**FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University**

MASTER THESIS

Noemi Kuželová

Multivariate Cox point processes

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	RNDr. Jiří Dvořák, Ph.D.
Study programme:	Mathematics
Study branch:	Probability, Mathematical Statistics and Econometrics

Prague 2022

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

Dedication. I would like to express my sincere gratitude to my patient and supportive supervisor RNDr. Jiří Dvořák, Ph.D. for his guidance and ideas that made this work happen, for the time and energy he devoted to me and especially for his friendly approach, thanks to which I could turn to him for help and advice at any time without worries. I would also like to thank my family and friends for their continued support. Special thanks to Jana, who has received the greatest number of my complaints and has never hesitated to listen to me.

Title: Multivariate Cox point processes

Author: Noemi Kuželová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jiří Dvořák, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The Log-Gaussian Cox process is an important example of the use of spatial modeling and spatial statistics in practice. It is useful for describing many real-world situations, from modeling tree growth in the rainforests, to trying to understand the occurrence of precipitation and earthquakes, to examining the expansion of the Greenland seal population. In this work we focus mainly on the multivariate form of this point process. Specially in such form that allows to describe at the same time inhomogeneity, clustering and environmental effects in the investigated system. When the parameters of multivariate LGCP process are estimated, the minimum contrast method is usually used. However, we investigate the possibility of using composite likelihood estimation instead. We consider the composite likelihood criterion as a limit of the likelihoods in approximating discrete models. We compare it with an established approach of constructing the composite likelihood based on multiplication of likelihoods for pairs of points.

Keywords: Log-Gaussian Cox process; Multivariate point process; Composite likelihood

Contents

Introduction

We find the inspiration for this thesis in the article Waagepetersen et al. [2016], which aims to explore the data describing the location of trees in Lansing Woods. The authors use the multivariate log-Gaussian Cox process for this. In order to capture as many influences as possible that affect the growth of trees in this area, the intensity function of these processes consists of a deterministic function that describes the inhomogeneity in the area and also two independent Gaussian fields. One of them explains clustering within one tree species and the other describes environmental variables that affect all trees, but this effect may vary from species to species. This article further discusses how to estimate the parameters in this model choosing the least squares method. At this point, we decided to take a different approach to the article. However, instead of the minimum contrast estimator we exploit the composite likelihood method.

The article inspired some further work. Before we focus in more detail on what this thesis offers, let us mention some of them. For example, the article Jalilian et al. [2020] performs a detailed analysis of the multivariate log-Gaussian Cox process also on data on the occurrence of different tree species, but this time in the Hyrcanian Forest. A very interesting application is offered in an article Jullum et al. [2020], which aims to estimate the annual production of seal pups in the Greenland Sea. The authors use the multivariate log-Gaussian Cox process to describe the occurrence of pups in aerial photographs. The paper Hessellund et al. [2020] assumes a semiparametric model for multidimensional intensity functions containing an unspecified complex factor common to all types of points. It is used to create their own composite likelihood function, which the authors estimate the parameters of the second-order for log-Gaussian Cox process. If we want to focus on how to find numerically stable and efficient estimates, especially for highly multivariate log-Gaussian Cox processes, we can start with the article Choiruddin et al. [2019]. On a bit another note, article Mateu and Jalilian [2022] proposes to examine point processes not using classical statistical analysis, but to involve neural networks. Like the others, it also suggests their use for the multivariate log-Gaussian process presented here.

Coming back to the content of this thesis. In the first chapter we build a basic theory of point processes. We introduce their first- and second-order characteristics and state a few of their basic relations. In addition to single-species processes, we do the same for multivariate point processes. Next, we introduce the Poisson point process, and this chapter culminates in the definition of the Cox point process. It is the basis of the processes we want to focus on in this thesis.

Chapter 2 deals with the univariate log-Gaussian Cox process. It presents its definition and basic characteristics, specifically what its intensity function, second-order product density and also the K-function look like.

In Chapter 3, we establish a multivariate log-Gaussian Cox process. That is exactly the one from the already mentioned Waagepetersen et al. [2016] article. We will carefully define this process here and derive in detail what the species-specific and inter-species covariance structure looks like. And also the resulting intensity functions, second-order product density, cross pair correlation function and last but not least the K-function and cross K-function.

In Chapter 4, we begin to consider the appropriate composite likelihood function by which we can estimate the parameters of the log-Gaussian Cox process, whether in the univariate or multivariate case. We are inspired by Waagepetersen [2007], whose idea is to divide the observation window into individual cells and approximate the probability of whether a point occurs in the cell by the intensity function multiplied by the cell area. Subsequently, we refine such partition to the limit until we get the expression of the composite likelihood function. However, since we are not satisfied only with the characteristics of the first order, we develop this idea to examine the probability of occurrence of a pair of points in cell pairs and its subsequent approximation using the second-order product density. We elaborate on both of these cases in detail. Next, we derive how to develop this method so that we can apply it to a multivariate process. We also use the Dvořák and Prokešová [2012] article, which advise us on how to avoid the numerically too demanding integrations that would await us if we left the composite likelihood function in its basic form and did not modify it further. We show what the respective composite likelihood functions look like for both univariate and multivariate log-Gassian Cox process.

Chapter 5 then introduces the popular form of composite likelihood function from the Guan [2006] article. We do not discuss it in as much detail as the composite likelihood function derived in the previous chapter. We present only its basic idea and form in which we can use it for univariate and multivariate log-Gaussian Cox process.

The last Chapter 6 is devoted to a simulation study. On the simulated data we examine and compare how well both composite likelihood functions from Chapters 4 and 5 estimate the parameters of the models. We deal with both univariate and multivariate log-Gaussian Cox processes.

1. Basic definitions

In the very beginning of this thesis, we need to set basic definitions. Next few sections will deal with introduction of descriptive characteristics of points processes and Cox point process. However, it is a long way to work on the Cox process, which is at the heart of this work.

1.1 Spatial point process

First of all, we need to define a point process. There are two ways to look at it. The first one, which [Møller and Waagepetersen, 2004, Chapter 2] state, gives us a nice intuitive idea of what to imagine under this term. We can think about spatial point process X as a random countable subset of space E, provided that $E \subset \mathbb{R}^d$.

The second way of looking at the point process is more robust and can be a bit more difficult to get oriented in. However, it allows us to move in the world of point processes with mathematical accuracy. That is why we will mention it as well. To do this, however, we must first introduce the concept of the space of all locally finite measures. This and other definitions in this subsection are based on [Cressie, 1993, Section 8.3].

Definition 1. *Let (E, ρ) be a separable metric space where every closed bounded set is compact. Further, let μ be a measure on $(E, \mathcal{B})$, where $\mathcal{B} = \mathcal{B}(E)$ stands for Borel sets on E. Then μ is called locally finite measure if $\mu(K) < \infty$ for every $K \in \mathcal{B}$ compact.*
Additionally, we define set of all locally finite measures on $(E, \mathcal{B})$:

$$\mathcal{M}(E) = \mathcal{M} = \{\mu \text{ measure on } (E, \mathcal{B});\ \mu \text{ is locally finite}\}.$$

In the theses, we will consider $E = \mathbb{R}^d$ and ρ an the Euclidean metric on $\mathbb{R}^d$. In addition, we will focus on counting processes and therefore introduce the appropriate set for the locally finite measures.

Definition 2. *Let M be set of all locally finite measures on $(E, \mathcal{B})$. Then we denote set of all locally finite counting measures on $(E, \mathcal{B})$ as following*

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

The next step is to assign suitable σ-algebras to sets $\mathcal{M}$ and $\mathcal{N}$. The one-dimensional projection will help us with that. For a fixed set $B \in \mathcal{B}$, we define it as follows.

$$\begin{aligned} \pi_B : \mathcal{M} &\longrightarrow [0, \infty] \\ \mu &\longmapsto \mu(B) \end{aligned}$$

Taking the smallest σ-algebra of images of all such projections that are additionally measurable, we will get the coveted σ-algebra on the space $\mathcal{M}$:

$$\mathfrak{M} = \sigma\{\pi_B^{-1}(A);\ B \in \mathcal{B},\ A \in [0, \infty]\}.$$

Since $\mathcal{N}$ is clearly a subset of $\mathcal{M}$, we can define its σ-algebra as a trace of $\mathfrak{M}$:

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

With the correctness of this definition we can refer to [Cressie, 1993, Section 8.3].

Now we can finally define a point process and also its generalization a random measure.

Definition 3 (Random measure, point process)**.** *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space. Then, a measurable mapping* $\Psi : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathcal{M}, \mathfrak{M})$ *is called a random measure.*
Furthermore, we call a measurable mapping $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathcal{N}, \mathfrak{N})$ *a point process.*

To work with variously distributed random measures, we introduce their probability distribution.

Definition 4 (Probability distribution)**.** *Let* $\Psi : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathcal{M}, \mathfrak{M})$ *be a random measure. Then its probability distribution is a probability measure* Q *on* $(\mathcal{M}, \mathfrak{M})$ *given by formula*

$$Q(\mathcal{U}) = \mathbb{P}\left(\omega \in \Omega;\ \Psi(\omega) \in \mathcal{U}\right),\ \mathcal{U} \in \mathfrak{M}.$$

As is customary in spatial statistics, we will use abbreviated notation $\Psi(B)$ instead of $\Psi(\omega)(B)$.

We should realize $\forall\ B \in \mathcal{B} :\ \Psi(B)$ is a random variable if and only if Ψ is a random measure.

1.2 Marked point process

In the next step, we would like to assign different marks to the individual points of the point process. It is conceivable that we would assign different marks to trees randomly growing in the forest. For example, an assigned tag can represent the height of a tree or its species. In this way, we get the so-called marked point process. We will formally introduce it using the following definition.

Here, as in the rest of the work, we take advantage of the point of view which treats X as a random set. We will denote individual events of the point process X by $x \in X$.

Definition 5 (Marked point process)**.** *Let* X *be a point process on* $E \subset \mathbb{R}^d$*. Let* $\mathbb{M}$ *be a complete separable locally compact metric space such that for each* x *in* X*, there is a random variable* $m_x \in \mathbb{M}$*. Then* $X_M = \{(x, m_x); x \in X\}$ *is called marked point process with points in* E *and mark space* $\mathbb{M}$*. Further, the elements* m_x *are called marks.*

If marked point process X_M has points of k different types, i.e. its marked space is $\mathbb{M} = \{1, \dots, k\}$, $k \in \mathbb{N}$, we speak about *multitype point process.* We can equivalently look at it as k-tuple $(X_1, \dots, X_k)$ of point process $X_1, \dots, X_k$ corresponding to types of points $1, \dots, k$. Then we speak about k-dimensional *multivariate point process.*

1.3 Summary statistics for point processes

In this section, let X be a point process on $E = \mathbb{R}^d$. We will denote the corresponding Borel sigma algebra by $\mathcal{B}^d = \mathcal{B}(\mathbb{R}^d)$.

The first introduced characteristics will be intensity measure (of a point process). It indicates the number of process points that can be expected to be observed in a given set.

Definition 6 (Intensity measure)**.** *Let X be a point process on $\mathbb{R}^d$. We define intensity measure Λ by relation*

$$\Lambda(B) = \mathbb{E}N_X(B), \ \ B \in \mathcal{B}^d,$$

where $N_X(B)$ denotes random variable defined as count of points of X on B, i.e. $N_X(B) = \sum_{\boldsymbol{u} \in X} \mathbf{1}_{[\boldsymbol{u} \in B]}$.

If it exists, we can define Radon–Nikodym derivative with respect to Lebesgue measure for the intensity measure. In this context, it is called the intensity function.

Definition 7 (Intensity function)**.** *Let X be a point process on $\mathbb{R}^d$ with intensity measure Λ. If there exist a non-negative function ρ such that $\Lambda(B) = \int_B \rho(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$ for each $B \in \mathcal{B}^d$, then ρ is called intensity function.*

If ρ is constant, then X is *homogeneous* (or *first-order stationary*) point process with *intensity* ρ. Otherwise, X is called *inhomogeneous* point process.

For a homogeneous point process, we can interpret intensity ρ as mean number of points per unit volume.

Similarly to intensity measure, we can also define the basic characteristics of the second order called second-order factorial moment measure. It indicates the expected number of process point pairs in a certain set.

Definition 8 (second-order factorial moment measure)**.** *Let X be a point process on $\mathbb{R}^d$. Then the following formula defines the second-order factorial moment measure*

$$\alpha^{(2)}(C) = \mathbb{E} \sum_{\boldsymbol{u},\boldsymbol{v} \in X}^{\neq} \mathbf{1}_{[(\boldsymbol{u},\boldsymbol{v}) \in C]}, \ \ C \subset \mathcal{B}\left(\mathbb{R}^d \times \mathbb{R}^d\right).$$

Let us define a density of $\alpha^{(2)}$.

Definition 9 (second-order product density)**.** *Let X be a point process on $\mathbb{R}^d$ with second-order factorial moment measure $\alpha^{(2)}(C)$. If there exists a non-negative function $\rho^{(2)}$ for which it holds $\alpha^{(2)}(C) = \int\int \mathbf{1}_{[(\boldsymbol{u},\boldsymbol{v}) \in C]} \rho^{(2)}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}$, $C \subset \mathcal{B}\left(\mathbb{R}^d \times \mathbb{R}^d\right)$. Then $\rho^{(2)}$ is called second-order product density.*

Just defined statistics ρ and $\rho^{(2)}$ allow us to introduce the concept of pair correlation function so that we can examine the relationships between pairs of points.

Definition 10 (Pair correlation function). *Let X be a point process on $\mathbb{R}^d$ such that intensity function ρ and second-order product density $\rho^{(2)}$ exist. We define pair correlation function:*

$$g(\boldsymbol{u},\boldsymbol{v}) = \frac{\rho^{(2)}(\boldsymbol{u},\boldsymbol{v})}{\rho(\boldsymbol{u})\rho(\boldsymbol{v})},\ \boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^d.$$

If $\rho(x) = 0$ for some $x \in X$, we use convention $c/0 = 0$ for $c \geq 0$.

We should also add an important second-order functional characteristic, the so-called K-function. It is often used to compare whether the examined dataset is more or less clustered compared to the Poisson point process that will be introduced in the Section 1.5 .

Definition 11 (K-function). *Let X be a stationary and isotropic point process on $\mathbb{R}^d$ with non-zero intensity ρ. Then, we define K-function for $t > 0$ as*

$$K(t) = \rho^{-1}\mathbb{E}[\textit{number of extra events within distance } t \textit{ of an arbitrary event}].$$

Note that whenever we talk about the K-function argument in this thesis, we will always consider it as a positive number. Although we do not always state this explicitly.

As shown by [Diggle, 2013, Chapter 4], the following relationship holds between the K-function and the pair correlation function.

Lemma 1. *Let X be a stationary and isotropic point process on $\mathbb{R}^d$ with pair correlation function $g(\boldsymbol{u},\boldsymbol{v}) = g(\boldsymbol{u}-\boldsymbol{v}) = g(\|\boldsymbol{u}-\boldsymbol{v}\|)$ and K-function K. Then*

$$K(t) = \int_{b(0,t)} g(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u},\ t > 0.$$

Respectively,

$$g(t) = \frac{K'(t)}{\sigma_d t^{d-1}},\ t > 0,$$

where $K'(t)$ is the derivative of K and σ_d is the surface area of unit sphere in $\mathbb{R}^d$, i.e. $\sigma_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$. Futher, $b(0,t)$ denotes ball centred in zero with radius $t > 0$.

1.4 Summary statistics for multivariate point processes

Since we do not want to be satisfied with just examining intraspecific relationships, we will first be forced to perform generalizations of the function $g(x,y)$ from definition 10 to multivariate processes. Respectively, generalize the factorial moment measure $\alpha^{(2)}(C)$ and second-order product density $\rho^{(2)}(x,y)$ from definitions 8 and 9.We will use the definition stated by [Møller and Waagepetersen, 2004, Chapter 4.4] to do this.

