Spatial Modelling (NMTP438)

April 16, 2018

Contents

1. Spatial models on lattices

1.1 Markov random fields

Let $L \subseteq \mathbb{R}^d$ be a finite 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). The elements of L (called *sites*) can also represent some regions (an example is shown in Figure 1 right). The number of sites will be denoted by $n = |L|$.

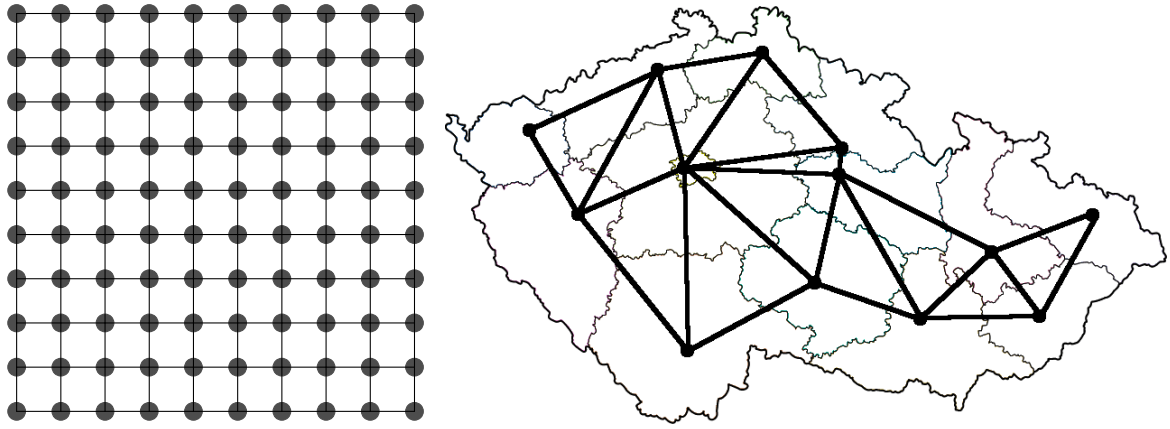

Figure 1. Two examples of lattices in the plane. Left: regular square grid 10×10. Right: capitals of 13 regions of the Czech Republic, two sites are connected if the regions share a common border.

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 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.

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. 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 together with relation $\sim$ generates 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.

Definition 3. A set $A \subseteq L$ is called a *clique* w.r.t. $\sim$ if $i \sim j$ for any $i, j \in A$, $i \neq j$.

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 the 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^A-a.a. $\boldsymbol{z} \in S^A$ satisfying $p(\boldsymbol{z}_A) > 0$ and for any $i \in L$ and $\partial i \subseteq A \subseteq L$.

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*

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

where $\mathcal{C} = \{C \subseteq L : C \text{ is a clique}\}$.

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

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

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

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

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 $p(\boldsymbol{z}) = \frac{1}{Z} \exp\{-E/T\}$. 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 Z is a normalizing constant called the *partition function*,

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

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 $Z = \mathrm{e}^{\Phi_\emptyset}$.

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 the 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:

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

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 the lattice L represent pixels, $z_i=1$ usually denotes black colour and $z_i=0$ white colour of the pixel i. 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)

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

where

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

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 is proportional to $\mathrm{e}^{\beta k}$. 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.

This random field is known as the *Ising model* [6]. 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 [10]), at which the phase transition occurs. For $\beta\leq\beta_c$ the Gibbs distribution is unique, while 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.

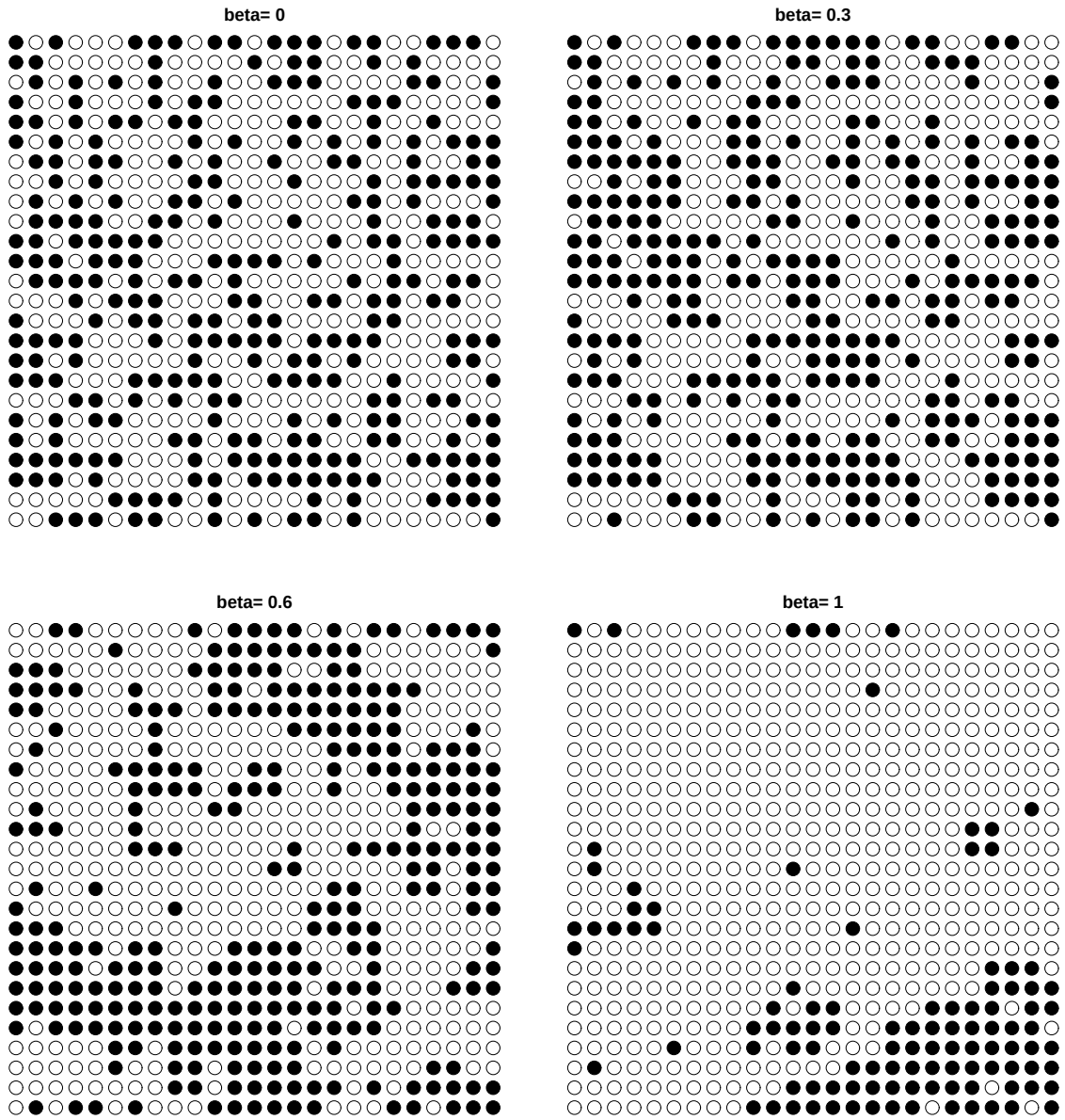

Figure 2. Simulations of the Ising model on the rectangular lattice 25×25 for $\beta \in \{0; 0.3; 0.6; 1\}$.

