int_{B_2} \int_{B_1} \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} f(m_1, m_2)\, P_{x_1,x_2}(\mathrm{d}m_1, \mathrm{d}m_2)\, \alpha^{(2)}(\mathrm{d}x_1, \mathrm{d}x_2).\end{aligned}$$

From this we see that

$$\lambda_f^{(2)}(x,y) = \lambda^{(2)}(x,y) \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} f(m_1, m_2)\, P_{x,y}(\mathrm{d}m_1, \mathrm{d}m_2).$$

Consequently,

$$\kappa_f(r) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} f(m_1, m_2)\, P_{or}(\mathrm{d}m_1, \mathrm{d}m_2).$$

□

The function $\kappa_f(r)$ represents the mean of f-function of two marks corresponding to the points of the process at distance r. It is also seen that the normalization c_f is chosen so that $k_f(r)$ is equal to 1 whenever the marks at distance r are independent. By taking special choices of the function f, different second-order characteristics are obtained. The most often used choices of f are the following:

- $f(m_1, m_2) = c(m_1, m_2) = m_1 m_2$,
- $f(m_1, m_2) = e(m_1, m_2) = m_1$,
- $f(m_1, m_2) = e^*(m_1, m_2) = m_2$,
- $f(m_1, m_2) = v(m_1, m_2) = m_1^2$,
- $f(m_1, m_2) = v^*(m_1, m_2) = m_2^2$,
- $f(m_1, m_2) = \gamma(m_1, m_2) = \frac{1}{2}(m_1 - m_2)^2$.

These choices of f in κ_f lead to the functions κ_c, κ_e, κ_{e^*}, κ_v, κ_{v^*}, and κ_γ. Note that $\kappa_e = \kappa_{e^*}$ and $\kappa_v = \kappa_{v^*}$.

Definition 111. The function $\kappa_e(r)$ is sometimes called the *E-function*,

$$E(r) = \mathbb{E}_{or} M(o), \quad r > 0.$$

We define the *V-function* as

$$V(r) = \kappa_v(r) - \kappa_e(r)^2 = \mathbb{E}_{or}(M(o) - E(r))^2, \quad r > 0.$$

Furthermore, we define the *mark covariance function* as

$$\mathrm{cov}(r) = \kappa_c(r) - \kappa_e(r)^2 = \mathbb{E}_{or}(M(o) - E(r))(M(r) - E(r)), \quad r > 0,$$

and the *mark correlation function* as

$$\mathrm{cor}(r) = \frac{\mathrm{cov}(r)}{V(r)}, \quad r > 0.$$

Finally, the *mark variogram* is

$$\kappa_\gamma(r) = \frac{1}{2}\mathbb{E}_{or}(M(o) - M(r))^2 = \kappa_v(r) - \kappa_c(r), \quad r > 0.$$

Examples of normalized f-mark correlation functions are the following.

Definition 112. If the mean typical mark is positive, we define *Stoyan's mark correlation function* as

$$k_c(r) = \frac{\mathbb{E}_{or} M(o)M(r)}{(\mathbb{E}M_0)^2}, \quad r > 0,$$

i.e. it is $k_f(r)$ with $f(m_1, m_2) = c(m_1, m_2) = m_1 m_2$, and the *r-mark function* as

$$k_e(r) = \frac{\mathbb{E}_{or} M(o)}{\mathbb{E}M_0}, \quad r > 0,$$

i.e. $k_e(r) = k_f(r)$ with $f(m_1, m_2) = e(m_1, m_2) = m_1$. Sometimes different notation is used in the literature: $k_c(r) = k_{mm}(r)$ and $k_e(r) = k_{m\cdot}(r)$.

Remark 67. The function $E(r)$ gives the mean mark of the point for which there is another point of the process at distance r. When marks and locations are not independent, the existence of another point at distance r may influence the magnitude of the mark of a given point. By normalizing with the mean typical mark, we obtain the function $k_e(r)$.

Now we extend Definition 106.