Definition 12 (Multivariate factorial moment measure, multivariate second-order product density). *Let X be a marked point process on $\mathbb{R}^d$ with marks $1,\ldots,p$. Let*

$C \subset \mathbb{R}^d \times \mathbb{R}^d$. *Then we define multivariate factorial moment measure of process* X *as*

$$\alpha_{i,j}^{(2)}(C) = \mathbb{E} \sum_{\substack{\boldsymbol{u} \in X_i, \\ \boldsymbol{v} \in X_j}}^{\neq} \mathbf{1}_{[(\boldsymbol{u},\boldsymbol{v}) \in C]}, \; i, j = 1, \ldots, p.$$

If such Radon-Nikodym derivative exists, we define multivariate second-order product density $\rho_{i,j}^{(2)}$ *of* X *by relation*

$$\alpha_{i,j}^{(2)}(C) = \int \int \mathbf{1}_{[(\boldsymbol{u},\boldsymbol{v}) \in C]} \rho_{i,j}^{(2)}(\boldsymbol{u}, \boldsymbol{v}) \, \mathrm{d}(\boldsymbol{u}) \mathrm{d}(\boldsymbol{v}).$$

We then get the so-called cross pair correlation function as the ratio of multivariate second-order product density and the appropriate intensity functions.

Definition 13 (Cross pair correlation function)**.** *Let* X *be a marked point process on* $\mathbb{R}^d$ *with marks* $1, \ldots, p$. *Let* ρ_i *be its intensity function for mark* $i = 1, \ldots, p$ *and* $\rho_{i,j}^{(2)}$ *multivariate second-order product density. Then we define cross pair correlation function as*

$$g_{i,j}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\rho_{i,j}^{(2)}(\boldsymbol{u}, \boldsymbol{v})}{\rho_i(\boldsymbol{u}) \rho_j(\boldsymbol{v})}, \; \boldsymbol{u}, \; \boldsymbol{v} \in \mathbb{R}^d, \; i, j = 1, \ldots, p.$$

Of course, we cannot miss the multivariate variant of K-function either.

Definition 14 (Cross K-function)**.** *Let* X *be a marked point process on* $\mathbb{R}^d$ *with marks* $1, \ldots, p$. *Let* ρ_i *be its intensity function for mark* $i = 1, \ldots, p$. *Suppose that measure*

$$\mathcal{K}_{ij}(B) = \frac{1}{|A|} \mathbb{E} \sum_{\substack{\boldsymbol{u} \in X_i, \\ \boldsymbol{v} \in X_j}} \frac{\mathbf{1}_{[\boldsymbol{u} \in A, \boldsymbol{v} - \boldsymbol{u} \in B]}}{\rho_i(\boldsymbol{u}) \rho_j(\boldsymbol{v})}, \; B \subset \mathbb{R}^d, \; i, j = 1, \ldots, p.$$

does not depend on the choice of $A \subset \mathbb{R}^d$ *such that* $0 < |A| < \infty$. *Then, we define the cross* K*-function* K_{ij} *as a measure of* $\mathcal{K}_{ij}$ *on a ball:*

$$K_{ij}(r) = \mathcal{K}_{ij} \left(b\left(0, r\right) \right), \; r > 0.$$

As in the case of the K-function, for the cross K-function we will always assume that the argument is positive.

For isotropic processes we then obtain the relation of the cross K-function and the cross-pair correlation function. It is no surprise, just an exact analogy of the relationship given in the Lemma 1. For the proof of the statement, we refer again to [Møller and Waagepetersen, 2004, Chapter 4.4].

Lemma 2. *Let* X *be a stationary and isotropic marked point process on* $\mathbb{R}^d$ *with marks* $1, \ldots, p$, *with cross pair correlation function* $g_{ij}(\boldsymbol{u}, \boldsymbol{v}) = g_{ij}(\boldsymbol{u} - \boldsymbol{v}) = g_{ij}(\|\boldsymbol{u} - \boldsymbol{v}\|)$ *and cross* K*-funciton* K_{ij}. *Then*

$$K_{ij}(t) = \int_{b(0,t)} g_{ij}(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}, \; t > 0.$$

Respectively,

$$g_{ij}(t) = \frac{K'_{ij}(t)}{\sigma_d t^{d-1}}, \; t > 0,$$

where $K'_{ij}(t)$ *is the derivative of* K_{ij} *and* σ_d *is the surface area of unit sphere in* $\mathbb{R}^d$, *i.e.* $\sigma_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$.

1.5 Poisson point process

The basic processes that spatial statistics study are the Poisson point process. Their class represents ”no interaction” models. At the same time, we are able to perform many calculations for them. They also serve as reference processes when summary characteristics are studied. Last but not least, we derive more complex models from Poisson's point processes, including Cox's point processes, which we want to deal with primarily in this thesis, we will draw from [Møller and Waagepetersen, 2004, Section 4].

We can naturally look at the Poisson point process as a generalization of the well-known one-dimensional version of this process. The Poisson process on $\mathbb{R}$ with intensity λ is then defined by two properties. On the one hand, its number of points in each bounded interval $(a, b]$ has a Poisson distribution with mean value $\lambda(b-a)$. And secondly, for every collection of disjoint bounden intervals $(a_1, b_1], \ldots, (a_n, b_n]$, the numbers of points in them are independent random variables, $n = 2, 3, \ldots$, see Baddeley et al. [2007]. The following definition of the Poisson point process was set by [Rataj, 2006, Section 6].

Definition 15 (Poisson point process)**.** *Let X be a point process on $E \subset \mathbb{R}^d$ with intensity measure Λ that is finite on all compact sets. Let both following conditions hold.*

1. *$N_X(B)$ has a Poisson distribution with mean $\Lambda(B)$, i.e.*

$$N_X(B) \sim Po\left(\Lambda(B)\right), \forall\ B\ \in\ \mathcal{B}(E) \text{ bounded},$$

2. *$N_X(B_1),\ \ldots,\ N_X(B_n)$ are independent for each $n \in \mathbb{N}$ and $B_1,\ \ldots,\ B_n \in \mathcal{B}(E)$ pairwise disjoint bounded sets.*

Then, X is called Poisson point process with intensity measure Λ.

We will generalize this definition slightly. If $\Lambda(B) = 0$ then $N_X(B) \overset{a.s.}{=} 0,\ B \in \mathcal{B}(E)$. And for $\Lambda(B) = \infty,\ B \in \mathcal{B}(E)$, we define $N_X(B) \overset{a.s.}{=} \infty$.

Definition 16 ((In)homogeneous and standard Poisson point process)**.** *Consider Poisson point process X with intensity function ρ on $\mathbb{R}^d$. If the intensity function is constant, X is called homogeneous Poisson process on $\mathbb{R}^d$. Otherwise we say X is inhomogeneous. Furthermore, for $\rho = 1$, we call X standard Poisson point process on $\mathbb{R}^d$.*

Now we can move on to the Cox point process.

1.6 Cox point process

The next step is to extend Poisson point process by randomizing the intensity function. Thus obtained point process is called Cox point process, we can also speak about doubly stochastic Poisson process.

Definition 17 (Cox point process)**.** *Let us consider Poisson point process X_Λ on $\mathbb{R}^d$ with locally finite intensity measure $\Lambda \in \mathcal{M}$. Denote its distribution Π_Λ. Suppose Ψ be a random diffuse measure on $(\mathbb{R}^d, \mathcal{B}^d)$ (i.e. $\Psi(\{\boldsymbol{u}\}) = 0$ for all*

$\boldsymbol{u} \in \mathbb{R}^d$) with distribution Q_Ψ. Then, the Cox point process X with driving random measure Ψ is given by distribution

$$Q(\cdot) = \int_{\mathcal{M}} \Pi_\Lambda(\cdot) Q_\Psi(\mathrm{d}\Lambda). \tag{1.1}$$

To define the Cox point process, in addition to the driving random measure Ψ, we can also use the so-called driving random function φ, which is a Radon-Nikodym derivative (with respect to Lebesgue measure) of Ψ, if it exists. Then,

$$\Psi(B) = \int_B \varphi(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u},\ \ B \in \mathcal{B}.$$

One of the most basic properties of the Cox point process is that the intensity measure is equal to the mean value of its driving measure. Symbolically written:

$$\mathbb{E}N_X(B) = \mathbb{E}\Psi(B), B \in \mathcal{B}^d.$$

This equality can be obtained as follows.

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

2. Univariate log-Gaussian Cox process

The main model we will want to deal with in this thesis is the multivariate log-Gaussian Cox process. First, however, we introduce its basic form, the univariate log-Gaussian Cox process, which considers points from only one type of process instead of adding different marks to them. We will refer to [Møller and Waagepetersen, 2004, Section 5.6] and [Dvořák and Prokešová, 2012, Section 3] when defining the univariate model and setting out the basic properties.

The last thing we need to define before we can introduce a log-Gaussian Cox process is a Gaussian random field. We will use the definitions given by [Cressie, 1993, Section 2.3].

Definition 18 (Random field)**.** *Let $E \subset \mathbb{R}^d$. A random field is a collection of real random variables $\{Z(x);\ x \in E\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*

Definition 19 (Gaussian random field)**.** *A random field $\{Z(x);\ x \in E\}$ is called Gaussian random field if the random vector $\left(Z\left(x_1\right), \ldots, Z\left(x_n\right)\right)^\top$ has n-dimensional normal distribution for any $x_1, \ldots, x_n \in E$ and $n \in \mathbb{N}$.*

2.1 Model definition

Let $\{Z(\boldsymbol{u});\ \boldsymbol{u} \in \mathbb{R}^2\}$, be a Gaussian random field. Then Cox point process X driven by intensity function $\varphi(\boldsymbol{u}) = \exp\{Z(\boldsymbol{u})\}$, $\boldsymbol{u} \in \mathbb{R}^2$ is called log-Gaussian Cox process on $\mathbb{R}^2$. We will use the usual abbreviation LGCP.

The distribution of this process is fully determined by the mean and covariance function

$$\mu(\boldsymbol{u}) = \mathbb{E}Z(\boldsymbol{u}) \text{ and } c(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{cov}(Z(\boldsymbol{u}), Z(\boldsymbol{v})),\ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2.$$

2.2 First and second-order characteristics

Since we want to focus mainly on multivariate LGCP, we will leave the derivation of characteristics to the next section. For the univariate case, we will just state their form.

For the general univariate LGCP we can determine the intensity function, which is

$$\rho(\boldsymbol{u}) = \exp\left\{\mu(\boldsymbol{u}) + \frac{1}{2}c(\boldsymbol{u}, \boldsymbol{u})\right\},\ \boldsymbol{u} \in \mathbb{R}^2$$

and the second-order product density

$$\rho^{(2)}(\boldsymbol{u}, \boldsymbol{v}) = \rho(\boldsymbol{u})\rho(\boldsymbol{v}) \exp\left\{c(\boldsymbol{u}, \boldsymbol{v})\right\},\ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2.$$

However, it is often enough for us to consider a simplified model that is stationary and isotropic. Furthermore, we will consider only a family of models with an exponential covariance function. This is very advantageous, as these models are often used in applications. We will therefore study models with a covariance function equal to

$$c(\boldsymbol{u}, \boldsymbol{v}) = c(\|\boldsymbol{u} - \boldsymbol{v}\|) = \sigma^2 \exp\Big\{ - \delta \|\boldsymbol{u} - \boldsymbol{v}\| \Big\},\ \delta > 0,\ \sigma^2 > 0,\ \boldsymbol{u} \in \mathbb{R}^2. \quad (2.1)$$

Then $\mu(\boldsymbol{u}) = \mu,\ \forall\ \boldsymbol{u} \in \mathbb{R}^2$ and the intensity function equals to

$$\rho(\boldsymbol{u}) = \rho = \exp\Big\{\mu + \frac{\sigma^2}{2}\Big\},\ \boldsymbol{u} \in \mathbb{R}^2$$

and the second-order product density is

$$\rho^{(2)}(\boldsymbol{u}, \boldsymbol{v}) = \rho^{(2)}(\boldsymbol{u} - \boldsymbol{v}) = \rho^2 \exp\Big\{\sigma^2 \exp\Big\{ - \delta \|\boldsymbol{u} - \boldsymbol{v}\| \Big\}\Big\},\ \boldsymbol{u} \in \mathbb{R}^2.$$

Then we are also able to calculate K-function:

$$K(r) = 2\pi \int_0^r s \exp\Big\{\sigma^2 \exp\{-\delta s\}\Big\}\, \mathrm{d}s,\ r \geq 0.$$

Let us move on to the multivariate case.

3. Multivariate log-Gaussian Cox process

Now we can finally focus on multivariate LGCP. As we have already indicated, this is a generalization of univariate LGCP. Specifically, we add a random field describing interspecies relationships. We will build on the article [Waagepetersen et al., 2016, Section 2]. This article works with data from Lansing Woods, which shows the exact location of six different tree species.

3.1 Model definition

We will now introduce so called multivariate log-Gaussian Cox process.

Consider a multivariate point process $X = (X_1, \ldots, X_p)$, $p > 1$, in $\mathbb{R}^2$ where for each $i \in \{1, \ldots, p\}$ component X_i is a Cox point process with driving intensity function φ_i which is of the form

$$\varphi_i(\boldsymbol{u}) = \exp\{Z_i(\boldsymbol{u})\}, \text{ where } Z_i(\boldsymbol{u}) = \mu_i(\boldsymbol{u}) + Y_i(\boldsymbol{u}) + U_i(\boldsymbol{u}), \ \boldsymbol{u} \in \mathbb{R}^2.$$

Whereas μ_i is considered to be a deterministic function. In contrast, U_i and Y_i random fields, specifically, these are mutually independent zero-mean Gaussian fields. We will also add assumptions about their correlation structure.

To better understand what the individual variables in this model mean, imagine that the model describes the mentioned location data of different tree species. Then the random variable U_i describes clustering based on species-specific factors. For example, how far the seeds of tree species i usually spread. While, the independent field Y describes combined influence of various variables such as soil acidity, which affect all types of trees, but possibly in a different manner.

Assume U_i be independent on U_j for each $i \neq j$. Further, let U_i be stationary and isotropic with covariance function

$$\operatorname{cov}\left(U_i(\boldsymbol{u}), U_i(\boldsymbol{u}+\boldsymbol{h})\right) = \sigma_i^2 c_i(||\boldsymbol{h}||) \text{ for } \boldsymbol{u}, \boldsymbol{h} \in \mathbb{R}^2.$$

Therefore both mean and variance of $U_i(\boldsymbol{u})$ are constant for each $\boldsymbol{u} \in \mathbb{R}^2$: $\mathbb{E}U_i(\boldsymbol{u}) = 0$ and $\operatorname{var} U_i(\boldsymbol{u}) = \sigma_i^2$.

The inner structure of the other fields is significantly more complex. Suppose we can get $\boldsymbol{Y}(\boldsymbol{u}) = (Y_1(\boldsymbol{u}), \ldots, Y_p(\boldsymbol{u}))^\top$ in the form

$$\boldsymbol{Y}(\boldsymbol{u}) = \mathbb{A}\boldsymbol{E}(\boldsymbol{u}), \ \boldsymbol{u} \in \mathbb{R}^2,$$

where $\mathbb{A} = (a_{ij})_{i,j=1}^{p,q}$ is $p \times q$ real matrix of coefficients. Further, $\boldsymbol{E}(\boldsymbol{u}) = (E_1(\boldsymbol{u}), \ldots, E_q(\boldsymbol{u}))^\top$, $\boldsymbol{u} \in \mathbb{R}^2$ is zero-mean unit variance Gaussian process which is in addition stationary, isotropic and its components are independent. Its correlation structure is given by covariance functions $r_j(\cdot)$, $j = 1, \ldots, q$ and multivariate covariance function $R(\cdot)$:

$$\begin{aligned} r_j(t) &= \operatorname{cov}\left(E_j(\boldsymbol{u}), E_j(\boldsymbol{u}+\boldsymbol{h})\right), j = 1, \ldots, q, \\ R(t) &= \operatorname{cov}\left(\boldsymbol{E}(\boldsymbol{u}), \boldsymbol{E}(\boldsymbol{u}+\boldsymbol{h})\right) = diag(r_1(t), \ldots, r_q(t)), \ \|\boldsymbol{h}\| = t \geq 0, \boldsymbol{u}, \boldsymbol{h} \in \mathbb{R}^2. \end{aligned}$$

We use notation $diag(d_1, \ldots, d_n)$ for n-dimensional diagonal matrix with diagonal elements $d_1, \ldots, d_n$. We can see that all just defined covariance functions depend only on the length of the argument shift as it follows from stationarity and isotropy. Furthermore,we assume that the components are uncorrelated.

Let's now look at what all this means for the correlation structure of $\boldsymbol{Y}$.