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 the 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

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

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 shades of grey. 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 the 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 the *autologistic model*.

Example: Let us move to the countable state space: $S = \mathbb{N}_0$. In practice, this situation appears when we deal with count data, e.g. disease-incidence counts in some region. Consider a model where these counts have a 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:

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

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 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 n-dimensional Gaussian distribution with the 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 the mean

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

and variance $1/q_{ii}$ (cf. Theorem 54). 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. A random field $\boldsymbol{Z} = \{Z_i : i \in L\}$ with n-dimensional normal distribution with positive definite covariance matrix $\boldsymbol{\Sigma}$ is called a *Gaussian Markov random field*. It is a particular case of a Markov random field w.r.t. the relation defined by $i \sim j \Leftrightarrow q_{ij} \neq 0$, 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 for $\Sigma = \sigma^2 \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. Then any two distinct sites are not neighbours.

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 Gaussian 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 [13].

The Gaussian Markov random fields were defined by their joint density that is 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 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:*

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

Proof: From the relation

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

it follows that

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

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

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

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

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

□

The Brook lemma gives instructions how to get the joint density from the conditional densities. We fix some $\boldsymbol{y}$ and apply the Brook lemma 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 the 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 *conditional autoregressive model* which is abbreviated as CAR.

If we choose Gaussian conditional densities, then we obtain Gaussian CAR model. 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 Gaussian distribution with the mean $\sum_{j\in L} b_{ij}Z_j$ and variance τ_i^2. Enumerate the sites of L by $1,\dots,n$ and fix $\boldsymbol{y} = o$ as the null vector in $\mathbb{R}^n$. Then by Lemma 5 we have

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

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,

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

where q_{ij} are the elements of the matrix $\boldsymbol{Q} = (\boldsymbol{I}-\boldsymbol{B})\boldsymbol{D}^{-1}$. 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 the 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. In this case we obtain the Gaussian CAR model which 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 $\boldsymbol{Q}^{-1} = \boldsymbol{D}(\boldsymbol{I}-\boldsymbol{B})^{-1}$, 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 incooperate 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 the 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 11. Let $\boldsymbol{\varepsilon} = \{\varepsilon_i : i\in L\}$ be a Gaussian white noise. Again 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* 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 Gaussian. 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 $\mathrm{cov}(\boldsymbol{\varepsilon},\boldsymbol{Z}) = \sigma^2(\boldsymbol{I}-\boldsymbol{B}^{\mathrm{T}})^{-1}$, 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 12. 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 the 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 the 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 to model 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 usually closer observations are 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 occured 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 13. 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 $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$. 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 $\mathbb{P}(Z_i = 1, Z_j = 0) = \frac{2m(n-m)}{n(n-1)}$. 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 have tendency 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 14. 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 distinct 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 $n!$ permutation 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*

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

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

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

$$\begin{aligned}
\mathbb{E}_g M_2 &= \sum_{i\in L} \mathbb{E}_g Y_i^2 = (n-1)\sigma^2, \\
\mathbb{E}_g R &= \sum_{i\in L}\sum_{j\in L} w_{ij}\mathbb{E}_g Y_i Y_j = -\frac{\sigma^2}{n} w, \\
\mathbb{E}_g V &= \sum_{i\in L}\sum_{j\in L} w_{ij}\mathbb{E}_g (Z_i - Z_j)^2 = 2\sigma^2 w.
\end{aligned}$$

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 55. 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 55 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 the fact 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 different (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 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 have tendency 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 expectation. 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 the case when D is connected and has 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 [11].

2.1 Basic definitions

Definition 15. 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

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

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

Definition 16. We say that the system of distribution functions $\{F_{x_1,\ldots,x_n} : n \in \mathbb{N}, x_1,\ldots,x_n \in D\}$ is *projective* if for any $n \in \mathbb{N}$, $x_1,\ldots,x_n \in D$, $t_1,\ldots,t_n \in \mathbb{R}$ and a permutation $i_1,\ldots,i_n$ of numbers $1,\ldots,n$ we have

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

and

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

Distribution functions of finite-dimensional distributions of the random field clearly 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: [15], Theorem I.10.3 or [7], Theorem 6.16.

□

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

Definition 17. A random field $\{Z(x) : x \in D\}$ is called *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.

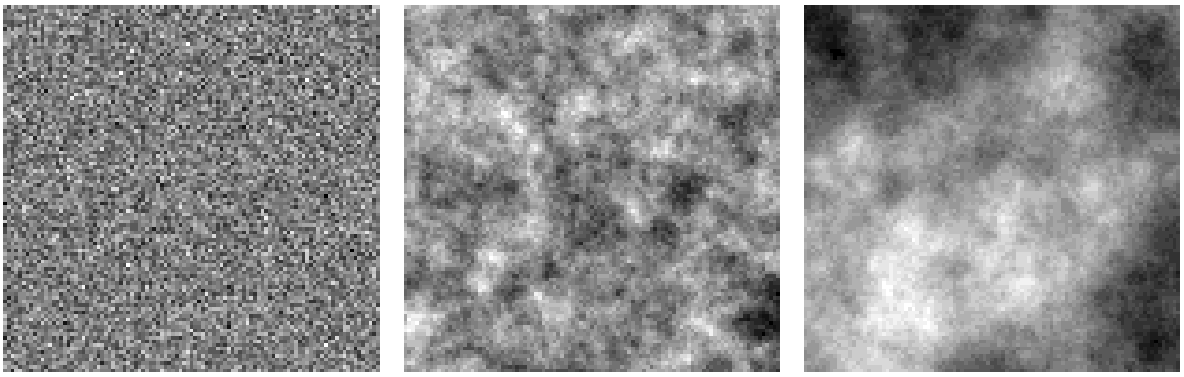

Figure 3. Simulated realizations of Gaussian random fields in the square $D = [0, 10]^2$. Constant mean and variance ($\mu(x) = 0$ and $\operatorname{var} Z(x) = 1$ for each $x \in D$) and three different autocovariance functions were considered. Left: there are no correlations ($C(x, y) = \mathbf{1}_{[x=y]}$). 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 18. 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 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 further kind of stationarity.

Definition 19. 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. The reverse implication 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 20. 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 21. 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 weak motion-invariant random field $\{Z(x) : x \in \mathbb{R}^d\}$ we have we have $C(x, y) = C(\|x - y\|)$ for any $x, y \in D$. Again we slightly abuse the notation and use C both for the function of two arguments x and y and for the function of one argument $\|x - y\| \in \mathbb{R}^+$.

Definition 22. 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 23. 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: [8], Problem 2.9. □

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

Definition 24. 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 $\{e_1, \dots, \mathrm{e}_d\}$ be the canonical basis of $\mathbb{R}^d$, then $Z'(x, e_j)$, $j = 1, \dots, d$, 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 so called variogram.

Definition 25. 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 the *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.