Definition 113. The *mark-weighted reduced second-order moment measure* $\mathcal{K}_f$ is defined by the relation

$$\lambda c_f \mathcal{K}_f(B) = \mathbb{E}_o^! \sum_{(X,M)\in \mathrm{supp}\,\Phi_{\mathrm{m}}} f(M(o), M)\mathbf{1}_B(X), \quad B \in \mathcal{B}^d.$$

The *mark-weighted K-function* is then defined as

$$K_f(r) = \mathcal{K}_f(b(o, r)), \quad r \geq 0.$$

Using (29) we can also write

$$\lambda c_f \mathcal{K}_f(B) = \int_{\mathbb{M}} \mathbb{E}^!_{(o,m)} \sum_{(X,M)\in \mathrm{supp}\,\Phi_{\mathrm{m}}} f(m, M)\mathbf{1}_B(X)\,\mathbb{Q}(\mathrm{d}m). \tag{31}$$

The following lemma provides an alternative way how to introduce the mark-weighted reduced second-order moment measure. It is analogous to Lemma 58.

Lemma 90. *Let Φ_{m} be a stationary simple marked point process with intensity $0 < \lambda < \infty$ and stationary mark distribution $\mathbb{Q}$. Assume that $c_f = \int\int f(m_1, m_2)\,\mathbb{Q}(\mathrm{d}m_1)\,\mathbb{Q}(\mathrm{d}m_2) > 0$. Then for arbitrary $A \in \mathcal{B}_0^d$ with $|A| > 0$, we have*

$$\mathcal{K}_f(B) = \frac{1}{\lambda^2 c_f |A|}\mathbb{E}\sum_{(X_1,M_1),(X_2,M_2)\in\operatorname{supp}\Phi_{\mathrm{m}}}^{\neq} f(M_1,M_2)\mathbf{1}_A(X_1)\mathbf{1}_B(X_2-X_1), \quad B\in\mathcal{B}^d.$$

Proof: Let

$$h_B(m,\nu) = \sum_{(y,\tilde{m})\in\operatorname{supp}\nu} f(m,\tilde{m})\mathbf{1}_B(y), \quad B\in\mathcal{B}^d,\ m\in\mathbb{M},\ \nu\in\mathcal{N}_{\mathrm{m}}^*.$$

From (28), stationarity of Φ_{m}, and (29), we obtain

$$\begin{aligned}
&\mathbb{E}\sum_{(X_1,M_1),(X_2,M_2)\in\operatorname{supp}\Phi_{\mathrm{m}}}^{\neq} f(M_1,M_2)\mathbf{1}_A(X_1)\mathbf{1}_B(X_2-X_1)\\
&\qquad = \mathbb{E}\sum_{(X_1,M_1)\in\operatorname{supp}\Phi_{\mathrm{m}}} h_{B+X_1}(M_1,\Phi_{\mathrm{m}}-\delta_{(X_1,M_1)})\mathbf{1}_A(X_1)\\
&\qquad = \lambda\int_A\int_{\mathbb{M}}\int_{\mathcal{N}_{\mathrm{m}}^*} h_{B+x}(m,t_x\nu)\,\mathbb{P}^!_{(o,m)}(\mathrm{d}\nu)\,\mathbb{Q}(\mathrm{d}m)\,\mathrm{d}x\\
&\qquad = \lambda\int_A\int_{\mathbb{M}}\int_{\mathcal{N}_{\mathrm{m}}^*} h_B(m,\nu)\,\mathbb{P}^!_{(o,m)}(\mathrm{d}\nu)\,\mathbb{Q}(\mathrm{d}m)\,\mathrm{d}x\\
&\qquad = \lambda\int_A\int_{\mathbb{M}}\mathbb{E}^!_{(o,m)}\sum_{(Y,\tilde{M})\in\operatorname{supp}\Phi_{\mathrm{m}}} f(m,\tilde{M})\mathbf{1}_B(Y)\,\mathbb{Q}(\mathrm{d}m)\,\mathrm{d}x\\
&\qquad = \lambda\int_A \lambda c_f\mathcal{K}_f(B)\,\mathrm{d}x = \lambda^2 c_f|A|\mathcal{K}_f(B).
\end{aligned}$$

□

The mark-weighted K-function is related to the f-mark correlation function.