3.2 Model properties

It is easy do derive the covariance function $C(\cdot)$:

$$\begin{aligned} C(t) &= \operatorname{cov}\left(\boldsymbol{Y}(\boldsymbol{u}), \boldsymbol{Y}(\boldsymbol{u}+\boldsymbol{h})\right) = \operatorname{cov}\left(\mathbb{A}\boldsymbol{E}(\boldsymbol{u}), \mathbb{A}\boldsymbol{E}(\boldsymbol{u}+\boldsymbol{h})\right) \\ &= \mathbb{A}\operatorname{cov}\left(\boldsymbol{E}(\boldsymbol{u}), \boldsymbol{E}(\boldsymbol{u}+\boldsymbol{h})\right)\mathbb{A}^\top = \mathbb{A}R(t)\mathbb{A}^\top = \sum_{i=1}^{q} r_i(t) a_{\cdot i} a_{\cdot i}^\top, \end{aligned}$$

for every $\|\boldsymbol{h}\| = t \geq 0,\ \boldsymbol{u}, \boldsymbol{h} \in \mathbb{R}^2$, where $a_{\cdot i}$ denotes $i-$th column of matrix $\mathbb{A}$, $i \in 1,\ \ldots, p$.

For a better idea of covariance function of Y_i, $i = 1, \ldots, p$, we will take advantage of matrix notation.

$$\boldsymbol{Y}(\boldsymbol{u}) = \begin{pmatrix} Y_1(\boldsymbol{u}) \\ \vdots \\ Y_p(\boldsymbol{u}) \end{pmatrix} = \mathbb{A}\boldsymbol{E}(\boldsymbol{u}) = \begin{pmatrix} \boxed{a_{1\cdot}} \\ \boxed{\vdots} \\ \boxed{a_{p\cdot}} \end{pmatrix} \begin{pmatrix} E_1(\boldsymbol{u}) \\ \vdots \\ E_q(\boldsymbol{u}) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{q} a_{1j} E_j(\boldsymbol{u}) \\ \vdots \\ \sum_{j=1}^{q} a_{pj} E_j(\boldsymbol{u}) \end{pmatrix},$$

where $a_{i\cdot}$ denotes i-th row of matrix $\mathbb{A}$. Now we can easily see the following form of the desired covariance function.

$$\begin{aligned} \operatorname{cov}\left(Y_i(\boldsymbol{u}), Y_i(\boldsymbol{u}+\boldsymbol{h})\right) &= \operatorname{cov}\left(\sum_{j=1}^{q} a_{ij} E_j(\boldsymbol{u}), \sum_{j=1}^{q} a_{ij} E_j(\boldsymbol{u}+\boldsymbol{h})\right) \\ &= \sum_{j=1}^{q}\sum_{k=1}^{q} a_{ij} a_{ik} \operatorname{cov}\left(E_j(\boldsymbol{u}), E_k(\boldsymbol{u}+\boldsymbol{h})\right) \\ &= \sum_{j=1}^{q} a_{ij}^2 \operatorname{cov}\left(E_j(\boldsymbol{u}), E_j(\boldsymbol{u}+\boldsymbol{h})\right) + \sum_{\substack{j=1 \\ j\neq k}}^{q} a_{ij} a_{ik} \operatorname{cov}\left(E_j(\boldsymbol{u}), E_k(\boldsymbol{u}+\boldsymbol{h})\right) \\ &= \sum_{j=1}^{q} a_{ij}^2 r_j(\|\boldsymbol{h}\|),\ \ \boldsymbol{u}, \boldsymbol{h} \in \mathbb{R}^2. \end{aligned}$$

The last equality comes from independence and subsequent zero correlation of components $E_j(\boldsymbol{u})$ and $E_k(\boldsymbol{u}+\boldsymbol{h})$ for different j and k.

So, using the assumption $r_i(0) = 1$, we can see that the variance of the random variable $Y_i(\cdot)$ equals to

$$\operatorname{var} Y_i(\boldsymbol{u}) = \sum_{j=1}^{q} a_{ij}^2 = a_{i\cdot} a_{i\cdot}^\top,\ \ \boldsymbol{u} \in \mathbb{R}^2,\ \ i = 1, \ldots, p.$$

Let's now focus on the properties and basic characteristics of the individual processes X_i, $i = 1, \ldots, p$. Firstly, we would like to explore the intensity function.

For this purpose, let's denote for each process $X_i, i = 1, \dots, p$ its intensity measure Λ_i and its driving random measure Ψ_i. Then the relation holds

$$\Psi_i(B) = \int_B \varphi_i(\boldsymbol{u})\mathrm{d}(\boldsymbol{u}),\ B \in \mathcal{B}^2.$$

Since X_i is a Cox point process, we know that its intensity measure and mean of driving measure are equal:

$$\Lambda_i(B) = \mathbb{E}N_{X_i}(B) = \mathbb{E}\Psi_i(B),\ B \in \mathcal{B}^2. \tag{3.1}$$

Substituting into the formula above and using the definition relationship $\Lambda_i(B) = \int_B \rho_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$ helps us to express the intensity function ρ_i in a more concrete form.

$$\int_B \rho_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \mathbb{E}\Psi_i(B) = \mathbb{E}\int_B \varphi_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u}.$$

If we further change the order of the mean value and the integral, we get

$$\int_B \rho_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \int_B \mathbb{E}\varphi_i(\boldsymbol{u})\mathrm{d}\boldsymbol{u},$$

so the following formula expressing the intensity function holds almost everywhere

$$\rho_i(\boldsymbol{u}) = \mathbb{E}\varphi_i(\boldsymbol{u}) = \mathbb{E}\exp\{\mu_i(\boldsymbol{u}) + U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u})\}. \tag{3.2}$$

We will use the non-randomness of the function μ and the independence of the random variables U_i and Y_i, so that we can write it in the following form.

$$\begin{aligned}\rho_i(\boldsymbol{u}) &= \mathbb{E}\exp\{\mu_i(\boldsymbol{u}) + U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u})\} \\ &= \exp\{\mu_i(\boldsymbol{u})\}\mathbb{E}\exp\{U_i(\boldsymbol{u})\}\mathbb{E}\exp\{Y_i(\boldsymbol{u})\},\ \boldsymbol{u} \in \mathbb{R}^2.\end{aligned}$$

In order to calculate the intensity function, it is enough to determine the mean values of random variables $\exp\{U_i(\boldsymbol{u})\}$ and $\exp\{Y_i(\boldsymbol{u})\}$. In both cases, it is an exponential transformation of a normally distributed random variable. In other words log-Gaussian random variables. So we can easily determine their mean value according to the formula given by the following Lemma, see [Crow, 1987, Chapter 4].

Lemma 3 (Log-Gaussian distribution)**.** *Let $\boldsymbol{Z} = (Z_1, \dots, Z_k)^\top$, $k \in \mathbb{N}$, be a $k-$dimensional random vector with multivariate normal distribution $\mathcal{N}_k(\boldsymbol{\nu}, \Sigma)$, where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^\top$ and $\Sigma = (\Sigma_{ij})_{ij=1}^k$. Then random vector obtained as exponential transformation of Z has multivariate log-normal distributions and with mean whose i-th element, $i = 1, \dots, k$, equals*

$$\mathbb{E}\exp\{Z_i\} = \exp\left\{\nu_i + \frac{1}{2}\Sigma_{ii}\right\}$$

and covariance matrix $\operatorname{Var}\exp\{\boldsymbol{Z}\}$ *whose element in position* $(i,j), i,j = 1, \dots, k$, *is shaped*

$$(\operatorname{Var}\exp\{\boldsymbol{Z}\})_{ij} = \exp\left\{\nu_i + \nu_j + \frac{1}{2}\left(\Sigma_{ii} + \Sigma_{jj}\right)\right\}\left(\exp\{\Sigma_{ij}\} - 1\right).$$

By substituting, we gain equalities

$$\mathbb{E}\exp\{U_i(\boldsymbol{u})\} = \exp\left\{\frac{1}{2}\sigma_i^2\right\}$$

and

$$\mathbb{E}\exp\{Y_i(\boldsymbol{u})\} = \exp\left\{\frac{1}{2}a_{i\cdot}a_{i\cdot}^\top.\right\}$$

All together, the intensity function of X_i equals

$$\rho_i(\boldsymbol{u}) = \exp\left\{\mu_i(\boldsymbol{u}) + \frac{1}{2}\sigma_i^2 + \frac{1}{2}a_{i\cdot}a_{i\cdot}^\top\right\}, \ \boldsymbol{u} \in \mathbb{R}^2. \tag{3.3}$$

Having gained a basic idea of the intensity with which the various points of our process occur, let's now focus on their interrelationships. The cross pair correlation function g_{ij} will be used for this. Definitions can be recalled in the Section 1.4.

Let us calculate the multivariate second order product density $\rho_{ij}^{(2)}$ for our case. Similarly to the relation (3.1) for the Cox process, the following relation applies to the second order:

$$\alpha_{i,j}^{(2)}(B_1 \times B_2) = \mathbb{E}\Psi_i(B_1)\Psi_j(B_2), \ B_1, B_2 \subset \mathbb{R}^2.$$

By disintegration for points of the process $\boldsymbol{u} \in X_i, \boldsymbol{v} \in X_j$ we then get the formula

$$\rho_{i,j}^{(2)}(\boldsymbol{u}, \boldsymbol{v}) = \mathbb{E}\varphi_i(\boldsymbol{u})\varphi_j(\boldsymbol{v}).$$

from which we can already calculate the rest. Since we are considering a stationary and isotropic process, these functions are given only by the distance of the points $\boldsymbol{u} \in X_i, \boldsymbol{u}+\boldsymbol{h} \in X_j$, which we will denote $\|\boldsymbol{h}\|$.

Please note that in this thesis we will use the usual abuse of notation, where we use the same symbol for both a function with a scalar argument and a vector or pair of vector arguments.

So, let's put it further

$$\begin{aligned}
\rho_{i,j}^{(2)}(\|\boldsymbol{h}\|) &= \rho_{i,j}^{(2)}(\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{h}) = \mathbb{E}\varphi_i(\boldsymbol{u})\varphi_j(\boldsymbol{u}+\boldsymbol{h}) \\
&= \mathbb{E}\exp\{\mu_i(\boldsymbol{u}) + U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u})\}\exp\{\mu_j(\boldsymbol{u}+\boldsymbol{h}) + U_j(\boldsymbol{u}+\boldsymbol{h}) + Y_j(\boldsymbol{u}+\boldsymbol{h})\} \\
&= \exp\{\mu_i(\boldsymbol{u}) + \mu_j(\boldsymbol{u}+\boldsymbol{h})\}\mathbb{E}\exp\{U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u}) + U_j(\boldsymbol{u}+\boldsymbol{h}) + Y_j(\boldsymbol{u}+\boldsymbol{h})\}.
\end{aligned}$$

Let's now look at the distribution of the sum of random variables in the argument of the exponential function, the individual distributions of which are these

$$\begin{gathered}
U_i(\cdot) \sim \mathcal{N}(0, \sigma_i^2), \ U_j(\cdot) \sim \mathcal{N}(0, \sigma_j^2) \\
Y_i(\cdot) \sim \mathcal{N}(0, a_{i\cdot}a_{i\cdot}^\top), \ Y_j(\cdot) \sim \mathcal{N}(0, a_{j\cdot}a_{j\cdot}^\top),
\end{gathered}$$

then

$$U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u}) \sim \mathcal{N}\left(0, \sigma_i^2 + a_{i\cdot}a_{i\cdot}^\top + 2\mathrm{cov}\left(U_i\left(\boldsymbol{u}\right), Y_i\left(\boldsymbol{u}\right)\right)\right) = \mathcal{N}\left(0, \sigma_i^2 + a_{i\cdot}a_{i\cdot}^\top\right),$$

similarly, of course

$$U_i(\boldsymbol{u}+\boldsymbol{h}) + Y_i(\boldsymbol{u}+\boldsymbol{h}) \sim \mathcal{N}\left(0, \sigma_i^2 + a_{i\cdot}a_{i\cdot}^\top\right).$$

And so

$$\begin{aligned} &U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u}) + U_j(\boldsymbol{u}+\boldsymbol{h}) + Y_j(\boldsymbol{u}+\boldsymbol{h}) \\ &\quad\sim \mathcal{N}\Big(0, \mathrm{var}\left(U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u})\right) + \mathrm{var}\left(U_j(\boldsymbol{u}+\boldsymbol{h}) + Y_j(\boldsymbol{u}+\boldsymbol{h})\right) \\ &\qquad\qquad + 2\mathrm{cov}\left(U_i\left(\boldsymbol{u}\right) + Y_i\left(\boldsymbol{u}\right), U_j\left(\boldsymbol{u}+\boldsymbol{h}\right) + Y_j\left(\boldsymbol{u}+\boldsymbol{h}\right)\right)\Big), \end{aligned}$$

where using the independence of fields $\mathbf{U}$ and $\mathbf{Y}$:

$$\begin{aligned} &\mathrm{cov}\left(U_i\left(\boldsymbol{u}\right) + Y_i\left(\boldsymbol{u}\right), U_j\left(\boldsymbol{u}+\boldsymbol{h}\right) + Y_j\left(\boldsymbol{u}+\boldsymbol{h}\right)\right) \\ &\quad= \mathrm{cov}\left(U_i\left(\boldsymbol{u}\right), U_j\left(\boldsymbol{u}+\boldsymbol{h}\right)\right) + \mathrm{cov}\left(U_i\left(\boldsymbol{u}\right), Y_j\left(\boldsymbol{u}+\boldsymbol{h}\right)\right) \\ &\qquad+ \mathrm{cov}\left(Y_i\left(\boldsymbol{u}\right), U_j\left(\boldsymbol{u}+\boldsymbol{h}\right)\right) + \mathrm{cov}\left(Y_i\left(\boldsymbol{u}\right), Y_j\left(\boldsymbol{u}+\boldsymbol{h}\right)\right) \\ &\quad= \mathbf{1}_{[i=j]}\sigma_i^2 c_i\left(\|\boldsymbol{h}\|\right) + 0 + 0 + \sum_{k=1}^{q} a_{ik}a_{jk}r_k\left(\|\boldsymbol{h}\|\right), \end{aligned}$$

so

$$\begin{aligned} &U_i(\boldsymbol{u}) + Y_i(\boldsymbol{u}) + U_j(\boldsymbol{u}+\boldsymbol{h}) + Y_j(\boldsymbol{u}+\boldsymbol{h}) \\ &\quad\sim \mathcal{N}\left(0, \sigma_i^2 + a_{i\cdot}a_{i\cdot}^\top + \sigma_j^2 + a_{j\cdot}a_{j\cdot}^\top + \mathbf{1}_{[i=j]}2\sigma_i^2 c_i\left(\|\boldsymbol{h}\|\right) + 2\sum_{k=1}^{q} a_{ik}a_{jk}r_k\left(\|\boldsymbol{h}\|\right)\right). \end{aligned}$$

And thus $\exp\{U_i(\boldsymbol{u})+Y_i(\boldsymbol{u})+U_j(\boldsymbol{u}+\boldsymbol{h})+Y_j(\boldsymbol{u}+\boldsymbol{h})\}$ has a log-normal distribution with mean value

$$\exp\left\{\frac{1}{2}\left(\sigma_i^2 + \sigma_j^2 + a_{i\cdot}a_{i\cdot}^\top + a_{j\cdot}a_{j\cdot}^\top\right) + \mathbf{1}_{[i=j]}\sigma_i^2 c_i\left(\|\boldsymbol{h}\|\right) + \sum_{k=1}^{q} a_{ik}a_{jk}r_k\left(\|\boldsymbol{h}\|\right)\right\}.$$

Therefore, the second order product density equals

$$\begin{aligned} \rho_{i,j}^{(2)}\left(\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{h}\right) = \exp\Bigg\{&\mu_i(\boldsymbol{u}) + \mu_j(\boldsymbol{u}+\boldsymbol{h}) + \frac{1}{2}\Bigg(\sigma_i^2 + \sigma_j^2 + \\ &+ a_{i\cdot}a_{i\cdot}^\top + a_{j\cdot}a_{j\cdot}^\top\Bigg) + \mathbf{1}_{[i=j]}\sigma_i^2 c_i\left(\|\boldsymbol{h}\|\right) + \sum_{k=1}^{q} a_{ik}a_{jk}r_k\left(\|\boldsymbol{h}\|\right)\Bigg\}. \end{aligned}$$

Finally, we can determine the desired cross pair correlation function using the already calculated intensity function (3.3):

$$\begin{aligned} g_{i,j}\left(\|\boldsymbol{h}\|\right) = g_{i,j}\left(\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{h}\right) &= \frac{\rho_{i,j}^{(2)}\left(\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{h}\right)}{\rho_i(\boldsymbol{u})\rho_j(\boldsymbol{u}+\boldsymbol{h})} \\ &= \exp\left\{\mathbf{1}_{[i=j]}\sigma_i^2 c_i\left(\|\boldsymbol{h}\|\right) + \sum_{k=1}^{q} a_{ik}a_{jk}r_k\left(\|\boldsymbol{h}\|\right)\right\}. \end{aligned}$$

In addition, we can calculate K-function, resp. cross K-function using the relation from the Lemmas 1 and 2:

$$
\begin{aligned}
K_i(r) &= 2\pi \int_0^r s \exp\left\{\sigma_i^2 c_i(s) + \sum_{k=1}^{q} a_{ik}^2 r_k(s)\right\} \mathrm{d}s, \\
K_{ij}(r) &= 2\pi \int_0^r s \exp\left\{\sum_{k=1}^{q} a_{ik} a_{jk} r_k(s)\right\} \mathrm{d}s.
\end{aligned}
$$

Finally, we should mention that, as in the case of the univariate model, we will want to work with models with the exponential covariance function, which is popular in practice. A proposal for its use can be found, for example, in [Møller and Waagepetersen, 2004, Chapter 4.2]. We will therefore consider these functions in the form of

$$
\begin{aligned}
c_i(h) &= \exp\{-\delta_i h\},\ \delta_i > 0,\ i = 1, \ldots, p, \\
r_k(h) &= \exp\{-\epsilon_k h\},\ \epsilon_k > 0,\ k = 1, \ldots, q.
\end{aligned}
$$

4. Composite likelihood estimation

In this chapter we will deal with estimating the model parameters. We will focus especially on composite likelihood method. We are inspired by the article [Waagepetersen, 2007, Sections 2 and 6] and also by [Møller and Waagepetersen, 2007, Section 8] which develops the idea. We will derive it for the first and second order of the unmarked point process. We then generalise the composite likelihood for the marked point process and calculate it for the investigated multivariate log-Gaussian Cox process presented in the previous chapter.