From definition it is clear that $\gamma(h) = \gamma(-h)$, $\gamma(o) = 0$ and $\gamma(h) \geq 0$. A weak motion-invariant random field has an isotropic (semi)variogram. The semivariogram of a weak stationary random field is bounded and it is related to the autocovariance function by the formula $\gamma(h) = C(o) - C(h)$ which follows from (9). Generally 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 so called *nugget effect*.

Definition 26. 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.

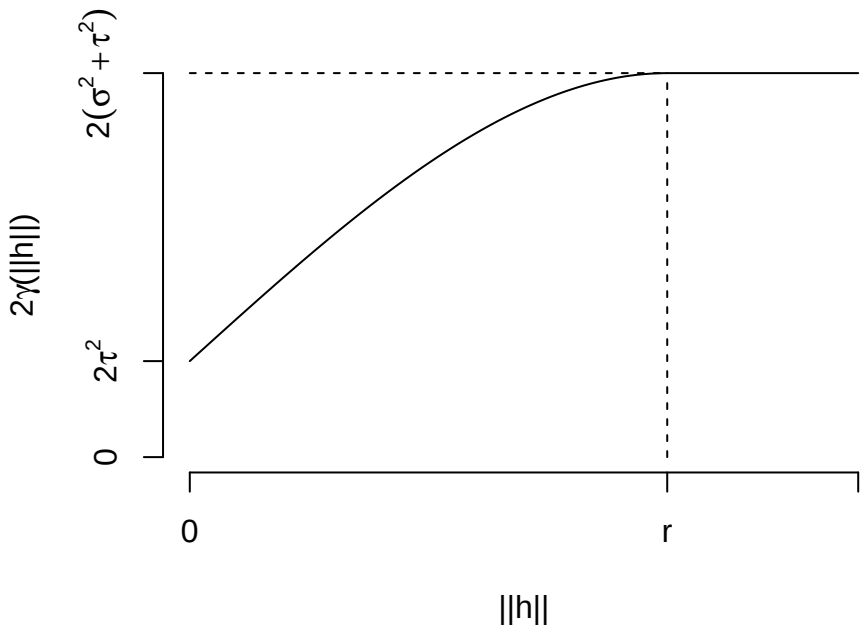

Figure 4. An illustration of the definition of the nugget, sill and range.

The nugget appears when the repeated measurements at the same location give different values. It happens, for example, when we are not observing directly the realization of a random field, 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 directly 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 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 $n \in \mathbb{N}$, $x_1, \ldots, x_n \in D$ and $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{R}$.

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 $n \in \mathbb{N}$, $x_1, \ldots, x_n \in D$ and $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{R}$ 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 27. 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 $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. 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 $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\beta_1, \ldots, \beta_n \in \mathbb{R}$ 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-values 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: [14], 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 with respect to 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 5.2). 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.

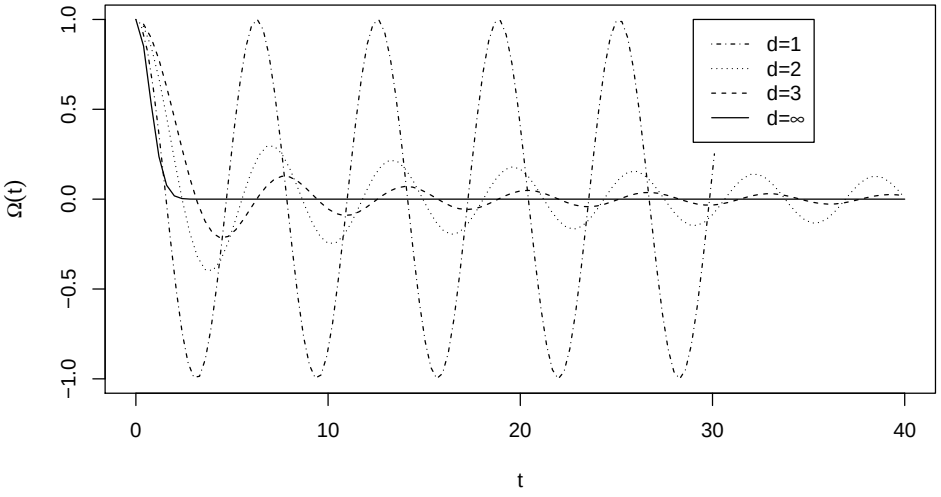

Figure 5. The basis functions $\Omega_d(t)$ for $d = 1$, $d = 2$, $d = 3$ and $d \to \infty$.

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

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

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

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: From the spectral decomposition it follows that

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

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 an isotropic semivariogram of an intrinsic stationary random field. In all these models one of the parameters is the nugget $2\tau^2 \geq 0$.

1. *power*:

$$\gamma(h) = \begin{cases} 0 & \text{pro } h = o, \\ \tau^2 + \sigma^2 \|h\|^\nu & \text{pro } h \neq o, \end{cases}$$

where $0 \leq \nu < 2$ and $\sigma > 0$. If $\nu > 0$, then this model does not have sill and the corresponding random field is not weak stationary. The case $\nu = 0$ is known as the *nugget model*. An example of a random field with this variogram is a white noise which is formed by uncorrelated random variables with constant mean and finite variance $\tau^2 + \sigma^2$. One realization of the white noise is shown in Figure 3 left. 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.

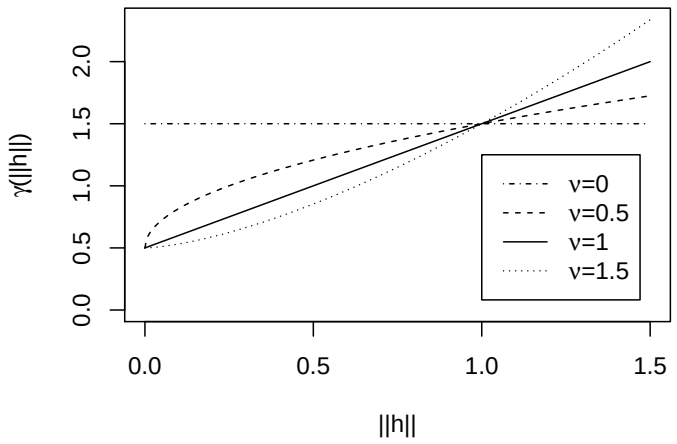

Figure 6. Power model of the isotropic semivariogram for $\tau^2 = 0.5$, $\sigma^2 = 1$ and different choices of the parameter ν.

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 $\sigma^2 \geq 0$, $\varrho > 0$. 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 the lower dimensions $d = 1$ and $d = 2$. However, the model that is valid in the 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*

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

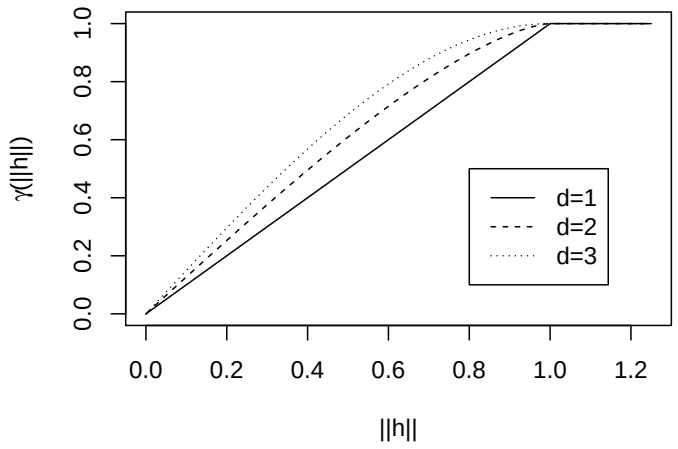

Figure 7. Spherical semivariogram in dimensions $d = 1$, $d = 2$ and $d = 3$ for the following choice of parameters: $\tau^2 = 0$, $\sigma^2 = 1$ and $r = 1$.