Lemma 91. *Let Φ_{m} be a motion-invariant simple marked point process. Assume that the pair correlation function g of the unmarked point process Φ exists. Then*

$$K_f'(r) = \sigma_d r^{d-1}k_f(r)g(r), \quad r\geq 0,$$

where $k_f(r)$ is the f-mark correlation function of Φ_{m}.

Proof: From Lemma 90 we obtain

$$\begin{aligned}
K_f(r) &= \frac{1}{\lambda^2 c_f|A|}\int\int f(m_1,m_2)\mathbf{1}_{[x_1\in A,\|x_2-x_1\|\leq r]}\,\alpha_{\mathrm{m}}^{(2)}(\mathrm{d}(x_1,m_1),\mathrm{d}(x_2,m_2))\\
&= \frac{1}{\lambda^2 c_f|A|}\int_{\mathbb{R}^d}\int_A \mathbf{1}_{[\|x_2-x_1\|\leq r]}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+} f(m_1,m_2)\,P_{x_1,x_2}(\mathrm{d}m_1,\mathrm{d}m_2)\,\alpha^{(2)}(\mathrm{d}x_1,\mathrm{d}x_2).\\
&= \frac{1}{\lambda^2 c_f|A|}\int_{\mathbb{R}^d}\int_A \mathbf{1}_{[\|x_2-x_1\|\leq r]}\kappa_f(\|x_2-x_1\|)\lambda^{(2)}(\|x_2-x_1\|)\,\mathrm{d}x_1\,\mathrm{d}x_2\\
&= \frac{1}{|A|}\int_A\int_0^r \sigma_d s^{d-1}k_f(s)g(s)\,\mathrm{d}s\,\mathrm{d}x_1 = \int_0^r \sigma_d s^{d-1}k_f(s)g(s)\,\mathrm{d}s.
\end{aligned}$$

We used Theorem 43, Theorem 81, motion-invariance of Φ_{m}, Lemma 89, and substitution $x_2 - x_1 = su$ with $s\in\mathbb{R}^+$ and $\|u\| = 1$. □

6. Appendix

6.1 Gaussian distribution

Theorem 92. *Assume that the random vector $\boldsymbol{X} = (X_1, \dots, X_n)^{\mathrm{T}}$ has the n-dimensional normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Sigma}$. For $k \in \{1, \dots, n-1\}$, divide the vector $\boldsymbol{X}$ into $\boldsymbol{X}_1 = (X_1, \dots, X_k)^{\mathrm{T}}$ and $\boldsymbol{X}_2 = (X_{k+1}, \dots, X_n)^{\mathrm{T}}$, the vector $\boldsymbol{\mu}$ into $\boldsymbol{\mu}_1 = (\mu_1, \dots, \mu_k)^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = (\mu_{k+1}, \dots, \mu_n)^{\mathrm{T}}$, and the matrix $\boldsymbol{\Sigma}$ into 4 submatrices $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$, and $\boldsymbol{\Sigma}_{22}$ of orders $k \times k$, $k \times (n-k)$, $(n-k) \times k$, and $(n-k) \times (n-k)$, respectively,*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Then the conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is the k-dimensional normal distribution with mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)$ and covariance matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Proof: It can be found e.g. in [1] as Theorem 4.12.