As usual, the vector of unknown parameters will be denoted by $\boldsymbol{\theta}$ from the parameter space Θ. We will also work with points of point process X, denoting $\boldsymbol{u} \in X$. Since we are looking for parameters on specific data set, i.e. a realization of the process X, we will look at the points as deterministic, not stochastic.

In order to develop this method, we will assume the existence of intensity function, second-order product density, pair correlation function and cross pair correlation function throughout the thesis.

4.1 Composite likelihood for the first order

4.1.1 Derivation of the general formula

In order to derive a formula for calculating composite likelihood, we need to introduce assumptions for the intensity function. Specifically, we need this to be continuous, integrable and bounded for each $\boldsymbol{\theta} \in \Theta$, that is

$$\int_W \rho_{\boldsymbol{\theta}}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} < \infty \text{ and } \exists\, M^1 \in \mathbb{R}\ \forall\, \boldsymbol{u} \in \mathbb{R}^2 :\ |\rho_{\boldsymbol{\theta}}(\boldsymbol{u})| \leq M^1.$$

Consider sequence of partition $(D^n)_{n\in\mathbb{N}}$ of observation window W with descending norms, with cells of the same size and property

$$\forall\, s \in S^n :\ \Big|\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\Big| < 1, \tag{4.1}$$

where C_s^n are cells of partition D^n, $\boldsymbol{u}_s^n$ a representing point in C_s^n and S^n the index set of D^n. By the norm of partition, we mean the maximum cell area of this partition.

Next, we denote the indicator that a point from the process X appears in the cell C_s^n is called $N_s^n = \mathbf{1}_{[N_X(C_s^n)>0]}$ and finally the probability of the truth of this condition for the given parameter $\boldsymbol{\theta} \in \Theta$ is $p_s^n(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}\left[N_s^n = 1\right]$. We assume that the process behaves independently in the cells. Then, the composite likelihood function of the discrete, approximating model is given by

$$L^n(\boldsymbol{\theta}) = \prod_{s\in S^n} p_s^n\left(\boldsymbol{\theta}\right)^{N_s^n} \left(1 - p_s^n\left(\boldsymbol{\theta}\right)\right)^{(1-N_s^n)}.$$

However, we simplify the situation and probability of the occurrence of a point in a given cell is approximated by the intensity function multiplied by the cell area:

$$p_s^n(\boldsymbol{\theta}) \approx \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|.$$

Then, the composite likelihood function is following

$$CL^n(\boldsymbol{\theta}) = \prod_{s\in S^n} \left(\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right)^{N_s^n} \left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right)^{(1-N_s^n)}.$$

To make the calculation even easier, let's move to the logarithmic form. This can be further broken down into four individual summands.

$$\begin{aligned}\log CL^n(\boldsymbol{\theta}) &= \sum_{s\in S^n} N_s^n \log\left(\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right) + (1-N_s^n)\log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right)\\ &= \sum_{s\in S^n} N_s^n \log \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n) + \sum_{s\in S^n} N_s^n \log|C_s^n|\\ &\quad + \sum_{s\in S^n} \log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right) - \sum_{s\in S^n} N_s^n \log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right).\end{aligned}$$

Let's go through each of these four sums now and see what their value will be if we send the cell sizes to zero.

- $\sum_{s\in S^n} N_s^n \log \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)$:

If we consider the limit shrinkage of cells C_s, then there can be at most one point in each cell. And non-zero are only those summands in which we consider a cell that contains at least one point of the process.

Next, we need to deal with the possibility that the intensity function $\rho_\theta(\cdot)$ equals to zero. And so its logarithm is equal to $-\infty$. However, from the definition of this function, this can only happen with zero probability. If we denote $A = \{u \in W; \rho_\theta(u) = 0\}$, then $\mathbb{E}N_X(A) = 0$ and $\mathbb{P}(N_X(A) \geq 1) = 0$. So we can write

$$\sum_{s\in S^n} N_s^n \log \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n) \xrightarrow[n\to\infty]{} \sum_{\boldsymbol{u}\in X} \log \rho_{\boldsymbol{\theta}}(\boldsymbol{u}).$$

Notice the limit sum does not depend on the cells area.

- $\sum_{s\in S^n} N_s^n \log|C_s^n|$:

The second sum does not depend on the unknown parameter $\boldsymbol{\theta}$. Therefore, it has no effect on maximizing the composite likelihood function.

Let us realize that although these elements tend to infinity, if we reduce area of cells C_s^n, we do not mind. Since we construct the composite likelihood function for one discrete model (with a given partition), the desired point of the maximum is not affected by this term. Therefore, we can consider an equivalent composite likelihood function denoted by a tilde, in which we omit this term.

- $\sum_{s\in S^n} \log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right)$:

For the third sum, we will use Taylor's expansion for the logarithmic function. It has the Maclaurin series

$$\log(1-x) = -\sum_{l=1}^{\infty} \frac{x^l}{l}, \ x \in [-1, 1). \tag{4.2}$$

We might apply this expansion for $x = \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|$ using property (4.1).

We will use only the first element of the Taylor series to approximate the logarithmic function. Therefore we get

$$\sum_{s\in S^n} \log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right) \approx -\sum_{s\in S^n} \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|$$

which we can also look at as an integral from the intensity function, if we send the cell sizes $|C_s^n|$ to zero:

$$-\sum_{s\in S^n} \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n| \xrightarrow[n\to\infty]{} -\int_W \rho_{\boldsymbol{\theta}}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.$$

The limit integral thus derived is Riemann. According to [Jarník, 1984, Theorem 157] we know that this exists if it is constructed on a bounded set that has zero measure of its boundary, and moreover, the integrated function is continuous up to the set of measure zero. Then there is also the Lebesgue integral and both integrals are equal, see [Jarník, 1984, Theorem 161]. The following two Theorems testify to this.

Theorem 4 (Existence of Riemann integral)**.** *Let $M \subset \mathbb{R}^n$ be a bounded set and f be a real function on $\mathbb{R}^n$ which is bounded on M. Then, Riemann integral of f on M exists if and only if both following conditions are satisfied:*

1. *The boundary of set M has zero Lebesgue measure.*
2. *The set of inner points of M where function f is not continuous has zero Lebesgue measure.*

Theorem 5 (Identity of Riemann and Lebesgue integrals)**.** *Let $M \subset \mathbb{R}^n$ set and f be a real function on $\mathbb{R}^n$. If Riemann integral of f on M exists then also Lebesgue integral of f on M exists and they are equal.*

The observation window, which we consider bounded and satisfies the zero-measure condition for its boundary. We also assumed the boudedness and continuity of the intensity function. Since all assumptions are met,the limit transition to the integral is completely justified.

Now, let us examine the behavior of the remaining elements of Taylor's expansion. We will show that it converges to zero if we shrink the cells C_s^n. To estimate the remaining elements of the series, we will use the Cauchy's form which we recall in Lemma 6 as stated by Darah [2020].

Lemma 6 (Cauchy's Form of the Remainder)**.** *Let $k \in \mathbb{N}$, I interval, $a,\ x \in I$ and f function such that it is continuous and $(k+1)$-st derivative of f, denoted by $f^{(k+1)}$ exists on I. Then there is ξ between point a and x such that*

$$f(x) - \left(\sum_{l=0}^{k} \frac{f^{(l)}(a)}{l!}(x-a)^l\right) = \frac{1}{k!} f^{(k+1)}(\xi)(x-\xi)^k(x-a).$$

We use this statement for $f(x) = \log(1-x),\ k = 1, a = 0$ and $x = \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|$. So

$$-\sum_{l=2}^{\infty} \frac{\rho_{\boldsymbol{\theta}}^l(\boldsymbol{u}_s^n)|C_s^n|^l}{l} = -\frac{1}{(1-\xi^n)^2}\left(\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n| - \xi^n\right)\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|,$$

where $\xi^n \in (0, \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|)$.

However, we need to explore sum of remainders over all cells in the partition resp. its limit behaviour when n tends to infinity. So, let us look at the expression

$$\sum_{s\in S^n} -\frac{1}{(1-\xi^n)^2}\left(\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n| - \xi^n\right)\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|.$$

Since $\xi^n \in (0, \rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|)$, and function $\rho_{\boldsymbol{\theta}}$ is non-negative and bounded, we can bound it from above and have a look at its limit.

$$\begin{aligned}
&\left|\sum_{s\in S^n} -\frac{1}{(1-\xi^n)^2}\left(\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n| - \xi^n\right)\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right| \\
&\quad\le \sum_{s\in S^n} \frac{1}{(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|)^2}\left|\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n| - \xi^n\right|\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n| \\
&\quad\le \sum_{s\in S^n} \frac{1}{(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|)^2}\rho_{\boldsymbol{\theta}}^2(\boldsymbol{u}_s^n)|C_s^n|^2 \\
&\quad\le \sum_{s\in S^n} \frac{1}{(1-\sup_{\boldsymbol{u}\in W}\rho_{\boldsymbol{\theta}}(\boldsymbol{u})|C_s^n|)^2}\sup_{\boldsymbol{u}\in W}\rho_{\boldsymbol{\theta}}^2(\boldsymbol{u})|C_s^n|^2 \\
&\quad= \sum_{s\in S^n} \frac{1}{(1-\sup_{\boldsymbol{u}\in W}\rho_{\boldsymbol{\theta}}(\boldsymbol{u})|W|)^2}\sup_{\boldsymbol{u}\in W}\rho_{\boldsymbol{\theta}}^2(\boldsymbol{u})|C_1^n|^2 \\
&\quad= |W|\frac{1}{(1-\sup_{\boldsymbol{u}\in W}\rho_{\boldsymbol{\theta}}(\boldsymbol{u})|W|)^2}\sup_{\boldsymbol{u}\in W}\rho_{\boldsymbol{\theta}}^2(\boldsymbol{u})|C_1^n| \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

In the last equation, we used the independence of summands on coefficient s and the fact that the size of the set S^n equal to $\frac{|W|}{|C_1^n|}$ which follows from uniformity of cell size in each partition.

As the norm of partition decreases with increasing n, the sum of Cauchy's remainders after Taylor approximation tend to zero and

$$\sum_{s\in S^n} \log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right) \xrightarrow[n\to\infty]{} -\int_W \rho_{\boldsymbol{\theta}}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.$$

- $\sum_{s\in S^n} N_s^n \log\left(1-\rho_{\boldsymbol{\theta}}(\boldsymbol{u}_s^n)|C_s^n|\right)$:

The fourth sum converges with decreasing areas $|C_s^n|$ to zero, because only the finite number of indicators N_s^n is non-zero and for such $s \in S^n$ the argument of the logarithm converges to one, i.e. the logarithm tends to zero.

So, overall, we can take the log composite likelihood function for unknown parameter $\boldsymbol{\theta} \in \Theta$ in the following form

$$\log \widetilde{CL}(\boldsymbol{\theta}) = \sum_{\boldsymbol{u}\in X}\left(\log \rho_{\boldsymbol{\theta}}(\boldsymbol{u})\right) - \int_W \rho_{\boldsymbol{\theta}}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.$$

4.1.2 first-order CL in multivariate LGCP

If we had additional information about the function μ_i, we could include in the vector of unknown parameters only σ_i^2 and elements of the matrix A. However, since we lack it, we also need to estimate this function. To do this, we use a simplification, namely assume stationarity, so μ_i is constant, i.e.

$$\mu_i(\boldsymbol{u}) = \mu_i \ \forall \boldsymbol{u} \in \mathbb{R}^2.$$

Therefore, in the LGCP model we examined, the vector of unknown parameters for process X_i equals:

$$\boldsymbol{\theta}_i = (\sigma_i^2, a_{i1}, \ldots, a_{ip}, \mu_i)$$

and the simplified intensity function has form

$$\rho_{i,\boldsymbol{\theta}}(\boldsymbol{u}) = \rho_{i,\boldsymbol{\theta}} = \exp\left\{\mu_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sum_{l=1}^{q} a_{il}^2\right\}.$$

This function is continuous, integrable and bounded on a bounded observation window W. This can be assumed as it is intended to describe a real situation in which it determines the expected number of points in the W, which is certainly not infinite. The assumptions used in this composite likelihood function are therefore met.

So the relevant log composite likelihood function equals

$$\begin{aligned}\log \widetilde{CL}(\boldsymbol{\theta}_i) = &\sum_{u\in X_i}\left(\mu_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sum_{l=1}^{q} a_{il}^2\right) - \int_W \exp\left\{\mu_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sum_{l=1}^{q} a_{il}^2\right\} \mathrm{d}\boldsymbol{u} \\ =&N_{X_i}(W)\left(\mu_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sum_{l=1}^{q} a_{il}^2\right) - |W| \exp\left\{\mu_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sum_{l=1}^{q} a_{il}^2\right\}.\end{aligned}$$

However, we see that this model has too many unknown parameters to use this composite likelihood in a multivariate LGCP. In addition, we have not yet looked at any interactions between points. So, move on to second-order composite likelihood.

4.2 Composite likelihood for the second order

Determining the parameters as we showed in the previous section is, of course, fully correct. However, this procedure does not take into account the interaction of points. Note that in composite likelihood it counts only one characteristic, namely the number of points. To improve the estimation of parameters, we will include second-order characteristics in the calculation.

4.2.1 Derivation of the formula

Again, we are inspired by the article [Møller and Waagepetersen, 2007, Section 8.1], which suggests how to include the second order in the calculation of composite likelihood function. For this purpose, we will consider the density derived from the probability that there is a pair of points in two given cells of the split observation window. We then estimate this probability using the second-order product density.

In order to build this procedure, we need to meet three assumptions. We have already met them in an analogous form at the first-order case, they are continuity, integrability and boundedness, this time applied to the second-order product density. I.e. for $\boldsymbol{\theta} \in \Theta$:

$$\lim_{(\boldsymbol{u},\boldsymbol{v})\to(\boldsymbol{u}_0,\boldsymbol{v}_0)} \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u},\boldsymbol{v}) = \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}_0,\boldsymbol{v}_0),$$
$$\int_W \int_W \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v} < \infty,$$
$$\exists\, M^2 \in \mathbb{R}\ \forall\, \boldsymbol{u},\ \boldsymbol{v} \in \mathbb{R}^2 :\ |\rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u},\boldsymbol{v})| \leq M^2.$$

The third one is stationarity of process X, which will ensure constant intensity function ρ.