3. *generalized exponential*:

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

where $\sigma^2 \geq 0$, $a > 0$ 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.

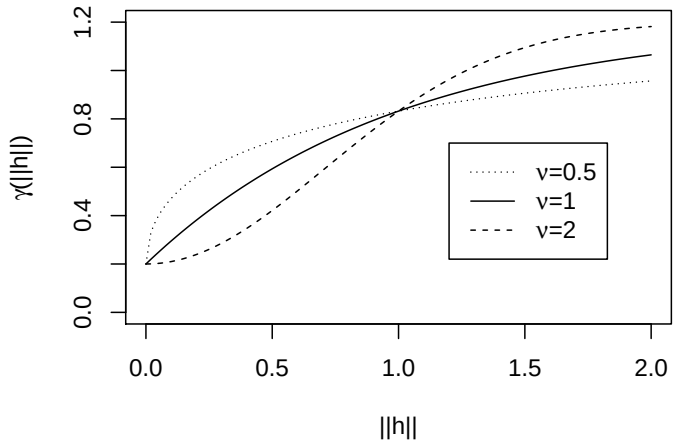

Figure 8. Generalized exponential model of the isotropic semivariogram for the following choice of parameters: $\tau^2 = 0.2$, $\sigma^2 = 1$, $a = 1$ and $\nu \in \{0.5, 1, 2\}$.

Both the spherical and the generalized exponential models have sill and thus they lead to a weak stationary random field with the autocovariance function $C(h) = \sigma^2 + \tau^2 - \gamma(h)$. The quite 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 5.2). 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 [16]. The spectal density of the autocovariance function (14) has the following form:

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

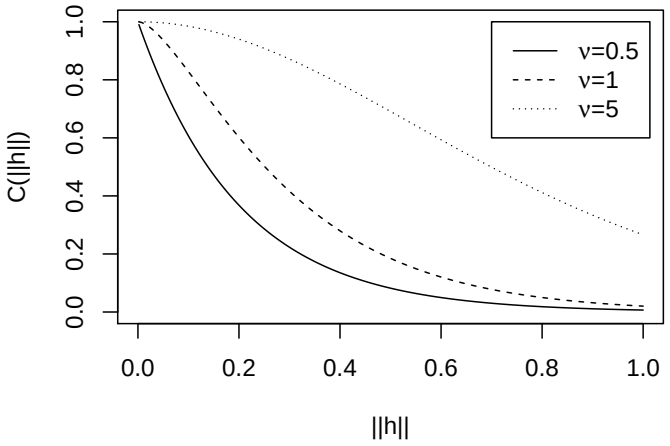

Figure 9. The Whittle-Matérn autocovariance function for $\alpha = 5$, $\sigma = 1$ and three values of ν.

3. Random measures

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 28. A measure μ on $(E, \mathcal{B})$ is *locally finite*, if it is finite on $\mathcal{B}_0$, i.e. $\mu(B) < \infty$ for any $B \in \mathcal{B}_0$. By $\mathcal{M} = \mathcal{M}(E)$ we denote the space of all locally finite measures on $(E, \mathcal{B})$ and by $\mathcal{M}_f = \mathcal{M}_f(E) = \{\mu \in \mathcal{M} : \mu(E) < \infty\}$ 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 $\mathcal{N}_f = \mathcal{M}_f \cap \mathcal{N}$ the space of all finite counting measures.

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

Proof: From our requirements on the space E it follows that there exist $K_n \in \mathcal{K}$ such that $E = \cup_n K_n$ with $\mu(K_n) < \infty$ for each $n \in \mathbb{N}$.

□

Definition 29. 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. Shortly we write

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

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 write shortly $\mathcal{M}_{B,[0,r)} = \mathcal{M}_{B,r}$ and $\mathfrak{M} = \sigma\{\mathcal{M}_{B,r} : B \in \mathcal{B}, r \in [0, \infty]\}$. Since $\mathcal{M}_f = \mathcal{M}_{E,\infty}$, it is clear that $\mathcal{M}_f \in \mathfrak{M}$. Consequently, $\mathcal{N}_f \in \mathfrak{N}$.

Lemma 18. *Let $\mathcal{S} \subseteq \mathcal{B}_0$ 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

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

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 56). 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 19. *It holds that $\mathcal{N} \in \mathfrak{M}$. Hence, $\mathfrak{N}$ is correctly defined.*

Proof: Consider the system $\mathcal{S}$ as in Lemma 18 that it is moreover countable. Then

$$\mathcal{N}_0 = \{\mu \in \mathcal{M} : \mu(A) \in \mathbb{N} \cup \{0\} \; \forall A \in \mathcal{S}\} \in \mathfrak{M},$$

because it is a countable intersection of measurable sets. Clearly $\mathcal{N} \subseteq \mathcal{N}_0$. Define $\mathcal{D}_n = \{B \in \mathcal{B} : \mu(B \cap A_n) \in \mathbb{N} \cup \{0\} \; \forall \mu \in \mathcal{N}_0\}$. It is a Dynkin system that contains $\mathcal{S}$. Hence, $\mathcal{D}_n = \mathcal{B}$ by Dynkin's theorem. 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 $B_1, \dots, B_n \in \mathcal{B}_0$ we denote the σ-algebra generated by the mappings $\pi_{B_1}, \dots, \pi_{B_n}$:

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

Lemma 20. *Let $\mathcal{S} \subseteq \mathcal{B}_0$ be a ring such that $\sigma\mathcal{S} = \mathcal{B}$. Then*

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

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

Proof: Since $\mathfrak{M}_{A_1,\dots,A_n}$ is a σ-algebra, the system $\mathfrak{M}_0$ is closed under complements and contains $\emptyset$ and $\mathcal{M}$. We can show that it is closed under finite unions (intersections) by considering sets $\mathcal{U} = \{\mu : \mu(A) \in D_1\} \in \mathfrak{M}_A$ and $\mathcal{V} = \{\mu : \mu(B) \in D_2\} \in \mathfrak{M}_B$. Then we have

$$\begin{aligned}\mathcal{U} \cap \mathcal{V} = \{\mu : \mu(A) \in D_1, \mu(B) \in D_2\} = \{\mu : \mu(A \setminus B) + \mu(A \cap B) \in D_1,\\ \mu(B \setminus A) + \mu(A \cap B) \in D_2\} \in \mathfrak{M}_{A\setminus B, A\cap B, B\setminus A} \subseteq \mathfrak{M}_0\end{aligned}$$

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 $\mathcal{U} \in \mathfrak{M}_{A_1,\dots,A_m}$ and $\mathcal{V} \in \mathfrak{M}_{B_1,\dots,B_n}$.