□

Theorem 93. *Let $X_1, \dots, X_n$ be independent identically distributed random variables with standard normal distribution $N(0,1)$. Let $h : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary measurable scale-invariant function, that is $h(ax_1, \dots, ax_n) = h(x_1, \dots, x_n)$ for every $a > 0$ and $x_1, \dots, x_n \in \mathbb{R}$. Then the random variables $H = h(X_1, \dots, X_n)$ and $Q = X_1^2 + \cdots + X_n^2$ are independent.*

Proof: Assume that there exists the moment generating function of the random variable H at some neighbourhood U of zero. Then the moment generating function of the random vector $(H, Q)^{\mathrm{T}}$ is

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{t_1 H + t_2 Q} &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{t_1 h(x_1,\dots,x_n)} \mathrm{e}^{-(1-2t_2)\sum_{i=1}^n x_i^2/2} \,\mathrm{d}(x_1, \dots, x_n) \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{t_1 h(y_1,\dots,y_n)} \mathrm{e}^{-\sum_{i=1}^n y_i^2/2} (1-2t_2)^{-n/2} \,\mathrm{d}(y_1, \dots, y_n) \\
&= (1-2t_2)^{-n/2} \mathbb{E}\mathrm{e}^{t_1 H}, \quad t_1 \in U, t_2 < 1/2,
\end{aligned}$$

which is in the form of the product of two moment generating functions. Therefore, H and Q are independent and Q has the χ^2-distribution with n degrees of freedom. If the moment generating function of H does not exist, we have to use the characteristic function to reach the same conclusion.

□

6.2 Bessel functions

The *Bessel function of the first kind* of order ν is defined as

$$J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(k+\nu+1)} \left(\frac{z}{2}\right)^{2k+\nu}, \quad z \in \mathbb{C}.$$

It is a solution of the Bessel differential equation

$$y'' + \frac{1}{z}y' + \left(1 - \frac{\nu^2}{z^2}\right) y = 0.$$

There exist two kinds of the *modified Bessel functions*. They are the solutions of the modified Bessel equation

$$w'' + \frac{1}{z}w' - \left(1 + \frac{\nu^2}{z^2}\right) w = 0.$$

The modified Bessel function of the first kind of order ν is

$$I_\nu(z) = \mathrm{i}^{-\nu} J_\nu(z\mathrm{i}) = \mathrm{e}^{-\frac{1}{2}\nu\pi\mathrm{i}} J_\nu(z\mathrm{i}) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+\nu+1)} \left(\frac{z}{2}\right)^{2k+\nu}$$

and the modified Bessel function of the second kind of order ν has the form

$$K_\nu(z) = \frac{\pi}{2} \cdot \frac{I_{-\nu}(z) - I_\nu(z)}{\sin \pi\nu}.$$

In particular, for $\nu = 1/2$, we get

$$I_{1/2}(z) = \sqrt{\frac{2}{\pi z}} \sinh z, \quad K_{1/2}(z) = \sqrt{\frac{\pi}{2}} \mathrm{e}^{-z} z^{-1/2}.$$

6.3 Measure theory

Let E be a given set. We recall some basic definitions of families of certain subsets of E.

Definition 114. A collection $\mathcal{R}$ is called a *ring* if it contains the empty set $\emptyset$ and it is closed under finite unions and set differences. If it additionally contains the whole space E, then it is called an *algebra*.

Definition 115. A collection $\mathcal{P}$ is a *π-system* if it is non-empty and closed under finite intersections ($A \cap B \in \mathcal{P}$ whenever $A, B \in \mathcal{P}$).

Definition 116. A collection $\mathcal{D}$ is called a *Dynkin system* or *λ-system* if
(i) $E \in \mathcal{D}$,
(ii) $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$,
(iii) $A_1, A_2, \ldots \in \mathcal{D}$ pairwise disjoint $\Rightarrow \cup_i A_i \in \mathcal{D}$.

Remark 68. A σ-algebra is both a π-system and a Dynkin system. Conversely, a class that is both a π-system and a Dynkin system is σ-algebra.

We often use the following result.

Theorem 94. *(Dynkin) Let $\mathcal{P}$ be a π-system and $\mathcal{D}$ a Dynkin system such that $\mathcal{P} \subseteq \mathcal{D}$. Then $\sigma\mathcal{P} \subseteq \mathcal{D}$.*

Proof: See Theorem 1.1 in [9].