So let us introduce notation for the construction of second-order composite likelihood. Consider the sequence of partitions the observation window $(D^n)_{n\in\mathbb{N}}$ with cells of the same area C^n_s indexed by s from the index set S^n. The areas of cells tend to zero when n increases. By $\boldsymbol{u}^n_s$ we understand a representing point from cell C^n_s.

Consider only such small partitions that for all $n \in \mathbb{N}$

$$\forall\, s,\ t \in S^n :\ \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}^n_s,\boldsymbol{u}^n_t)\,|C^n_s|\,|C^n_t| < 1. \tag{4.3}$$

.

Next, let's have indicators of the simultaneous occurrence of a pair of points of the process X in the cells C^n_s and C^n_t, where $s,\ t \in S^n$:

$$N^n_{s,t} = \mathbf{1}_{\left[N_X(C^n_s)>0\ \&\ N_X(C^n_t)>0.\right]}$$

We denote the probability that the point occurs in the two cells, for given parameter $\boldsymbol{\theta}$, as

$$p^n_{s,t}(\boldsymbol{\theta}) = \mathbb{P}\left[N^n_{s,t} = 1\right].$$

We assume that the individual indicators are independent. Then, the composite likelihood function has for unknown parameter $\boldsymbol{\theta} \in \Theta$ form

$$L^n(\boldsymbol{\theta}) = \prod_{s,t\in S^n}^{\neq} p^n_{s,t}(\boldsymbol{\theta})^{N^n_{s,t}} \left(1 - p^n_{s,t}(\boldsymbol{\theta})\right)^{\left(1-N^n_{s,t}\right)}.$$

The symbol $\prod^{\neq}_{a,b}$ denotes the product over all pairs of different points a and b.

As mentioned in the introduction to this section, we approximate probability $p^n_{s,t}$ using the second-order product density. Which makes good sense if we look at the second-order density as the density of a pair of points.

$$p^n_{s,t}(\boldsymbol{\theta}) \approx \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}^n_s,\boldsymbol{u}^n_t)|C^n_s||C^n_t|$$

The rest of the derivation will be very similar to the first order. Hence the form of the composite likelihood is

$$CL^n(\boldsymbol{\theta}) = \prod_{s,t\in S^n}^{\neq} \left(\rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}^n_s,\boldsymbol{u}^n_t)|C^n_s||C^n_t|\right)^{N^n_{s,t}} \left(1 - \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}^n_s,\boldsymbol{u}^n_t)|C^n_s||C^n_t|\right)^{\left(1-N^n_{s,t}\right)}.$$

And the log-composite likelihood function again consists of four sums, which we will examine separately.

$$
\begin{aligned}
\log CL^n(\boldsymbol{\theta}) = & \sum_{s,t \in S^n}^{\neq} N_{s,t}^n \log\left(\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \\
& + \left(1 - N_{s,t}^n\right) \log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \\
= & \sum_{s,t \in S^n}^{\neq} N_{s,t}^n \log \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n) + \sum_{s,t \in S^n}^{\neq} N_{s,t}^n \log\left(|C_s^n||C_t^n|\right) \\
& + \sum_{s,t \in S^n}^{\neq} \log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \\
& - \sum_{s,t \in S^n}^{\neq} N_{s,t}^n \log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right).
\end{aligned}
$$

So let us see how the individual sums behave if we send the cell size to zero.

- $\sum_{s,t \in S^n} N_{s,t}^n \log \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)$:

If we consider small enough cells into which no more than one point of the process falls, we can neglect the zero summands of the first sum.

As in the case of the first order, here we should mention the possibility that $\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n) = 0$, for which we would get summand equal to minus infinity. However, this will only happen with a probability of zero. So we can write that

$$
\sum_{s,t \in S^n}^{\neq} N_{s,t}^n \log \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n) \xrightarrow[n \to \infty]{} \sum_{\boldsymbol{u},\boldsymbol{v} \in X}^{\neq} \log \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v}).
$$

- $\sum_{s,t \in S^n}^{\neq} N_{s,t}^n \log\left(|C_s^n||C_t^n|\right)$:

This sum does not depend on the unknown parameter $\boldsymbol{\theta}$, we will therefore consider equivalent composite likelihood function in which we neglect it. The point of maximum that we find in this way will, of course, be the same as if we continued to consider this term.

- $\sum_{s,t \in S^n}^{\neq} \log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right)$:

For the logarithmic function we use the expansion into the Taylor's series (4.2). Assumption (4.3) ensures that the variable $\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|$ belongs to the desired interval $[-1, 1)$. We will use the first term of Taylor's expansion for the estimation:

$$
\log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \approx -\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|,
$$

respectively

$$
\sum_{s,t \in S^n}^{\neq} \log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \approx - \sum_{s,t \in S^n}^{\neq} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|,
$$

If we consider the limit of zero size of cells C_s^n and C_t^n, then we can look at the estimating sum as a double integral:

$$\lim_{n\to\infty} \sum_{s,t\in S^n}^{\neq} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n| = \int_W \int_W \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v}) \,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}$$

As in the case of the first order, it is necessary to stop here and pay attention to the limit transition from discrete series to the double integral. Again, we can use the Theorems 4 and 5 to guarantee that this integral exists, is well defined, and can be viewed not only as Riemann integral but also as a Lebesgue integral. All the assumptions are fulfilled. The observation window is bounded, its boundary has zero Lebesgue measure and the second-order product density is assumed to be bounded and continuous.

For the other elements of the power series we will use Cauchy's form of the remainder stated by Lemma 6. We will get out of here

$$\begin{aligned} -\sum_{l=2}^{\infty} \frac{\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)^l |C_s^n|^l |C_t^n|^l}{l} = {} & -\frac{1}{(1-\xi^n)^2}\left(\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n| - \xi^n\right) \\ & \cdot \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|, \end{aligned}$$

where $\xi^n \in \left(0, \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right)$.

Again, we will be interested in the limit of the sum of these remainders.

$$\begin{aligned} & \left| \sum_{s,t\in S^n}^{\neq} -\frac{1}{(1-\xi^n)^2}\left(\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n| - \xi^n\right) \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n| \right| \\ & \quad \le \sum_{s,t\in S^n}^{\neq} \frac{1}{(1-\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|)^2} \left(\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)\right)^2 |C_s^n|^2|C_t^n|^2 \\ & \quad \le \sum_{s,t\in S^n}^{\neq} \frac{1}{(1-\sup_{\boldsymbol{u},\boldsymbol{v}\in W} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})|C_s^n||C_t^n|)^2} \left(\sup_{\boldsymbol{u},\boldsymbol{v}\in W} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})\right)^2 |C_s^n|^2|C_t^n|^2 \\ & \quad \le \sum_{s,t\in S^n}^{\neq} \frac{1}{(1-\sup_{\boldsymbol{u},\boldsymbol{v}\in W} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})|W|^2)^2} \left(\sup_{\boldsymbol{u},\boldsymbol{v}\in W} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})\right)^2 |C_1^n|^4 \\ & \quad \le \frac{|W|^2}{(1-\sup_{\boldsymbol{u},\boldsymbol{v}\in W} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})|W|^2)^2} \left(\sup_{\boldsymbol{u},\boldsymbol{v}\in W} \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})\right)^2 |C_1^n|^2 \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

The last inequality results from the fact that the number of different pairs in the set S^n is this:

$$\sum_{s,t\in S^n}^{\neq} 1 = \frac{|S_n|(|S_n|-1)}{2} = \frac{|W|}{2|C_1^n|}\left(\frac{|W|}{|C_1^n|} - 1\right) \le \frac{|W|^2}{|C_1^n|^2}.$$

We have shown that the sum of Cauchy remainders converges to zero for partition whose norm tents to zero:

$$\lim_{n\to\infty} \sum_{s,t\in S^n}^{\neq} \log\left(1 - \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) = -\int_W \int_W \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v}) \,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}.$$

- $\sum_{s,t\in S^n}^{\neq} N_{s,t}^n \log\left(1-\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n,\boldsymbol{u}_t^n)|C_s^n||C_t^n|\right)$

The last sum consists only of finitely many non-zero elements. Due to boundedness of the second-order product density the argument of logarithmic function to one when cell size tends to zero. Hence

$$\lim_{n\to\infty}\sum_{s,t\in S^n}^{\neq} N_{s,t}^n \log\left(1-\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n,\boldsymbol{u}_t^n)|C_s^n||C_t^n|\right)=0.$$

We will therefore consider the logarithmic composite likelihood in the form

$$\log\widetilde{CL}(\boldsymbol{\theta})=\sum_{\boldsymbol{u},\boldsymbol{v}\in X}^{\neq}\left(\log\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u},\boldsymbol{v})\right)-\int_W\int_W\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}.$$

4.2.2 Avoiding numerical integration in the CL criterion

Until now, deriving composite likelihood has been analogous to first-order case. But now comes the question of how to maximize this function. There is a multiple integral in the formula we have just derived. This can be potentially very computationally intensive. We will therefore use an estimate to avoid this calculation.

Here we will finally use the assumption of stationarity and isotropy. In the summand containing the double integral, we use the expression second-order product density using correlation function g:

$$\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u},\boldsymbol{v})=g_{\boldsymbol{\theta}}(\boldsymbol{u},\boldsymbol{v})\rho_{\boldsymbol{\theta}}(\boldsymbol{u})\rho_{\boldsymbol{\theta}}(\boldsymbol{v}),$$

using the assumption of stationarity and isotropy we might write

$$\rho_{\boldsymbol{\theta}}^{(2)}\left(\|\boldsymbol{u}-\boldsymbol{v}\|\right)=g_{\boldsymbol{\theta}}\left(\|\boldsymbol{u}-\boldsymbol{v}\|\right)\rho_{\boldsymbol{\theta}}^2.$$

We can combine this with the expression of the correlation function using the K-function using formula $g(r)=\frac{K'(r)}{2\pi r}$ from Lemma 1, so we receive expression

$$\int_W\int_W\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}=\rho_{\boldsymbol{\theta}}^2\int_W\int_W\frac{K_{\boldsymbol{\theta}}'\left(\|\boldsymbol{u}-\boldsymbol{v}\|\right)}{2\pi\|\boldsymbol{u}-\boldsymbol{v}\|}\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}.$$

In addition, the article [Dvořák and Prokešová, 2012, Section 4.2] exhors us to consider only pairs of points that are not too far apart in our calculation. Points that lie too far apart do not to carry information about interaction parameters but they would bias the estimate. Then the relationship implies

$$\int_{W_{\ominus R}}\int_W\mathbf{1}_{[\|\boldsymbol{u}-\boldsymbol{v}\|<R]}\,\rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}=\rho_{\boldsymbol{\theta}}^2\,|W_{\ominus R}|\,K_{\boldsymbol{\theta}(R),\ R>0},$$

where $W_{\ominus R}=\{\boldsymbol{u}\in W:\,b(\boldsymbol{u},R)\subset W\}$. R needs to be chosen smaller than half of width of the observation window. However, we usually choose one quarter of the smaller side of W. Or the corresponding size if the observation window is not rectangular.

Then, for an appropriate fixed $R > 0$, we will look for the parameter $\boldsymbol{\theta}$ as the maximum argument of the simplified composite likelihood function

$$\log \widetilde{\widetilde{CL}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{u},\boldsymbol{v}\in X}^{\neq} \left(\log \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u},\boldsymbol{v})\right) - \rho^2_{\boldsymbol{\theta}} K(R)|W_{\ominus R}|. \tag{4.4}$$

4.3 CL on the univariate LGCP

Let us now look at what this composite likelihood function looks like if we want to use it to estimate parameters in the univariate LGCP model. Then, assuming stationarity and isotropy

$$\begin{aligned}
\log \widetilde{\widetilde{CL}}(\boldsymbol{\theta}) &= \sum_{\boldsymbol{u},\boldsymbol{v}\in X}^{\neq} \left(\log \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u},\boldsymbol{v})\right) - \rho^2_{\boldsymbol{\theta}} K(R)|W_{\ominus R}| \\
&= \sum_{\boldsymbol{u},\boldsymbol{v}\in X}^{\neq} \left(2\mu + \sigma^2 + \sigma^2 \exp\{-\delta\|\boldsymbol{u}-\boldsymbol{v}\|\}\right) \\
&\quad - \exp\{2\mu + \sigma^2\} 2\pi \int_0^r s \exp\left\{\sigma^2 \exp\{-\delta s\}\right\} \mathrm{d}s \, |W_{\ominus R}| \, .
\end{aligned}$$

4.4 CL on the multivariate LGCP

For estimates in a model where there is more than one type of point, we would like to take this fact into account and approach these individual processes in part separately. We will therefore apply a composite likelihood estimate in two steps.

In the first step, we will be interested in parameters that relate to only one type of process, species-specific parameters. So we will apply the composite likelihood function to each process X_i separately. In this way we a priori estimate all the parameters, but we keep only those that influence only one process. We discard the other estimates. Like this, we avoid that the estimation of these parameters is too biased by inter-species interactions. In the case of the multivariate LGCP model, to which we will soon apply the composite likelihood method, it turns out that it is advantageous to choose new parameters as a function of some species-specific and inter-species interaction parameters. Then, the dimensionality of the problem is reduced in this step, and there is no need to discard the estimated parameters at the end. Specifically, it is described in the Section 4.4.1.

In the second step, we will apply the procedure to pairs of points from different processes as well. Our goal will be to estimate inter-species interaction parameters. Since we will already have estimates of species-specific parameters from the previous step, we will insert them into the estimates as a fixed input. In addition, this approach will greatly reduce the dimensionality of the optimization problem.

Let us look at deriving a formula for the second step.

To work with marked processes, we will use the same sequence of partitions that we introduced in the Section 4.2.1. On the pairs of cells of the partition, we define the indicators of occurrence of the points of X_i and X_j, $i \neq j$:

$$N^{n,i,j}_{s,t} = \mathbf{1}_{\left[N_{X_i}(C^n_s)>0 \ \& \ N_{X_j}(C^n_t)>0\right]}.$$

We further define the probabilities of these events and a given parameter $\boldsymbol{\theta} \in \Theta$:

$$p_{s,t}^{n,i,j}(\boldsymbol{\theta}) = \mathbb{P}\left[N_{s,t}^{n,i,j} = 1\right].$$

Then, the composite likelihood function has form

$$L^n(\boldsymbol{\theta}) = \prod_{\substack{i,j=1 \\ i \leq j}}^{p} \prod_{s,t \in S^n}^{\neq} p_{s,t}^{n,i,j}(\boldsymbol{\theta})^{N_{s,t}^{n,i,j}} \left(1 - p_{s,t}^{n,i,j}(\boldsymbol{\theta})\right)^{\left(1 - N_{s,t}^{n,i,j}\right)}$$

We approximate the probability of a pair of points occurring in the respective cells by the multivariate second-order product density $\rho_{i,j,\boldsymbol{\theta}}^{(2)}$:

$$p_{s,t}^{n,i,j}(\boldsymbol{\theta}) \approx \rho_{i,j,\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|.$$

Again, we need the assumption of continuity, integrability and boundedness of the multivariate second-order product density.