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 57, 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 30. We say that a sequence of finite measures $\mu_n \in \mathcal{M}_f$ converges *weakly* to $\mu \in \mathcal{M}_f$ (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 point $n \in \mathbb{N}$. Then μ_n converges vaguely to zero measure but it does not converge weakly.

Definition 31. The *Prochorov distance* between two finite measures $\mu, \nu \in \mathcal{M}_f$ 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 measure 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 $\mathcal{M}_f$ and $\widehat{\varrho}_P$ defines a metric on $\mathcal{M}$ (see Exercise class).

Proposition 21.

a) *The space $(\mathcal{M}_f, \varrho_P)$ is a complete separable metric space and the metric generates the weak convergence of measures:*

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

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:*

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

Proof: [12], Theorem 2.2.

3.2 Random measures

Definition 32. 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 22. *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 22 remains true if we consider only Borel sets B that form a system $\mathcal{S}$ from Lemma 18.

Definition 33. 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 ([9], Theorem 8.5). An intensity measure does not have to be locally finite. There could exist $B \in \mathcal{B}_0$ such that $\Psi(B) < \infty$ but $\Lambda(B) = \mathbb{E}\Psi(B) = \infty$.

The distribution of a random measure is uniquely determined by the finite-dimensional projections.

Theorem 23. *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 random vectors $(\Psi_1(A_1), \dots, \Psi_1(A_n))^{\mathrm{T}}$ and $(\Psi_2(A_1), \dots, \Psi_2(A_n))^{\mathrm{T}}$ have equal distributions for any $n \in \mathbb{N}$ and $A_1, \dots, A_n \in \mathcal{S}$ 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 algebra $\mathfrak{M}_0$, which is the π-system and thus Theorem 58 states that Ψ_1 and Ψ_2 have the same distribution on $\sigma\mathfrak{M}_0$, which is equal to $\mathfrak{M}$ by Lemma 20.

□

Proposition 24. *Let $\{\xi_A, A \in \mathcal{B}_0\}$ be a family of non-negative and a.s. finite-valued random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying:*

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

Then there exists a random measure Ψ such that $\Psi(A) \overset{a.s.}{=} \xi_A$ for all $A \in \mathcal{B}_0$.

Proof: It is a slight generalization of the theorem stating the existence of a regular version of conditional expectation ([15], Theorem VI.1.21). The complete proof can be found in [5], 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 25. *(existence of random measure with given finite-dimensional projections) Let a probability measure $Q_{B_1,\dots,B_n}$ on $([0,\infty)^n, \mathcal{B}([0,\infty)^n))$ be given for any $n \in \mathbb{N}$ and $B_1, \dots, B_n \in \mathcal{B}_0$ pairwise disjoint. Assume that the following conditions are satisfied:*

1. *(projectivity) $Q_{B_1,\dots,B_n}(\cdot) = Q_{B_1,\dots,B_n,B_{n+1}}((\Pi_n^{n+1})^{-1}(\cdot))$, where $\Pi_n^{n+1} : [0,\infty)^{n+1} \to [0,\infty)^n$ is a canonical projection $(u_1, \dots, u_n, u_{n+1})^{\mathrm{T}} \mapsto (u_1, \dots, u_n)^{\mathrm{T}}$,*
2. *(symmetry) $Q_{B_1,\dots,B_n}(U_1 \times \cdots \times U_n) = Q_{B_{\pi(1)},\dots,B_{\pi(n)}}(U_{\pi(1)} \times \cdots \times U_{\pi(n)})$ for every permutation π of the set $\{1, \dots, n\}$,*
3. *(additivity) $Q_{A\cup B}(U) = Q_{A,B}(\{(u,v) \in [0,\infty)^2 : u + v \in U\})$ for $A \cap B = \emptyset$,*

4. (continuity in $\emptyset$) $A_n \in \mathcal{B}_0$, $A_n \searrow \emptyset \Rightarrow Q_{A_n} \xrightarrow[n\to\infty]{w} \delta_0$.

Then there exists a unique probability measure Q on $(\mathcal{M}, \mathfrak{M})$ such that $Q_{B_1,\dots,B_n}$ are the finite-dimensional distributions of Q.

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

$$Q_{A_1,\dots,A_n}(\mathcal{U}) = Q_{B_1,\dots,B_k}(\phi^{-1}\mathcal{U}), \quad \mathcal{U} \in \mathcal{B}([0,\infty)^n),$$

where

$$\phi : (x_1, \dots, x_k) \mapsto \left(\sum_{i\in I_1} x_i, \dots, \sum_{i\in I_n} x_i \right).$$

The correctness of the definition follows from the following fact: if A_i, B_j, I_i and the mapping ϕ are as above and if furthermore $A_1, \dots, A_n \in \mathcal{B}_0$ are pairwise disjoint, then

$$Q_{A_1,\dots,A_n}(\mathcal{U}) = Q_{B_1,\dots,B_k}(\phi^{-1}\mathcal{U}), \quad \mathcal{U} \in \mathcal{B}([0,\infty)^n).$$

This can be proved by induction on n using the assumed properties (projectivity, additivity and symmetry).

Now we verify that the distributions $\{Q_{A_1,\dots,A_n} : A_1, \dots, A_n \in \mathcal{B}_0, n \in \mathbb{N}\}$ form a consistency system. Let $C_1, \dots, C_l$ be a disjoint decomposition of $A_1, \dots, A_{n+1}$ such that $C_i \subseteq A_1 \cup \dots \cup A_n$ for $i = 1, \dots, l-1$ (i.e. $C_l = A_{n+1} \setminus \cup_{i=1}^n A_i$) and $A_i = \cup_{j\in I_i} C_j$, $i = 1, \dots, n+1$ (thus $l \in I_{n+1}$). If we define the mappings

$$\phi : (x_1, \dots, x_{l-1}) \mapsto \left(\sum_{i\in I_1} x_i, \dots, \sum_{i\in I_n} x_i \right)$$

and

$$\psi : (x_1, \dots, x_l) \mapsto \left(\sum_{i\in I_1} x_i, \dots, \sum_{i\in I_n} x_i, \sum_{i\in I_{n+1}} x_i \right),$$

then $\phi \circ \Pi_{l-1}^l = \Pi_n^{n+1} \circ \psi$. From the commutativity of the diagram

$$\begin{array}{ccc} Q_{C_1,\dots,C_l} & \xrightarrow{\Pi_{l-1}^l} & Q_{C_1,\dots,C_{l-1}} \\ \downarrow \psi & & \downarrow \phi \\ Q_{A_1,\dots,A_{n+1}} & \xrightarrow{\Pi_n^{n+1}} & Q_{A_1,\dots,A_n} \end{array}$$

the required projectivity follows:

$$Q_{A_1,\dots,A_n}(\cdot) = Q_{A_1,\dots,A_{n+1}}((\Pi_n^{n+1})^{-1}(\cdot)).$$

Therefore, we can apply the Daniell-Kolmogorov existence theorem. It assures the existence of a probability measure Q_0 on $([0,\infty)^{\mathcal{B}_0}, \mathcal{B}([0,\infty)^{\mathcal{B}_0})$ such that $Q_{A_1,\dots,A_n}$, for $A_1, \dots, A_n \in \mathcal{B}_0$, ara the finite-dimensional distributions of Q_0. Let $\{\xi_A : A \in \mathcal{B}_0\}$ be a family of random variables that has the distribution Q_0. The conditions 3. and 4. guarantee that the assumptions of Proposition 24 are satisfied. The fourth condition $Q_{A_n} \xrightarrow[n\to\infty]{w} \delta_0$ means that $\xi_{A_n} \xrightarrow[n\to\infty]{\mathcal{D}} 0$, which is equivalent to $\xi_{A_n} \xrightarrow[n\to\infty]{P} 0$, and hence $\xi_{A_{n_k}} \xrightarrow[k\to\infty]{a.s.} 0$ for some subsequence n_k. This already implies $\xi_{A_n} \xrightarrow[n\to\infty]{} 0$ because the sequence is monotone a.s. Therefore, by Proposition 24, there exists a random measure Ψ and its distribution Q on $(\mathcal{M}, \mathfrak{M})$ is the desired probability measure. It is unique by Theorem 23. This completes the proof.

□

3.3 Simple point processes

Definition 34. Let $\mathcal{N}^* = \{\nu \in \mathcal{N} : \nu(\{x\}) \le 1 \text{ for every } x \in E\}$. 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 35. 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 26. *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 $A_1^1 = \tilde{A}_1$, $A_2^1 = \tilde{A}_2 \setminus \tilde{A}_1$, $A_3^1 = \tilde{A}_3 \setminus (\tilde{A}_1 \cup \tilde{A}_2)$, ... 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 non-atomic 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 $\varepsilon(x, n) \le 1/2n$ 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 27. *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

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

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

□

Now we are ready to define a simple point process.

Definition 36. 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 $\mathcal{N}^*$.

Definition 37. 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. The following lemma states that the atoms may be enumerated in a measurable way.

Lemma 28. *There exist measurable mappings $\zeta_i : \mathcal{N} \to E$ such that*

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

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

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 $x \in A_{j_k(x)}^k$, $k \in \mathbb{N}$, 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 $\cup_{i=1}^{j_1(x)} A_i^1$. 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 η, 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

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

which shows the measurability of ζ_p.

□

For a simple measure $\nu \in \mathcal{N}^*$ 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 29. *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 18 we can take $\mathcal{S} = \mathcal{K} \subseteq \mathcal{B}_0$ 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 $\{\{A_1^n, A_2^n, \ldots\}, n \in \mathbb{N}\}$ we can write

$$\mathcal{N}_{K,r}^* = \bigcup_{n_0\in\mathbb{N}} \bigcap_{n\geq n_0} \bigcup_{i_1,\ldots,i_m} \left[(\mathcal{N}_{K\cap\bar{A}_{i_1}^n,0}^*)^c \cap (\mathcal{N}_{K\cap\bar{A}_{i_2}^n,0}^*)^c \cap \cdots \cap (\mathcal{N}_{K\cap\bar{A}_{i_m}^n,0}^*)^c \cap \mathcal{N}_{K\setminus(\bar{A}_{i_1}^n\cup\cdots\cup\bar{A}_{i_m}^n),0}^*\right],$$

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 30. *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 38. 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 the void probabilities.

Theorem 31. *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 $\Phi_1 \stackrel{\mathcal{D}}{=} \Phi_2$.*

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 29). Now it suffices to apply Theorem 58.

□

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 intervals. Once 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 39. A set $A \subseteq E$ is called *locally finite* when $A \cap B$ is a finite set for every $B \in \mathcal{B}_0(E)$. Obviously, every locally finite set is closed. Denote $\mathcal{F}_{lf} = \{A \in \mathcal{F} : A \text{ locally finite}\}$ the family of all locally finite sets.

Definition 40. The *support* of a locally finite measure $\mu \in \mathcal{M}(E)$ is defined as the smallest closed subset A of E such that $\nu(E \setminus A) = 0$. It is denoted by $\operatorname{supp}\nu$ and it can be written as

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

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

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

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

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

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

□

Definition 42. A *random locally finite set* in E is a measurable mapping $\Xi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{F}_{lf}, \mathfrak{F}_{lf})$, where $\mathfrak{F}_{lf} = \{\mathcal{U} \cap \mathcal{F}_{lf} : \mathcal{U} \in \mathfrak{F}\}$ is the trace of the σ-algebra $\mathfrak{F}$ on $\mathcal{F}_{lf}$.

Lemma 33. *The map $i : \nu \mapsto \operatorname{supp} \nu$ is a bijection between $(\mathcal{N}^*, \mathfrak{N}^*)$ and $(\mathcal{F}_{lf}, \mathfrak{F}_{lf})$. 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 $i^{-1}(\mathcal{F}^K \cap \mathcal{F}_{lf}) = \{\nu \in \mathcal{N}^* : \operatorname{supp} \nu \cap K = \emptyset\} = \{\nu \in \mathcal{N}^* : \nu(K) = 0\} \in \mathfrak{N}^*$.

Corollary 34. *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 35. *(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 are also coincide on $\mathfrak{F}$ by Theorem 58. 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 31 is a consequence of Theorem 35 and Lemma 33.

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 22, δ_X is indeed a point process.

Definition 43. 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 identically distributed (ν-uniformly) 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 *a binomial point process* of n points in B according to ν.

Remark 22. The measurability of $\Phi^{(n)}$ is clear (it is 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.

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

Definition 44. Let ν be a diffuse 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 $\Phi = 0$.

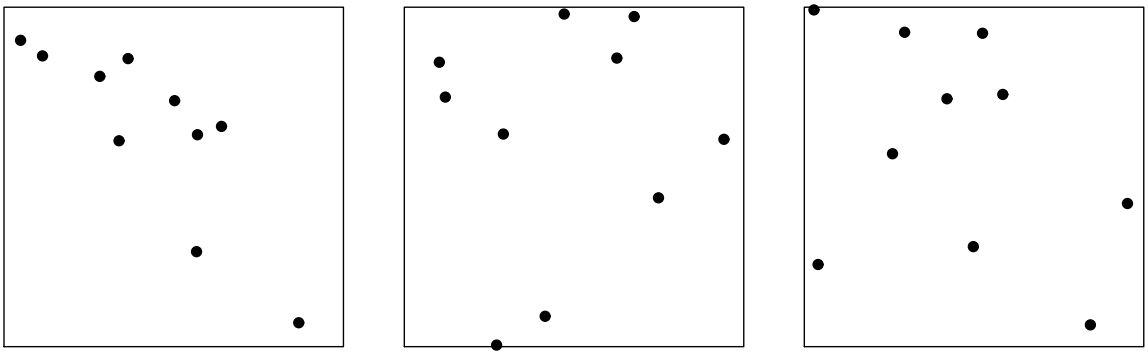

Figure 10. Three different realizations of a binomial point process of 10 uniformly distributed points in the unit square window.

The Poisson point process serves as 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 occuring 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 45. 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.