Definition 117. We say that $\mathcal{M}$ is a *monotone system* if
(i) $A_1 \subseteq A_2 \subseteq \cdots \in \mathcal{M} \Rightarrow \cup_i A_i \in \mathcal{M}$,
(ii) $A_1 \supseteq A_2 \supseteq \cdots \in \mathcal{M} \Rightarrow \cap_i A_i \in \mathcal{M}$.

Remark 69. A Dynkin system is a monotone system.

Theorem 95. *(monotone class theorem) If $\mathcal{R}$ is a ring and $\mathcal{M}$ is a monotone system such that $\mathcal{R} \subseteq \mathcal{M}$, then $\sigma\mathcal{R} \subseteq \mathcal{M}$.*

Proof: See 11.4 in [11].

Theorem 96. *(uniqueness of measure extension) Let $(E, \mathcal{E})$ be a measurable space and $\mathcal{S} \subseteq \mathcal{E}$ be a π-system such that $\sigma\mathcal{S} = \mathcal{E}$. If two σ-finite measures μ and ν coincide on $\mathcal{S}$, then $\mu = \nu$.*

Proof: See Lemma 1.17 in [9] or Lemma 2.2 in [14].

□

References

[1] J. Anděl (2011): *Základy matematické statistiky*, in Czech, 3. vydání, Matfyzpress, Praha.

[2] Z. Ciesielski and A. Kamont (1995): Levy's fractional Brownian random field and function spaces, *Acta Sci. Math. (Szeged)* **60**, 99–118.

[3] A. D. Cliff and J. K. Ord (1981): *Spatial Processes; Models and Applications*, Pion Limited, London.

[4] C. Comas, P. Delicado and J. Mateu (2011): A second order approach to analyse spatial point patterns with functional marks, *TEST* **20**, 503–523.

[5] D. J. Daley and D. Vere-Jones (2003): *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*, second edition, Springer-Verlag, New York.

[6] D. J. Daley and D. Vere-Jones (2008): *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*, second edition, Springer-Verlag, New York.

[7] U. Hahn, E. B. Vedel Jensen, M.-C. van Lieshout and L. S. Nielsen (2003): Inhomogeneous spatial point processes by location-dependent scaling, *Adv. Appl. Prob. (SGSA)* **35**, 319–336.

[8] E. Ising (1925): Beitrag zur Theorie des Ferromagnetismus, *Z. Physik* **31**, 253–258.

[9] O. Kallenberg (2002): *Foundations of Modern Probability*, 2nd edition, Springer, New York.

[10] I. Karatzas and S. E. Shreve (1988): *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

[11] J. Lukeš and J. Malý (2005): *Measure and Integral*, Matfyzpress, Prague.

[12] L. Onsager (1944): Crystal statistics. I. A two-dimensional model with an order-disorder transition, *Phys. Rev.* **65**, 117–149.

[13] Z. Prášková (2004): *Základy náhodných procesů II*, in Czech, Karolinum, Praha.

[14] J. Rataj (2006): *Bodové procesy*, in Czech, Karolinum, Praha.

[15] H. Rue and L. Held (2005): *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC, Boca Raton.

[16] Z. Sasvári (1994): *Positive Definite and Definitizable Functions*, Akademie Verlag, Berlin.

[17] R. Schneider, W. Weil (2008): *Stochastic and Integral Geometry*, Springer-Verlag, Berlin.

[18] D. J. Strauss (1975): A model for clustering, *Biometrika* **62**, 467–475.

[19] J. Štěpán (1987): *Teorie pravděpodobnosti – matematické základy*, in Czech, Academia, Praha.

[20] E. B. Vedel Jensen and L. S. Nielsen (2000): Inhomogeneous Markov point processes by transformation, *Bernoulli* **6**, 761–782.

[21] P. Whittle (1954): On stationary processes in the plane, *Biometrika* **41**, 434–449.