When we move to log composite likelihood function, we get the formula

$$\begin{aligned}
\log CL^n(\boldsymbol{\theta}) = & \sum_{\substack{i,j=1 \\ i \leq j}}^{p} \left(\sum_{s,t \in S^n}^{\neq} N_{s,t}^{n,i,j} \log \rho_{i,j,\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n) + \sum_{s,t \in S^n}^{\neq} N_{s,t}^{n,i,j} \log\left(|C_s^n||C_t^n|\right) \right. \\
& + \sum_{s,t \in S^n}^{\neq} \log\left(1 - \rho_{i,j,\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \\
& \left. - \sum_{s,t \in S^n}^{\neq} N_{s,t}^{n,i,j} \log\left(1 - \rho_{i,j,\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}_s^n, \boldsymbol{u}_t^n)|C_s^n||C_t^n|\right) \right).
\end{aligned}$$

We will no longer derive in detail all the limit and approximation relationships that will lead to the following composite likelihood. It would be a direct analogy of the unmarked case.

$$\log \widetilde{CL}(\boldsymbol{\theta}) = \sum_{\substack{i,j=1 \\ i \leq j}}^{p} \left(\sum_{\substack{\boldsymbol{u} \in X_i, \\ \boldsymbol{v} \in X_j}} \left(\log \rho_{i,j,\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})\right) - \int_W \int_W \rho_{i,j,\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v}) \, \mathrm{d}\boldsymbol{u} \, \mathrm{d}\boldsymbol{v} \right).$$

We will simplify this formula further for the purposes of easier calculations. We will use the cross K-function K_{ij}, its relations with cross pair correlation function g_{ij} given in Section 1.4 and assumption of stationarity and isotropy, which means $\rho_{i,\boldsymbol{\theta}}(\boldsymbol{u}) = \rho_{i,\boldsymbol{\theta}} \ \forall \boldsymbol{u} \in X_i, \ i = 1, \dots, p$. So we get formula

$$\log \widetilde{\widetilde{CL}}(\boldsymbol{\theta}) = \sum_{\substack{i,j=1 \\ i \leq j}}^{p} \left(\sum_{\substack{\boldsymbol{u} \in X_i, \\ \boldsymbol{v} \in X_j}} \left(\log \rho_{\boldsymbol{\theta}}^{(2)}(\boldsymbol{u}, \boldsymbol{v})\right) - \rho_{i,\boldsymbol{\theta}} \rho_{j,\boldsymbol{\theta}} K_{ij}(R)|W_{\ominus R}| \right). \tag{4.5}$$

4.4.1 second-order CL in multivariate LGCP

Let us apply the derived method to the multivariate LGCP model introduced in Section 3. As we have outlined, we will proceed in two steps. However, we will still have to simplify our model a bit to be able to estimate its parameters. The current model has so many parameters whose effects are intertwined that we have trouble to distinguish them from each other.

In addition, we will use one more idea, which is represented by [Kopecký and Mrkvička, 2016, Section 3.2], for example. This is the use of intensity estimate beyond the composite likelihood. We get it as the ratio of the number of points to the size of the observation window, i.e.

$$\widehat{\rho_{i,\boldsymbol{\theta}_i}} = \frac{N_{X_i}(W)}{|W|}.$$

Model simplification

We will assume that the internal exponential covariance structure of the Gaussian fields $\boldsymbol{U}$ and $\boldsymbol{Y}$ across all marginals is given by the same parameter. Specifically, we introduce parameter δ such that

$$\forall\ i = 1, \dots, p\ \forall\ l = 1, \dots, q: \ \ \delta = \delta_i = \epsilon_l.$$

To estimate this parameter correctly, we will not maximize the composite likelihood functions separately for each species in the first step, but will maximize their sum

$$\sum_{i=1}^{p} \log \widetilde{\widetilde{CL}}_i(\boldsymbol{\theta}_i).$$

Species-specific parameters

In this step we will use also reparametrization. Let us introduce a new parameter $\nu = (\nu_1, \dots, \nu_p)^\top$ as

$$\nu_i = \sigma_i^2 + \sum_{l=1}^{q} a_{il}^2.$$

Then, recalling the assumption of stationarity and isotropy, we can express the intensity and second-order product density as follows:

$$\rho_{i,\boldsymbol{\theta}_i} = \rho_{i,\boldsymbol{\theta}_i}(\boldsymbol{u}) = \exp\left\{\mu_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sum_{l=1}^{q} a_{il}^2\right\} = \exp\left\{\mu_i + \frac{1}{2}\nu_i\right\} \tag{4.6}$$

and

$$\begin{aligned}\rho^{(2)}_{i,i,\boldsymbol{\theta}_i}(h) = \rho^{(2)}_{i,i,\boldsymbol{\theta}_i}(\boldsymbol{u}, \boldsymbol{v}) &= \exp\Bigg\{2\mu_i + \sigma_i^2 + \sum_{l=1}^{q} a_{il}^2 + \sigma_i^2 \exp\{-\delta_i h\} \\ &\qquad + \sum_{l=1}^{q} a_{il}^2 \exp\{-\epsilon_l h\}\Bigg\} \\ &= \exp\left\{2\mu_i + \nu_i + \nu_i \exp\{-\delta h\}\right\},\end{aligned} \tag{4.7}$$

where $h = \|\boldsymbol{u} - \boldsymbol{v}\|$.

In addition, thanks to the intensity estimate, we can express $\mu_i = \log \rho_{i,\boldsymbol{\theta}_i} - \frac{1}{2}\nu_i$ and therefore plug it in the composite likelihood function in form

$$\mu_i = \log \frac{N_{X_i}(W)}{|W|} - \frac{1}{2}\nu_i.$$

Therefore, in the first step we will search for the parameters $\nu_1, \ldots, \nu_p, \delta$.

The exact form of logarithmic composite likelihood is obtained using the intensity function and the second-order product density.

Note that even in this case there is no problem with integrability or boundedness of functions $\rho_{i,\boldsymbol{\theta}_i}$ and $\rho_{i,\boldsymbol{\theta}_j}$, so we can use the previously derived composite likelihood. By plugging into the (4.5) we get the exact form for our model

$$\begin{aligned}\log \widetilde{\widetilde{CL}}_i(\boldsymbol{\theta}_i) = &\sum_{\boldsymbol{u},\boldsymbol{v}\in X_i}^{\neq} \left(2\mu_i + \sigma_i^2 + \sum_{l=1}^{q} a_{il}^2 + \sigma_i^2 \exp\{-\delta_i h\} + \sum_{l=1}^{q} a_{il}^2 \exp\{-\epsilon_l h\}\right) \\ &- |W_{\ominus R}| \exp\left\{2\mu_i + \sigma_i^2 + \sum_{l=1}^{q} a_{il}^2\right\} \\ &\cdot 2\pi \int_0^r s \exp\left\{\sigma_i^2 \mathrm{e}^{-\delta_i s} + \sum_{l=1}^{q} a_{il}^2 \mathrm{e}^{-\epsilon_l s}\right\} \mathrm{d}s\end{aligned}$$

which for our simplified model means

$$\begin{aligned}\log \widetilde{\widetilde{\widetilde{CL}}}_i(\boldsymbol{\theta}_i) = &\sum_{\boldsymbol{u},\boldsymbol{v}\in X_i}^{\neq} \left(2\left(\log \frac{N_{X_i}(W)}{|W|} - \frac{1}{2}\nu_i\right) + \nu_i + \nu_i \exp\{-\delta h\}\right) \\ &- |W_{\ominus R}| \exp\left\{2\left(\log \frac{N_{X_i}(W)}{|W|} - \frac{1}{2}\nu_i\right) + \nu_i\right\} \\ &\cdot 2\pi \int_0^r s \exp\left\{\nu_i \exp\{-\delta h\}\right\} \mathrm{d}s \\ = &\sum_{\boldsymbol{u},\boldsymbol{v}\in X_i}^{\neq} \left(2\log \frac{N_{X_i}(W)}{|W|} + \nu_i \exp\{-\delta h\}\right) \\ &- |W_{\ominus R}| \exp\left\{2\log \frac{N_{X_i}(W)}{|W|}\right\} 2\pi \int_0^r s \exp\left\{\nu_i \exp\{-\delta s\}\right\} \mathrm{d}s,\end{aligned}$$

where $h = \|\boldsymbol{u} - \boldsymbol{v}\|$.

Therefore, we will maximize the function

$$\begin{aligned}\sum_{i=1}^{p} \log \widetilde{\widetilde{\widetilde{CL}}}_i(\boldsymbol{\theta}_i) = &\sum_{i=1}^{p} \left(\sum_{\boldsymbol{u},\boldsymbol{v}\in X_i}^{\neq} \left(2\log \frac{N_{X_i}(W)}{|W|} + \nu_i \exp\{-\delta h\}\right)\right. \\ &\left. - |W_{\ominus R}| \exp\left\{2\log \frac{N_{X_i}(W)}{|W|}\right\} 2\pi \int_0^r s \exp\left\{\nu_i \exp\{-\delta s\}\right\} \mathrm{d}s\right).\end{aligned}$$

When we get estimates $\widehat{\nu_i}$ and $\widehat{\delta}$, we can also estimate the parameter μ_i as

$$\widehat{\mu_i} = \log \frac{N_{X_i}(W)}{|W|} - \frac{1}{2}\widehat{\nu_i}.$$

Inter-species interaction parameters

In this step, we want to include the relationships between the processes X_i and X_j in the estimation of parameters $\mathbb{A}$. We will now take the parameters δ, ν_i and $\mu_i, i = 1, \dots, p$, we firstly estimated as given.

We will now derive how exactly the respective composite likelihood function will look like. We will consider a vector of unknown parameters $\boldsymbol{\theta} = (a_{11}, \dots, a_{pq})$.

In order to express this formula (4.5) for our model, let us recall that multivariate second-order product density for points from of type i and j using the assumption of stationarity and isotropy:

$$\rho_{i,j}^{(2)}(\boldsymbol{u}, \boldsymbol{v}) = \exp\left\{\mu_i + \mu_j + \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) + \frac{1}{2}\sum_{k=1}^{q}\left(a_{ik}^2 + a_{jk}^2\right)\right. \\ \left. + \mathbf{1}_{[i=j]}\sigma_i^2 \exp\{-\delta_i \|\boldsymbol{u}-\boldsymbol{v}\|\} + \sum_{k=1}^{q} a_{ik}a_{jk}\exp\{-\epsilon_k\|\boldsymbol{u}-\boldsymbol{v}\|\}\right)\Bigg\}. \tag{4.8}$$

Plugging into the formula for composite likelihood function:

$$\log \widetilde{\widetilde{CL}}(\boldsymbol{\theta}) = \sum_{\substack{i,j=1\\ i\le j}}^{p}\left(\sum_{\substack{\boldsymbol{u}\in X_i,\\ \boldsymbol{v}\in X_j}}\left(\mu_i + \mu_j + \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2 + \sum_{k=1}^{q} a_{ik}^2 + \sum_{k=1}^{q} a_{jk}^2\right)\right.\right. \\ \left. + \mathbf{1}_{[i=j]}\sigma_i^2 \exp\{-\delta_i\|\boldsymbol{u}-\boldsymbol{v}\|\} + \sum_{k=1}^{q} a_{ik}a_{jk}\exp\{-\epsilon_k\|\boldsymbol{u}-\boldsymbol{v}\|\}\right) \\ - \exp\left\{\mu_i + \mu_j + \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) + \frac{1}{2}\sum_{l=1}^{q}\left(a_{il}^2 + a_{jl}^2\right)\right\} \\ \left. \cdot 2\pi \int_0^r s\exp\left\{\mathbf{1}_{[i=j]}\sigma_i^2 \mathrm{e}^{-\delta_i s} + \sum_{l=1}^{q} a_{il}a_{jl}\mathrm{e}^{-\epsilon_l s}\right\}\mathrm{d}s |W_{\ominus R}|\right).$$

Using the reparametrization relation $\sigma_i^2 = \nu_i - \sum_{i=l}^{q} a_{il}^2$ we receive the function to maximize. Let us recall that we will maximize it only for the parameters $a_{11}, \dots, a_{pq}$, the others have fixed values obtained by the estimates in the previous step.

$$
\begin{aligned}
\log \widetilde{\widetilde{CL}}(\boldsymbol{\theta}) = \sum_{\substack{i,j=1\\ i\leq j}}^{p} \Bigg(& \sum_{\substack{\boldsymbol{u}\in X_i,\\ \boldsymbol{v}\in X_j}} \Bigg(\mu_i + \mu_j + \frac{1}{2}\nu_i + \frac{1}{2}\nu_j \\
& + \mathbf{1}_{[i=j]} \left(\nu_i - \sum_{i=l}^{q} a_{il}^2 \right) \exp\{-\delta\|\boldsymbol{u}-\boldsymbol{v}\|\} \\
& + \sum_{k=1}^{q} a_{ik}a_{jk} \exp\left\{-\delta\|\boldsymbol{u}-\boldsymbol{v}\|\right\} \Bigg) \\
& - \exp\left\{ \mu_i + \mu_j + \frac{1}{2}\nu_i + \frac{1}{2}\nu_j \right\} \\
& \cdot 2\pi \int_0^r s \exp\left\{ \mathbf{1}_{[i=j]} \left(\nu_i - \sum_{i=l}^{q} a_{il}^2 \right) \mathrm{e}^{-\delta s} + \sum_{l=1}^{q} a_{il}a_{jl}\mathrm{e}^{-\delta s} \right\} \mathrm{d}s |W_{\ominus R}| \Bigg)
\end{aligned}
$$

Finally, we can estimate the last of the parameters, the parameter σ_i^2 as

$$
\widehat{\sigma_i^2} = \widehat{\nu_i} - \sum_{i=l}^{q} \widehat{a_{il}}^2. \tag{4.9}
$$

Thus, we have obtained estimates of all parameters from the original parameterization of the model.

5. Guan's composite likelihood

At present, if we wanted to find the parameters $\boldsymbol{\theta}$ for the log-Gaussian Cox process using a moment-based method, we would probably turn to the composite likelihood derived by Yongtao Guan in his article [Guan, 2006, Section 2].

In order to determine the accuracy of both methods, Guan's and the one in the Section 4 we will compare them on simulated data, first from a univariate log-Gaussian Cox process, and then from the extended multivariate log-Gaussian Cox process we presented in Section 3.

5.1 Composite likelihood by Guan

Contrary to the approach we demonstrate here, Guan's method is based on the idea that the density for two events from stationary and isotropic point process X can be expressed as the second-order product density, $\rho^{(2)}_{\boldsymbol{\theta}}$. It is the density describing the occurrence of a pair of points of the process at locations $\boldsymbol{u}$ and $\boldsymbol{v}$. In order to talk about density, we must of course standardize it. So

$$f(\boldsymbol{u},\boldsymbol{v}) = \frac{\rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}-\boldsymbol{v})}{\int_W \int_W \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}}, \ \ \boldsymbol{u},\boldsymbol{v} \in X.$$

Therefore, if we consider the product of such partial densities for each pair of points, we obtain a composite likelihood function. An introduction to this approach can be found in Lindsay [1988]. In addition, when we add the idea that only pairs of points to the distance R carry significant information, where R makes sense to think at most equal to quarter the shorter side of the observation window W, for $\theta \in \Theta$ we get the expression

$$L(\boldsymbol{\theta}) = \prod_{\substack{\boldsymbol{u}\neq\boldsymbol{v}\in X\cap W \\ \|\boldsymbol{u}-\boldsymbol{v}\|<R}} \frac{\rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}-\boldsymbol{v})}{\int_W \int_W \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}}.$$

From here we can move on to log composite likelihood function:

$$\log CL(\boldsymbol{\theta}) = \sum_{\substack{\boldsymbol{u}\boldsymbol{v}\in X\cap W \\ \|\boldsymbol{u}-\boldsymbol{v}\|<R}} \log \frac{\rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}-\boldsymbol{v})}{\int_W \int_W \rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}}.$$

The article [Dvořák and Prokešová, 2012, Section 4.2] then suggests to use the modification of this composite likelihood function, using the K-function and estimate the parameters from formula

$$\log \widetilde{CL}(\boldsymbol{\theta}) = \sum_{\substack{0<\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X\cap W_{\ominus R} \\ \boldsymbol{v}\in X\cap W}} \log \frac{\rho^{(2)}_{\boldsymbol{\theta}}(\boldsymbol{u}-\boldsymbol{v})}{\rho^2_{\boldsymbol{\theta}}|W_{\ominus R}|K(R)},$$

where $\rho_{\boldsymbol{\theta}}$ is according to our standard notation intensity function, K denotes K-function and $W_{\ominus R}$ denotes the observation window reduced by the edge of width R. We thus avoid the need to calculate the four-dimensional integral, which significantly reduces the computational complexity.

5.2 Guan's CL on univariate LGCP

If we substitute the form of characteristics for univariate LGCP from the Section 2.2 into this formula for composite likelihood, we get the following form.

$$
\begin{aligned}
\log \widetilde{CL}(\boldsymbol{\theta}) &= \sum_{\substack{0<\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X\cap W_{\ominus R} \\ \boldsymbol{v}\in X\cap W}} \log \frac{\rho_{\boldsymbol{\theta}}^2 \exp\left\{c(\boldsymbol{u},\boldsymbol{v})\right\}}{\rho_{\boldsymbol{\theta}}^2 |W_{\ominus R}| K(R)} \\
&= \sum_{\substack{0<\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X\cap W_{\ominus R} \\ \boldsymbol{v}\in X\cap W}} \log \frac{\exp\left\{\sigma^2 \exp\{-\delta\|\boldsymbol{u}-\boldsymbol{v}\|\}\right\}}{|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\sigma^2 \exp\{-\delta s\}\right\} \mathrm{d}s} \\
&= \sum_{\substack{0<\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X\cap W_{\ominus R} \\ \boldsymbol{v}\in X\cap W}} \left(\sigma^2 \exp\{-\delta\|\boldsymbol{u}-\boldsymbol{v}\|\}\right) \\
&\qquad - N_{X(2)}^R \log\left(|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\sigma^2 \exp\{-\delta s\}\right\} \mathrm{d}s\right),
\end{aligned}
$$

where $N_{X(2)}^R$ denotes number of pairs of different points $(\boldsymbol{u}, \boldsymbol{v})$ such that $\boldsymbol{u} \in X \cap W_{\ominus R}$, $\boldsymbol{v} \in X \cap W$ and $\|\boldsymbol{u}-\boldsymbol{v}\| < R$.