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

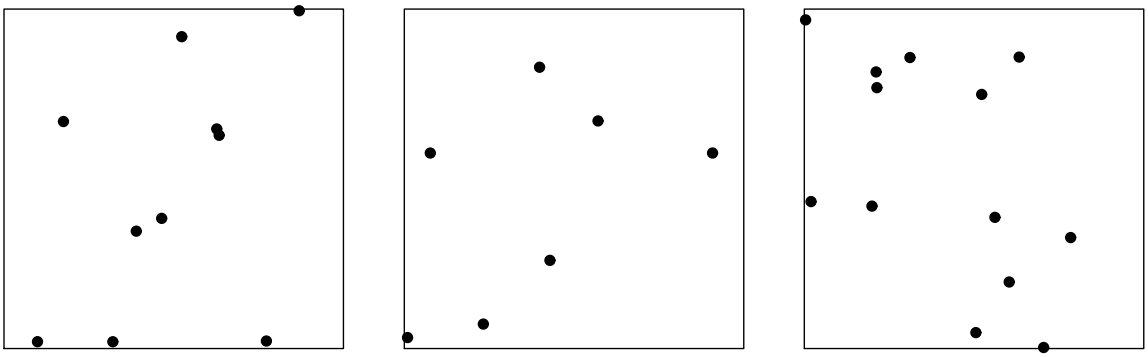

Figure 11. Three different realizations of a Poisson point process in the unit square window. The intensity measure is a multiple of the Lebesgue measure. The expected number of points in the window is 10. The actual observed numbers are 10, 7 and 13.

The uniqueness and existence is ensured by Theorem 25.

Corollary 36. *Let Λ be a locally finite measure on E. Then there exists the Poisson point process with intensity measure Λ and its distribution is uniquely determined.*

Proof: For $B_1, \ldots, B_n \in \mathcal{B}_0$ pairwise disjoint the finite dimensional projections have by definition the distribution $Q_{B_1,\ldots,B_n} = P_{B_1} \times \cdots \times P_{B_n}$, where P_{B_i} is the Poisson distribution with parameter $\Lambda(B_i)$. It suffices to use Theorem 25.

□

Lemma 37. *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 $B_1^{(k)}, \ldots, B_k^{(k)} \in \mathcal{B}$ such that $K = \cup_{i=1}^k B_i^{(k)}$ and $\Lambda(B_i^{(k)}) = \varepsilon/k$, $i = 1, \ldots, k$. There must be a number $i \in \{1, \ldots, k\}$ 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 37 and Theorem 31, 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 38. *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 31.

□

Theorem 38 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).

Definition 46. Denote by P_Λ the distribution of the Poisson point process Φ_Λ on E with locally finite intensity measure $\Lambda \in \mathcal{M}$. Let Ψ be a random diffuse measure on E with distribution Q_Ψ. The *Cox point process* Φ_Ψ with *driving random measure* Ψ has distribution given as the mixture $Q(\cdot) = \int_{\mathcal{M}} P_\Lambda(\cdot)\, Q_\Psi(\mathrm{d}\Lambda)$.

Remark 23. The previous definition means that conditionally on $\Psi = \Lambda$, Φ_Ψ is the Poisson point process with intensity measure Λ. Therefore, Φ_Ψ is sometimes also called *doubly stochastic Poisson process*. We have assumed that Ψ is diffuse to ensure that Φ_Ψ is simple.

Lemma 39. *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

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

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

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

□

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 47. Consider a deterministic measure $\Lambda \in \mathcal{M}$. Let $P_{t\Lambda}$ be the distribution of a Poisson point process with intensity measure $t\Lambda$. Let Y be a non-negative random variable with distribution R. A point process with the distribution $Q = \int_0^\infty P_{t\Lambda}\, R(\mathrm{d}t)$ 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 39, the intensity measure of the mixed Poisson point process is $\Lambda(\cdot)\mathbb{E}Y$.

3.5 Moment measures

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

Definition 48. The *n-th order moment measure* of the 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,

$$M^{(n)}(A_1 \times \cdots \times A_n) = \mathbb{E}\Psi(A_1) \cdots \Psi(A_n), \quad A_1, \ldots, A_n \in \mathcal{B}.$$

Remark 24. The measure $M^{(n)}$ is in fact the intensity measure of the random measure Ψ^n on E^n. For the point process Φ on E, Φ^n is the point process on $\mathbb{E}^n$ 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 $\mathcal{B}^n$ in $E^{[n]}$ will be denoted by $\mathcal{B}^{[n]}$. For $\mu \in \mathcal{M}$ let $\mu^{[n]} = \mu_n|_{E^{[n]}}$.

Definition 49. The *n-th order factorial moment measure* of the random measure Ψ is defined as

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

Remark 25. First order moment measures coincide with the intensity measure: $M^{(1)} = \alpha^{(1)} = \Lambda$. The relation between the n-th order moment measure and the moments of the numbers of points is the following:

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

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.

Let Φ be a simple point process. Then

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

and

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

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

Theorem 40. *Let Φ be a Poisson point process with diffuse intensity measure Λ. 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 45) 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 58 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 26. The factorial moment measure of the Cox point process with driving measure Ψ is $\alpha^{(n)} = \mathbb{E}\Psi^n$.

The following theorem will be often very useful.

Theorem 41. *(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 48. The rest of the proof proceeds by a standard argument of measure theory.

□

Definition 50. The *Laplace transform* of a random measure Ψ is the functional L_Ψ defined by the relation

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

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

Remark 27. In particular, if we choose $f = \sum_{i=1}^n t_i \mathbf{1}_{B_i}$ to be a simple function, then

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

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 42. *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 23.

□

Lemma 43.

(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 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 $f = \sum_{i=1}^n t_i \mathbf{1}_{B_i}$, where $B_i \in \mathcal{B}$ are pairwise disjoint, 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

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

where Q_t is the distribution of the Poisson point process with intensity measure $t\Lambda$ and R is the distribution of Y.

□

3.6 Palm distribution

Definition 51. 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 44. *(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 $\mu(A \times B) \leq \nu(A)$. 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 $\mu(A \times B) \leq \nu(A)$). 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 24. The addivity $\xi_{B \cup B'} = \xi_B + \xi_{B'}$ 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 $\xi_{B_1} \geq \cdots \geq \xi_{B_n} \geq \cdots$, they converge also a.s. Now by Proposition 24 we obtain a random measure Ψ on T such that $\Psi(B) = \xi_B$ a.s. for every $B \in \mathcal{T}$. Morevover, 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

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

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 52. 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 28. From

$$C(B \times \mathcal{U}) = \int_{\mathcal{M}} \int_E \mathbf{1}_B(x)\mathbf{1}_{\mathcal{U}}(\mu)\, \mu(\mathrm{d}x)\, Q(\mathrm{d}\mu),$$

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 f on $E\times\mathcal{M}$.

Corollary 45. *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 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 21), we can use Theorem 44 with $S=E$, $T=\mathcal{M}$, $\nu=\Lambda$ and $\mu=C$.

□

Definition 53. If P is the Markov kernel from Corollary 45, then the distribution $P_x(\cdot)=P(x,\cdot)$ is called *Palm distribution* of the random measure Ψ at point $x\in E$.