Note that we do not estimate parameter μ in this procedure. We will not obtain it using composite likelihood, but from estimating the intensity function as the ratio of the number observed to the size of the observation window: $\widehat{\rho_{\boldsymbol{\theta}}} = \frac{N_X(W)}{|W|}$. Then we express the estimate of μ as from the intensity calculation formula (2.1), so

$$
\widehat{\mu} = \log \widehat{\rho_{\boldsymbol{\theta}}} - \frac{\widehat{\sigma^2}}{2}.
$$

5.3 Guan's CL on multivariate LGCP

Now we want to apply Guan's composite likelihood to the multivariate LGCP model we are studying.

As in the previous section, to use Guan's composite likelihood, we simplify the model by substituting a single parameter into the exponential covariance structures. So, for $i = 1, \ldots, p$ and $l = 1, \ldots, q$:

$$
\delta = \delta_i = \epsilon_l.
$$

And we will also take advantage of substituting

$$
\nu_i = \sigma_i^2 + \sum_{l=1}^{q} a_{il}^2.
$$

As in the case of the previous composite likelihood function, we will now proceed in two steps. Firstly, we estimate parameters that depend on only one type of the process. In the second step, we will use these estimates as input for estimating interaction parameters.

Species-specific parameters

Since the intensity for stationary isotropic process equals to (4.6) and the second-order product density is (4.7), the marginal composite likelihood with unknown parameters $\boldsymbol{\theta}_i = (\nu_i, \delta)$ equals to

$$
\begin{aligned}
\log \widetilde{CL}_i(\boldsymbol{\theta}_i) &= \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X_i\cap W_{\ominus R} \\ \boldsymbol{v}\in X_i\cap W}} \log \frac{\rho^{(2)}_{i,i,\boldsymbol{\theta}_i}(h)}{\rho^2_{i,\boldsymbol{\theta}_i}|W_{\ominus R}|K_i(R)} \\
&= \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X_i\cap W_{\ominus R} \\ \boldsymbol{v}\in X_i\cap W}} \log \frac{\exp\left\{\sigma_i^2 \exp\{-\delta_i h\} + \sum_{l=1}^q a_{il}^2 \exp\{-\epsilon_l h\}\right\}}{|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\sigma_i^2 \mathrm{e}^{-\delta_i s} + \sum_{l=1}^q a_{il}^2 \mathrm{e}^{-\epsilon_l s}\right\} \mathrm{d}s} \\
&= \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X_i\cap W_{\ominus R} \\ \boldsymbol{v}\in X_i\cap W}} \left(\sigma_i^2 \exp\{-\delta_i h\} + \sum_{l=1}^q a_{il}^2 \exp\{-\epsilon_l h\}\right) \\
&\qquad - N_i^R \log\left(|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\sigma_i^2 \mathrm{e}^{-\delta_i s} + \sum_{l=1}^q a_{il}^2 \mathrm{e}^{-\epsilon_l s}\right\} \mathrm{d}s\right) \\
&= \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X_i\cap W_{\ominus R} \\ \boldsymbol{v}\in X_i\cap W}} (\nu_i \exp\{-\delta h\}) \\
&\qquad - N_i^R \log\left(|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\nu_i \exp\{-\delta s\}\right\} \mathrm{d}s\right),
\end{aligned}
$$

where N_i^R denotes the number of different pairs of process points X_i limited to the reduced window $W_{\ominus R}$, i.e. the elements of the set $\{(\boldsymbol{u}, \boldsymbol{v});\ 0 < \|\boldsymbol{u} - \boldsymbol{v}\| < R,\ \boldsymbol{u} \in X_i \cap W_{\ominus R}, \boldsymbol{v} \in X_i\}$.

So, overall, we want to maximize function

$$
\begin{aligned}
\sum_{i=1}^p \log \widetilde{CL}_i(\boldsymbol{\theta}_i) = &\sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R \\ \boldsymbol{u}\in X_i\cap W_{\ominus R} \\ \boldsymbol{v}\in X_i\cap W}} (\nu_i \exp\{-\delta h\}) \\
&\qquad - N_i^R \log\left(|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\nu_i \exp\{-\delta s\}\right\} \mathrm{d}s\right),
\end{aligned}
$$

Then, from the resulting estimate of ν_i, we receive estimate of μ_i as

$$\widehat{\mu_i} = \log \frac{N_{X_i}(W)}{|W|} - \frac{1}{2}\widehat{\nu_i}.$$

Inter-species interaction parameters

In the next step we will estimate the remaining parameters, specifically elements of matrix $\mathbb{A}$. From these and using the estimate of ν_i, we can express the estimate of σ_i^2 then. So now we will work with $\boldsymbol{\theta} = \{a_{11}, \ldots, a_{pq}\}$. We will use second-order product density in the form (4.8) and relation $\sigma_i^2 = \nu_i - \sum_{i=l}^q a_{il}^2$

$$
\begin{aligned}
\log \widetilde{CL}(\boldsymbol{\theta}) &= \sum_{\substack{i,j=1\\ i\leq j}}^{p} \widetilde{CL}_{i,j}(\boldsymbol{\theta}) = \sum_{\substack{i,j=1\\ i\leq j}}^{p} \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R\\ \boldsymbol{u}\in X_i\cap W_{\ominus R}\\ \boldsymbol{v}\in X_j\cap W}} \log \frac{\rho^{(2)}_{i,j,\boldsymbol{\theta}}(h)}{\rho_{i,\boldsymbol{\theta}}\rho_{j,\boldsymbol{\theta}}|W_{\ominus R}|K_{ij}(R)} \\
&= \sum_{\substack{i,j=1\\ i\leq j}}^{p} \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R\\ \boldsymbol{u}\in X_i\cap W_{\ominus R}\\ \boldsymbol{v}\in X_j\cap W}} \log \frac{\exp\left\{\mathbf{1}_{[i=j]}\sigma_i^2 \mathrm{e}^{-\delta_i h} + \sum_{l=1}^{q} a_{il}a_{jl}\mathrm{e}^{-\epsilon_l s}\right\}}{|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\mathbf{1}_{[i=j]}\sigma_i^2 \mathrm{e}^{-\delta_i h} + \sum_{l=1}^{q} a_{il}a_{jl}\mathrm{e}^{-\epsilon_l s}\right\} \mathrm{d}s} \\
&= \sum_{\substack{i,j=1\\ i\leq j}}^{p} \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R\\ \boldsymbol{u}\in X_i\cap W_{\ominus R}\\ \boldsymbol{v}\in X_j\cap W}} \left(\mathbf{1}_{[i=j]}\sigma_i^2 \exp\{-\delta_i h\} + \sum_{l=1}^{q} a_{il}a_{jl} \exp\{-\epsilon_l h\}\right) \\
&\quad - N_{ij}^R \log\left(|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\mathbf{1}_{[i=j]}\sigma_i^2 \mathrm{e}^{-\delta_i s} + \sum_{l=1}^{q} a_{il}a_{jl}\mathrm{e}^{-\epsilon_l s}\right\} \mathrm{d}s\right) \\
&= \sum_{\substack{i,j=1\\ i\leq j}}^{p} \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R\\ \boldsymbol{u}\in X_i\cap W_{\ominus R}\\ \boldsymbol{v}\in X_j\cap W}} \left(\mathbf{1}_{[i=j]}\left(\nu_i - \sum_{i=l}^{q} a_{il}^2\right) \exp\{-\delta h\} + \sum_{l=1}^{q} a_{il}a_{jl} \exp\{-\delta h\}\right) \\
&\quad - N_{ij}^R \log\left(|W_{\ominus R}| 2\pi \int_0^r s \exp\left\{\mathbf{1}_{[i=j]}\left(\nu_i - \sum_{i=l}^{q} a_{il}^2\right) \mathrm{e}^{-\delta s} + \sum_{l=1}^{q} a_{il}a_{jl}\mathrm{e}^{-\delta s}\right\} \mathrm{d}s\right),
\end{aligned}
$$

where N_{ij}^R indicates the appropriate number of point pairs, specifically $N_{ij}^R = \sum_{\substack{0<h=\|\boldsymbol{u}-\boldsymbol{v}\|<R\\ \boldsymbol{u}\in X_i\cap W_{\ominus R}\\ \boldsymbol{v}\in X_j\cap W}} 1$.

And very lastly, we can calculate estimate of σ_i^2 as

$$
\widehat{\sigma_i^2} = \widehat{\nu}_i - \sum_{i=l}^{q} \widehat{a_{il}}^2.
$$

6. Comparison of composite likelihoods

As we have already outlined, we would now like to see how well our composite likelihood can estimate the parameters of the LGCP model. On the simulated data, we compare the method presented by us in Chapter 4 with Guan's composite likelihood function as set in Chapter 5.

We perform the comparison on both univariate and multivariate log-Gaussian Cox point processes.

We will determine the distance of our estimates from the true value of the parameters according to two measures, relative bias and relative mean squared error. We define the relative bias of the parameter θ in n simulations as

$$\mathrm{relBias}\,(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{\theta}_i - \theta_0}{|\theta_0|}$$

and the mean squared error is defined as

$$\mathrm{relMSE}\,(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\widehat{\theta}_i - \theta_0\right)^2}{\theta_0^2}$$

where $\widehat{\theta}_i$ is estimate of the parameter θ in the i-th simulation and θ_0 is its true value.

Due to the presence of non-negligible number of outlying estimates, we decided not to calculate these characteristics from all estimates. Instead, we do not use 5% of the simulations those relative mean squared error would be the highest. Even in practical applications, we would not hesitate to discard measurements that would be obvious to be outliers. The results from the full range of simulations can be found in Appendix A. In order not to confuse these characteristics, we will be the one for whom we have chosen 95% of the best estimates denote by $\mathrm{relBias}_{95}$, resp. relMSE_{95}.

To simulate data from appropriate models, we used the `spatstat package` for `R` (Baddeley et al. [2015]). Optimization of the composite likelihood functions was also performed using the `R` (R Core Team [2017]), specifically using the `optim` function based on the Nelder-Mead algorithm. The scripts can be found in the electronic appendix of this thesis.

6.1 Univariate LGCP

First, let us look at the simulation in the univariate LGCP model. There were 100 runs, all with the choice $R = 0.25$. We see their results in terms of relative bias and relative mean squared error in Table 6.1. These results are computed without the five percent worst realizations. The notation ρ_0, σ_0^2 and δ_0 indicates that the estimates in the right part of the table were calculated from data simulated from a univariate LGCP model with these true parameter values.

We can observe that both methods estimate the parameter μ reasonably well, but it is interesting to observe that while our method tends to overestimate it,

			Our CL						Guan's CL					
			relBias$_{95}$			relMSE$_{95}$			relBias$_{95}$			relMSE$_{95}$		
ρ_0	σ_0^2	δ_0	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$
150	1	10	0.14	-0.04	-0.39	0.03	0.01	17.46	-0.13	-0.12	0.11	0.02	0.56	0.27
	0.5	10	0.09	0.90	-1.00	0.01	0.83	1.00	-0.09	-0.22	0.06	0.01	0.46	0.06
250	1	10	0.14	-0.08	0.77	0.03	0.03	44.13	-0.17	-0.17	0.08	0.03	0.38	0.21
	0.5	10	0.08	0.85	-1.00	0.01	0.78	1.00	-0.17	-0.15	0.10	0.03	0.38	0.06

Table 6.1: Relative biases and relative mean squared errors of the estimates of the univariate LGCP computed from 95% of the best simulations. Using $R = 0.25$.

Guan's method underestimates it. This is due to the use of non-parametric estimation.

We also find the estimation of σ^2 to be satisfactory for Guan's method. However, our method estimated it with sufficiently small error only when its true value was 1. When it decreased, our method was no longer able to estimate the parameter well.

Guan's method estimates the parameter δ quite well, at least for higher intensities. On the other hand, our method failed here. In many cases it tends to estimate the delta very close to zero, and in the case of smaller σ^2 it estimates practically equal to zero in all realizations. Overall, the quality of the estimates deteriorates as the intensity decreases. If the actual value of σ^2 is low, then it is harder to estimate the value of δ, since the effect of both parameters on the exact configuration of the points is much weaker than with high σ^2. Guan's method turns out to be more sensitive than ours in this respect and it is therefore possible to distinguish this effect even at lower σ^2.

6.2 Multivariate LGCP

For the multivariate LGCP model, we first had to decide how large model to choose. Based on the experience from the univariate model, we decided to examine the results on as small structure as possible to have a chance of obtaining relevant estimates. Thus, we used a model with two types of points, thus we chose $p = 2$, and also with two environmental effects, i.e. $q = 2$.

We also simulated the model 100 times. Again, we use the option $R = 0.25$. Table 6.2 offers a comparison of the results in terms of relative bias, while Table 6.3 talks about the relative mean squared error. Similarly to the univariate approach, we use the notation ρ_0, σ_0^2, δ_0 and $\mathbb{A}_0$ to indicate the true values of the parameters indicating the simulations from which the estimates were computed.

Note that both methods estimate the parameter $\mu = (\mu_1, \mu_2)$ very well. This good estimate is due to the fact that both CL methods estimate μ in a non-parametric way. Our method seems to work even a little better than Guan's. Similarly to the univariate LGCP model, our method tends to underestimate μ, while Guan's tends to do the opposite in most cases.

Another similarity with the univariate case is shown by the estimation of the parameter δ. Guan's method has no significant problems with it. While ours strongly underestimates it, essentially failing to estimate it at all in certain cases.

Similar conclusions can be drawn for the estimates of the elements of the matrix $\mathbb{A}$ if both its non-diagonal elements are positive. Guan's method can determine them satisfactorily well. However, our method strongly underestimates the elements on the diagonal in the vast majority of cases. Even though the non-diagonal parameters are estimated quite satisfactorily, the negative effect of the estimates of the matrix $\mathbb{A}$ on the estimation of σ^2 cannot be denied. It turns out that in the case of biased estimates of the elements of $\mathbb{A}$, computing the estimate of σ^2 from the formula (4.9) does not ensure that this estimate is positive. In cases where one of the elements of the matrix $\mathbb{A}$ is negative, Guan's method also fails. Neither method is able to identify the negativity of this parameter. However, the corresponding estimate of σ^2 is not negatively affected, since the elements of $\mathbb{A}$ are squared in its calculation.

					relBias$_{95}$			
ρ_0	σ_0^2	δ_0	$\mathbb{A}_0$	method	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\widehat{\mathbb{A}}$
(200, 400)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.05, 0.09)	(-0.53, -13.12)	-0.61	$\begin{pmatrix} 0.06 & 0.24 \\ 0.26 & 1.90 \end{pmatrix}$
				Guan's	(0.08, -0.18)	(-0.30, 1.55)	0.16	$\begin{pmatrix} -0.65 & 0.30 \\ 0.16 & -0.56 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.00, -0.03)	(0.95, 0.54)	-0.97	$\begin{pmatrix} -0.96 & 2.30 \\ 0.18 & -0.86 \end{pmatrix}$
				Guan's	(0.06, 0.04)	(-0.98, -1.19)	0.17	$\begin{pmatrix} 0.15 & 2.21 \\ 0.19 & 0.16 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.06, -0.07)	(2.38, -1.27)	-0.92	$\begin{pmatrix} -0.86 & 0.30 \\ 0.18 & -0.67 \end{pmatrix}$
				Guan's	(0.03, 0.04)	(-1.71, -1.97)	0.31	$\begin{pmatrix} 0.19 & 0.20 \\ 0.20 & 0.19 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.07, -0.05)	(2.52, -0.46)	-0.91	$\begin{pmatrix} -0.90 & 2.30 \\ 0.18 & -0.71 \end{pmatrix}$
				Guan's	(0.02, 0.04)	(-1.50, -1.74)	0.27	$\begin{pmatrix} 0.15 & 2.20 \\ 0.19 & 0.15 \end{pmatrix}$
(300, 300)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.02, -0.01)	(0.99, 0.25)	-0.98	$\begin{pmatrix} -0.98 & 0.30 \\ 0.18 & -0.83 \end{pmatrix}$
				Guan's	(0.04, 0.08)	(-0.93, -1.18)	0.23	$\begin{pmatrix} 0.19 & 0.20 \\ 0.20 & 0.18 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.01, -0.02)	(0.93, 0.53)	-1.00	$\begin{pmatrix} -0.95 & -.30 \\ 0.18 & -0.87 \end{pmatrix}$
				Guan's	(0.06, 0.05)	(-1.06, -1.22)	0.24	$\begin{pmatrix} 0.17 & 2.20 \\ 0.20 & 0.17 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.05, -0.06)	(2.65, -0.48)	-0.97	$\begin{pmatrix} -0.90 & 0.30 \\ 0.18 & -0.72 \end{pmatrix}$
				Guan's	(0.03, 0.05)	(-1.76, -2.15)	0.28	$\begin{pmatrix} 0.19 & 0.20 \\ 0.20 & 0.20 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.05, -0.07)	(2.61, -0.86)	-1.00	$\begin{pmatrix} -0.89 & 2.30 \\ 0.18 & -0.69 \end{pmatrix}$
				Guan's	(0.04, 0.03)	(-1.51, -1.78)	0.24	$\begin{pmatrix} 0.17 & 2.20 \\ 0.20 & 0.15 \end{pmatrix}$

Table 6.2: Relative biases of the multivariate LGCP computed from 95% of the best simulations. Using $R = 0.25$.

					relMSE$_{95}$			
ρ_0	σ_0^2	δ_0	$\mathbb{A}_0$	method	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\widehat{\mathbb{A}}$
(200, 400)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.02)	(4.64, 4.92)	0.60	$\begin{pmatrix} 0.26 & 0.66 \\ 1.39 & 0.06 \end{pmatrix}$
				Guan's	(0.03, 0.08)	(1.00, 9.21)	0.11	$\begin{pmatrix} 0.54 & 0.09 \\ 0.03 & 0.76 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(1.20, 2.52)	0.97	$\begin{pmatrix} 0.98 & 5.29 \\ 0.03 & 1.08 \end{pmatrix}$
				Guan's	(0.03, 0.01)	(2.08, 2.19)	0.11	$\begin{pmatrix} 0.09 & 4.87 \\ 0.04 & 0.08 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(7.75, 298.14)	0.91	$\begin{pmatrix} 0.91 & 0.09 \\ 0.03 & 2.38 \end{pmatrix}$
				Guan's	(0.01, 0.01)	(5.88, 7.49)	0.21	$\begin{pmatrix} 0.04 & 0.04 \\ 0.04 & 0.05 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(8.66, 218.95)	0.91	$\begin{pmatrix} 0.92 & 5.30 \\ 0.03 & 2.06 \end{pmatrix}$
				Guan's	(0.01, 0.01)	(6.14, 6.45)	0.14	$\begin{pmatrix} 0.08 & 4.86 \\ 0.04 & 0.09 \end{pmatrix}$
(300, 300)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.00, 0.01)	(0.99, 12.73)	0.98	$\begin{pmatrix} 0.97 & 0.09 \\ 0.03 & 1.30 \end{pmatrix}$
				Guan's	(0.02, 0.02)	(2.23, 2.51)	0.16	$\begin{pmatrix} 0.05 & 0.04 \\ 0.04 & 0.05 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(1.05, 2.37)	1.00	$\begin{pmatrix} 0.98 & 5.29 \\ 0.03 & 1.07 \end{pmatrix}$
				Guan's	(0.02, 0.01)	(2.16, 2.42)	0.17	$\begin{pmatrix} 0.06 & 4.86 \\ 0.04 & 0.08 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(7.71, 232.52)	0.97	$\begin{pmatrix} 0.91 & 0.09 \\ 0.03 & 2.07 \end{pmatrix}$
				Guan's	(0.01, 0.01)	(6.09, 6.80)	0.19	$\begin{pmatrix} 0.05 & 0.04 \\ 0.04 & 0.05 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(7.92, 238.05)	1.00	$\begin{pmatrix} 0.91 & 5.3 \\ 0.03 & 2.20 \end{pmatrix}$
				Guan's	(0.01, 0.01)	(5.15, 8.3)	0.16	$\begin{pmatrix} 0.07 & 4.86 \\ 0.04 & 0.08 \end{pmatrix}$

Table 6.3: Relative mean squared errors of the multivariate LGCP computed from 95% of the best simulations. Using $R = 0.25$.

Conclusion

We began this thesis by building the necessary theory and describing the univariate and multivariate log-Gaussian Cox process. We based this on the definition of the multivariate LGCP as introduced in article Waagepetersen et al. [2016]. For these processes we derived the form of their first- and second-order characteristics.

Our aim was to estimate the parameters in the LGCP models using the composite likelihood method. Thus, we took inspiration from article Waagepetersen [2007] and constructed composite likelihood functions of first and second order as limit of the likelihoods of approximating discrete models. Furthermore, we generalized this for multivariate models. Moreover, thanks to the article Dvořák and Prokešová [2012], we were able to avoid the need for numerically demanding multiple integration when maximizing composite likelihood functions.

Furthermore, we showed how the well-established Guan's CL method could be applied to multivariate log-Gaussian Cox process.

We then compared the two methods in a simulation study. Both of them approximate very well the parameters that give the frequency of points of each species. For the parameters expressing the internal correlation structures and variability, the Guan's CL method then performed well. However, the way in which the variability parameter is calculated here cannot ensure that its estimate is positive, as we would expect. It might be a matter for further research to find out how to resolve this inconvenience, e.g. by modifying the optimization procedure to incorporate the positivity constraint.

This work can certainly serve as inspiration for much further research. It would certainly be interesting to investigate the properties of the composite likelihood method we derived, specifically the consistency and asymptotic normality of its estimates. However, it would be even more attractive to compare the estimates obtained using the composite likelihood method, either ours or Guan's, with those obtained using the minimum contrast method, which has been suggested earlier in the literature for multivariate log-Gaussian Cox process.

Bibliography

A. Baddeley, I. Bárány, and R. Schneider. Spatial point processes and their applications. *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pages 1–75, 2007.

A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015.

A. Choiruddin, F. Cuevas, J.-F. Coeurjolly, and R. Waagepetersen. Regularized estimation for highly multivariate log Gaussian Cox processes. 30:649–662, 05 2019.

N. Cressie. *Statistics for Spatial Data*. Revised Edition. Wiley, Iowa State University, 1993.

E. Crow. *Lognormal Distributions: Theory and Applications*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1987.

A. Darah. Mixed; Lagrange's and Cauchy's remainders form. *International Journal of Science and Research (IJSR)*, Volume 9:1232–1236, 02 2020.

P. J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2013.

J. Dvořák and M. Prokešová. Moment estimation methods for stationary spatial cox processes-a comparison. *Kybernetika*, 48(5):1007–1026, 2012.

Y. Guan. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476):1502–1512, 2006. ISSN 01621459.

K. Hessellund, G. Xu, Y. Guan, and R. Waagepetersen. Second order semiparametric inference for multivariate log Gaussian Cox processes. 71:244–268, 12 2020. ISSN 01621459.

A. Jalilian, A. Safari, and H. Sohrabi. Modeling spatial patterns and species associations in a hyrcanian forest using a multivariate log-Gaussian Cox process. 1:59–76, 03 2020.

V. Jarník. *Diferenciální počet II*. Academia, 1984.

M. Jullum, T. Thorarinsdottir, and F. Bachl. Estimating seal pup production in the greenland sea by using bayesian hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69, 01 2020.

J. Kopecký and T. Mrkvička. On the Bayesian estimation for the stationary Neyman-Scott point processes. *Applications of Mathematics*, 61:503–514, 08 2016.

B Lindsay. Composite likelihood. *Contemporary Mathematics*, 80:221–239, 01 1988.

J. Mateu and A. Jalilian. Spatial point processes and neural networks: A convenient couple. *Spatial Statistics*, 02 2022. ISSN 2211-6753.

J. Møller and R. P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CHAPMAN HALL/CRC, Boca Raton London New York Washington, D.C., 2004.

J. Møller and R.P. Waagepetersen. Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34:643 – 684, 12 2007.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

J. Rataj. *Bodové procesy*. Second Edition. Karolinum, Praha, 2006.

R. Waagepetersen, Y. Guan, A. Jalilian, and J. Mateu. Analysis of multispecies point patterns by using multivariate log-Gussian Cox process. *J. R. Stat. Soc. C*, (65):77–96, 2016.

R. P. Waagepetersen. An estimating function approach to inference for inhomogeneous neyman-scott processes. *Biometrics*, 63(1):252–258, 2007. ISSN 0006341X, 15410420.

List of Tables

A. Model comparison on all simulations

To provide a complete picture of the results of the simulations performed, you can find here tables showing the results from all simulations. That is, without excluding the 5% of outliers.

Table A.1 shows the results for the univariate LGCP model. Table A.2 shows the relative bias for the multivariate LGCP model and in the last Table A.3 reported here we find the relative MSE for the multivariate LGCP model.

The conclusions drawn from them are comparable to those presented in section 6.

			Our CL						Guan's CL					
			relBias			relMSE			relBias			relMSE		
ρ_0	σ_0	δ_0	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\delta}$
150	1	10	0.16	-0.07	1.88	0.03	0.02	128.27	-0.14	0.02	0.27	0.03	0.73	1.81
	0.5	10	0.10	0.92	-1.00	0.01	0.86	1.00	-0.10	-0.06	0.03	0.01	1.00	0.09
250	1	10	0.16	-0.09	5.3	0.03	0.10	558.57	-0.18	-0.04	0.21	0.04	0.71	0.60
	0.5	10	0.09	0.90	-0.44	0.01	0.93	15.71	-0.17	-0.02	0.11	0.03	0.70	0.09

Table A.1: Relative biases and relative mean squared errors of the estimates of the univariate LGCP computed from all simulations. Using $R = 0.2$.

					relBias			
ρ_0	σ_0^2	δ_0	$\mathbb{A}_0$	method	$\widehat{\mu}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\widehat{\mathbb{A}}$
(200, 400)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.06, 0.11)	(-1.54, -2.15)	-0.63	$\begin{pmatrix} 0.34 & 0.50 \\ 0.89 & 0.11 \end{pmatrix}$
				Guan's	(0.07, -0.22)	(-0.11, 1.34)	0.23	$\begin{pmatrix} -0.67 & 0.30 \\ 0.17 & -0.40 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.02, -0.04)	(0.57, -0.72)	-0.97	$\begin{pmatrix} -0.83 & 2.30 \\ 0.19 & -0.63 \end{pmatrix}$
				Guan's	(0.06, 0.05)	(-1.01, -1.27)	0.20	$\begin{pmatrix} 0.10 & 2.21 \\ 0.19 & 0.11 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.07, -0.07)	(2.53, -6.4)	-0.92	$\begin{pmatrix} -0.87 & 0.30 \\ 0.18 & -0.33 \end{pmatrix}$
				Guan's	(0.04, 0.04)	(-1.85, -2.20)	0.36	$\begin{pmatrix} 0.15 & 0.21 \\ 0.20 & 0.17 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.08, -0.05)	(2.65, -4.46)	-0.92	$\begin{pmatrix} -0.89 & 2.30 \\ 0.18 & -0.41 \end{pmatrix}$
				Guan's	(0.03, 0.05)	(-1.66, -1.88)	0.31	$\begin{pmatrix} 0.10 & 2.21 \\ 0.20 & 0.09 \end{pmatrix}$
(300, 300)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.02, -0.01)	(0.69, -1.59)	-0.98	$\begin{pmatrix} -0.87 & 0.30 \\ 0.19 & -0.53 \end{pmatrix}$
				Guan's	(0.05, 0.09)	(-1.03, -1.26)	0.27	$\begin{pmatrix} 0.13 & 0.21 \\ 0.20 & 0.12 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.02, -0.03)	(0.51, -0.98)	-1.00	$\begin{pmatrix} -0.83 & 2.30 \\ 0.19 & -0.61 \end{pmatrix}$
				Guan's	(0.07, 0.06)	(-1.14, -1.31)	0.29	$\begin{pmatrix} 0.11 & 2.21 \\ 0.20 & 0.13 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.06, -0.07)	(2.56, -4.32)	-0.97	$\begin{pmatrix} -0.88 & 0.30 \\ 0.18 & -0.42 \end{pmatrix}$
				Guan's	(0.04, 0.06)	(-1.75, -2.28)	0.32	$\begin{pmatrix} 0.14 & 0.21 \\ 0.20 & 0.15 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(-0.05, -0.08)	(2.65, -4.55)	-1.00	$\begin{pmatrix} -0.90 & 2.30 \\ 0.18 & -0.40 \end{pmatrix}$
				Guan's	(0.03, 0.02)	(-1.43, -1.66)	0.3	$\begin{pmatrix} 0.11 & 2.21 \\ 0.2 & 0.09 \end{pmatrix}$

Table A.2: Relative biases of the multivariate LGCP computed from all simulations. Using $R = 0.25$.

ρ_0	σ_0^2	δ_0	$\mathbb{A}_0$	method	relMSE $\widehat{\mu}$	relMSE $\widehat{\sigma^2}$	relMSE $\widehat{\delta}$	relMSE $\widehat{\mathbb{A}}$
(200, 400)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.03)	(5.7, 5.93)	0.62	$\begin{pmatrix} 0.31 & 0.89 \\ 1.58 & 0.09 \end{pmatrix}$
				Guan's	(0.04, 0.16)	(1.6, 12.3)	0.28	$\begin{pmatrix} 0.56 & 0.09 \\ 0.03 & 1.08 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(3.49, 43.02)	0.97	$\begin{pmatrix} 1.08 & 5.29 \\ 0.04 & 1.88 \end{pmatrix}$
				Guan's	(0.03, 0.02)	(2.44, 2.47)	0.14	$\begin{pmatrix} 0.13 & 4.91 \\ 0.04 & 0.13 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.02, 0.01)	(8.92, 861.37)	0.91	$\begin{pmatrix} 0.91 & 0.09 \\ 0.03 & 4.25 \end{pmatrix}$
				Guan's	(0.01, 0.01)	(6.64, 10.21)	0.27	$\begin{pmatrix} 0.07 & 0.04 \\ 0.04 & 0.11 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.02, 0.01)	(9.57, 530.82)	0.91	$\begin{pmatrix} 0.92 & 5.30 \\ 0.03 & 3.44 \end{pmatrix}$
				Guan's	(0.02, 0.01)	(6.89, 7.19)	0.19	$\begin{pmatrix} 0.13 & 4.89 \\ 0.04 & 0.14 \end{pmatrix}$
(300, 300)	(1, 1)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(2.87, 79.13)	0.98	$\begin{pmatrix} 1.05 & 0.09 \\ 0.04 & 2.54 \end{pmatrix}$
				Guan's	(0.02, 0.03)	(2.51, 2.78)	0.23	$\begin{pmatrix} 0.09 & 0.05 \\ 0.04 & 0.10 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(4.00, 53.50)	1.00	$\begin{pmatrix} 1.11 & 5.29 \\ 0.04 & 2.06 \end{pmatrix}$
				Guan's	(0.02, 0.02)	(2.44, 2.79)	0.24	$\begin{pmatrix} 0.11 & 4.89 \\ 0.04 & 0.14 \end{pmatrix}$
	(0.5, 0.5)	10	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(9.11, 519.37)	0.97	$\begin{pmatrix} 0.92 & 0.09 \\ 0.03 & 3.39 \end{pmatrix}$
				Guan's	(0.02, 0.01)	(7.80, 7.53)	0.24	$\begin{pmatrix} 0.07 & 0.04 \\ 0.04 & 0.08 \end{pmatrix}$
			$\begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}$	Ours	(0.01, 0.01)	(8.09, 503.99)	1.00	$\begin{pmatrix} 0.92 & 5.3 \\ 0.03 & 3.45 \end{pmatrix}$
				Guan's	(0.01, 0.02)	(6.54, 10.31)	0.25	$\begin{pmatrix} 0.11 & 4.89 \\ 0.04 & 0.13 \end{pmatrix}$

Table A.3: Relative mean squared errors of the multivariate LGCP computed from all simulations. Using $R = 0.25$.