Remark 29. It actually does not make sense to speak about the Palm distribution at one particular point because this could be defined arbitrarily. Nevertheless, Corollary 45 assures that the family $\{P_x : x\in E\}$ of Palm distributions is uniquely determined for Λ-a.a. x. If $\{P_x\}$ and $\{\tilde{P}_x\}$ would be 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 46. *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 $f(x,\mu)=\mathbf{1}_A(x)\mathbf{1}_{[\mu(\{x\})\geq 1]}$ for arbitary $A\in\mathcal{B}_0$. 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 the 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 30. 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 54. 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 arbitrary non-negative measurable function g.

Theorem 47. *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 31. 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

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

where $b(x,\varepsilon)$ denotes the ball of center x and radius ε. Lemma 7.2 in [12] 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 32. 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 48. *(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 theorey arguments. For $h(x,\nu) = \mathbf{1}_A(x)\mathbf{1}_{\mathcal{U}}(\nu)$ we have

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

The second equation follows from Corollary 45. The second relation can be deduced from the first one and Definition 54.

□

Theorem 49. *(Slivnyak theorem) Let Φ be a Poisson point process having distribution Π and intensity measure $\Lambda \in \mathcal{M}$. Then $P_x = \Pi * \delta_{(\delta_x)}$, i.e. $P_x^! = \Pi$, 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 [5], Proposition 13.1.VII or [12], Theorem 7.3. Here we show the proof for diffuse Λ. In this case, Lemma 37 states that $\Phi + \delta_x$ is a simple point process. Denote by $\tilde{\Phi}_x$ a simple point process with distribution P_x. According to Theorem 31 it suffices to verify that $\Phi + \delta_x$ and $\tilde{\Phi}_x$ have the same void probabilities. Let $A \in \mathcal{B}_0$ and $K \in \mathcal{K}$ be arbitrary. Recall that $\mathcal{N}_{K,0} = \{\nu : \nu(K) = 0\}$. We show that

$$\int_A P_x(\mathcal{N}_{K,0})\, \Lambda(\mathrm{d}x) = \int_A \mathbb{P}((\Phi + \delta_x)(K) = 0)\, \Lambda(\mathrm{d}x).$$

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

$$\begin{aligned}\int_A \mathbb{P}(\Phi(K) = 0, x \notin K)\, \Lambda(\mathrm{d}x) &= \int_{A \setminus K} \int_{\mathcal{N}} \mathbf{1}_{[\nu(K)=0]}\, \Pi(\mathrm{d}\nu)\, \Lambda(\mathrm{d}x) \\ &= \Lambda(A \setminus K)\mathbb{P}(\Phi(K) = 0) = \mathbb{E}\Phi(A \setminus K)\mathbf{1}_{[\Phi(K)=0]},\end{aligned}$$

where in the last step we have used the independence of $\Phi(A \setminus K)$ and $\Phi(K)$. Clearly, $\mathbb{E}\Phi(A)\mathbf{1}_{[\Phi(K)=0]} = \mathbb{E}\Phi(A \setminus K)\mathbf{1}_{[\Phi(K)=0]}$, 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 $\mathcal{B}^d$. A point process on $\mathbb{R}^d$ will be referred to as the *spatial point process*.

4.1 Stationary spatial point processes

Definition 55. 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 33. If $\mu = \sum_{i=1}^{\tau} \delta_{x_i}$, then $t_z\mu = \sum_{i=1}^{\tau} \delta_{x_i+z}$.

Definition 56. 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 57. A random measure is called *motion-invariant* if it is both stationary and isotropic.

Lemma 50. *If Ψ is a 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 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 measure on $(\mathbb{R}^d, \mathcal{B}^d)$, see Exercise 26.6 in [9].

□

Definition 58. The multiple $\lambda \geq 0$ from Lemma 50 is called *intensity* of a 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 59. 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 50 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 60. 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 34. Every homogeneous Poisson point process is both motion-invariant (see Exercise class).

Definition 61. 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 35. The first order product density coincides with the intensity function and we write $\lambda^{(1)} = \lambda$.

Remark 36. A heuristic interpretation of the n-th order product density of a spatial point process is the following. Consider n infinitesimally small disjoint balls with centres $x_1, \ldots, x_n$ and volumes $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$. Then $\lambda^{(n)}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$ is the probability that each of these ball contains a point of the process.

Corollary 51. *If there exists the intensity function λ of the Poisson point process Φ, then its n-th order product density satistfies $\lambda^{(n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \lambda(x_i)$, $x_1, \ldots, x_n \in \mathbb{R}^d$.*

Proof: The assertion follows from Theorem 40.

□

From the definition of stationarity (Definition 55) it follows that the moment measures of a stationary point process are invariant under diagonal shifts, i.e.

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

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 52. *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 50). 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 52. 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 53. *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 37. The expectation with respect to P_o will be denoted by the symbol $\mathbb{E}_o$. Similarly, $\mathbb{E}_o^!$ stands for the expectation with respect to $P_o^!$. It means that

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

5. Appendix

5.1 Gaussian distribution

Theorem 54. *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].

□

Theorem 55. *Let $X_1, \dots, X_n$ be independent identically distributed random variables with standard Gaussian 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.

□

5.2 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

$$I_{1/2}(z) = \sqrt{\frac{2}{\pi z}} \sinh z, \quad K_{1/2}(z) = \sqrt{\frac{\pi}{2}} \mathrm{e}^{-z} z^{-1/2}.$$

5.3 Measure theory

Let E be a given set. We recall some basic definitions of families of certain subsets of E.

Definition 62. 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 63. 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 64. A collection $\mathcal{D}$ is called a *Dynkin system* or *λ-system* if
(i) $\emptyset \in \mathcal{D}$, $E \in \mathcal{D}$,
(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 38. 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 a σ-algebra.

We often use the following result.

Theorem 56. *(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 [7].

Definition 65. 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 39. A Dynkin system is a monotone system.

Theorem 57. *(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 [9].

Theorem 58. *(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 [7] or Lemma 2.2 in [12].

□

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] 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.

[5] 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.

[6] E. Ising (1925): Beitrag zur Theorie des Ferromagnetismus, *Z. Physik* **31**, 253–258.

[7] O. Kallenberg (2002): *Foundations of Modern Probability*, 2nd edition, Springer, New York.

[8] I. Karatzas and S. E. Shreve (1988): *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

[9] J. Lukeš and J. Malý (2005): *Measure and Integral*, Matfyzpress, Prague.

[10] L. Onsager (1944): Crystal statistics. I. A two-dimensional model with an order-disorder transition, *Phys. Rev.* **65**, 117–149.

[11] Z. Prášková (2004): *Základy náhodných procesů II*, in Czech, Karolinum, Praha.

[12] J. Rataj (2006): *Bodové procesy*, in Czech, Karolinum, Praha.

[13] H. Rue and L. Held (2005): *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC, Boca Raton.

[14] Z. Sasvári (1994): *Positive Definite and Definitizable Functions*, Akademie Verlag, Berlin.

[15] J. Štěpán (1987): *Teorie pravděpodobnosti – matematické základy*, in Czech, Academia, Praha.

[16] P. Whittle (1954): On stationary processes in the plane, *Biometrika* **41**, 434–449.