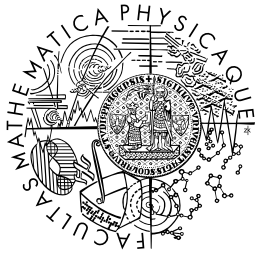


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Filip Kulla

Point processes of objects with random lifetime

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: RNDr. Jiří Dvořák, Ph.D.

Study programme: Probability, Mathematical Statistics and Econometrics

Study branch: MPSP

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

I would like to express my gratitude to RNDr. Jiří Dvořák, Ph.D. for supervising this thesis, for his time, patience and a lot of useful advice. Similarly, I would like to thank my parents and my girlfriend Majka for supporting me throughout my studies.

Title: Point processes of objects with random lifetime

Author: Filip Kulla

Department: Department of Probability and Mathematical Statistics

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

Abstract: The thesis deals with point processes of objects with random lifetime. The form of the likelihood function of an observed spatial-temporal pattern with random lifetimes is derived, where the formula is subsequently generalised to the case of censored lifetimes. Moreover, some simple parametric models are introduced and conditions under which they are non-explosive are derived. In addition, aspects of our implementation of the algorithm which generates a realisation of a given spatial-temporal point process with random lifetimes and of the likelihood-based estimation are discussed. The thesis contains a simulation study in which the use of the (partial) likelihood on simulated data is demonstrated and properties of resulting estimates are discussed. Furthermore, it contains an application of the partial likelihood to the real data, where the question of interest is the spatial dynamics of propagation of an observed population of flowers.

Keywords: point process, space-time process, marked process, lifetime, likelihood inference

Contents

Introduction

Many real world phenomena have temporal or even spatial-temporal character. For instance, we can consider behaviour of a population of flowers in a meadow. In this example, the temporal character (points in time when individual flowers occur) is indeed complemented by the spatial one as well (positions in the meadow in which these flowers occur). It is important to be capable of analysing the data created by these spatial-temporal phenomena and of obtaining insights about the underlying mechanisms. Fundamental tools for dealing with these types of data, whether it is predicting future events or indeed generating insights, are stochastic processes modelling the phenomena of interest. In this thesis we will concentrate mostly on modelling of spatial-temporal phenomena in which individual events/objects have a random lifetime. For instance, in the aforementioned example of flowers in a meadow, every individual flower is born at some time T and position X, it lives and influences its surroundings for some random period M and at time $T + M$ it eventually dies. When analysing real or simulated spatial-temporal point patterns (with random lifetimes), we will use the maximum likelihood estimation (the exact form of the likelihood function will be derived in the thesis).

In the first chapter we introduce some fundamental concepts, which will be used throughout the whole thesis, in context of purely temporal point processes. In the second chapter we generalise concepts from the first chapter to spatial-temporal point processes with random lifetimes. We introduce an appropriate stochastic process to model these processes and discuss generation of its realisations. Furthermore, in this chapter we derive the form of the likelihood function. In the third chapter we introduce some simple parametric models, state conditions under which they are non-explosive and present some examples. We show that the likelihood function can be in special cases maximised analytically. In the fourth chapter we deal with more realistic assumption of censored lifetimes and we derive the form of the likelihood function in this context. A part of the thesis is our own implementation of the algorithm which generates a realisation of a given spatial-temporal point process with random lifetimes in addition to the implementation of the likelihood-based estimation. The relevant source codes are provided in the electronic attachment to the thesis. Details of our implementations are given in the fifth chapter. In the sixth chapter we demonstrate the use of the derived methods on simulated data and discuss properties of the corresponding estimates. Finally, in the seventh chapter we analyse the spatial dynamics in the real data consisting of observations of Succisa pratensis (a flowering plant). The plants were observed once a year, which causes discretisation and presents further challenges to the analysis. We propose two types of models, one with multiplicative and one with additive structure of interactions, and use a Monte Carlo based goodness of fit test to assess how well these models describe the observed data.

1. Temporal point processes

In order to develop concepts which will be useful in the study of spatial-temporal point processes, which are of our primary interest, we will start by considering purely temporal processes first. This chapter is based on Rasmussen [2018]. Presented theory can also be found in Daley and Vere-Jones [2003] in a much more technical setting of random measures.

A temporal point process is, in essence, just a collection of times of events. It can be modelled as a measurable mapping Φ from some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ to the space $\mathcal{N}$ – the space of locally finite counting measures operating on $(\mathbb{R}, \mathcal{B})$, which is endowed with a σ-algebra $\mathfrak{N}$ – the smallest σ-algebra with respect to which all projections $\pi_B : \mathcal{N} \to [0, \infty], B \in \mathcal{B}$, defined as $\pi_B(\mu) = \mu(B)$, are measurable. We assume that we have a simple point process – $\mathsf{P}(\Phi \in \{\nu \in \mathcal{N} : \nu(\{t\}) \leq 1, \forall t \in \mathbb{R}\}) = 1$, i.e. no two events are allowed to coincide.

Alternatively, and because of the existence of the total ordering on $\mathbb{R}$ (the usual "less than or equal to") perhaps more conveniently, it is possible to model a temporal point process as a stochastic process $(T_n)_{n=1}^{\infty}$, where T_n is a non-negative random variable representing the time of the n-th event. Indeed, since we have this order of time it is natural to define a temporal point process by specifying a stochastic model for the time of the $(n+1)$-th event T_{n+1} given that we know times of all the previous events $T_1, \ldots, T_n$. That is, we can specify the conditional probability density function of T_{n+1} given $T_1, \ldots, T_n$, i.e. we can specify $f_{T_{n+1} \,|\, T_1, \ldots, T_n}(t_{n+1} \,|\, t_1, \ldots, t_n)$.

Remark. We will usually prefer abbreviated notation of densities in which we try to suppress names of random variables when they are clear from the context, e.g. instead of $f_{T_{n+1} \,|\, T_1, \ldots, T_n}(t_{n+1} \,|\, t_1, \ldots, t_n)$ we would write just $f(t_{n+1} \,|\, t_1, \ldots, t_n)$. Furthermore, instead of explicitly writing the whole condition $t_1, \ldots, t_n$, we will prefer $\mathcal{H}_{t_n}$, which will stand for this whole history of the process up to (and including) time of the n-th event t_n, e.g. the aforementioned density could be written as $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$.

As a consequence of specifying all conditional densities $f(t_{n+1} \,|\, \mathcal{H}_{t_n}), n \in \mathbb{N}_0$, the system of all finite-dimensional distributions of the stochastic process $(T_n)_{n=1}^{\infty}$ is specified as well, since

$$\prod_{i=1}^{n} f(t_i \,|\, \mathcal{H}_{t_{i-1}}) = f(t_1, \ldots, t_n).$$

Remark. Here, $\mathcal{H}_{t_0}$ stands for no condition at all. By $f(t_1 \,|\, \mathcal{H}_{t_0})$ we mean simply the unconditional density of T_1 at the point t_1, i.e. the density $f(t_1)$.

Nevertheless, in order to define the conditional distribution of T_{n+1} given $\mathcal{H}_{t_n}$, it is sometimes more intuitive to do it by specifying the so-called conditional intensity function $\lambda^*(t)$ instead of the conditional density $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$. This conditional intensity function is defined as

$$\lambda^*(t) = \frac{f(t \,|\, \mathcal{H}_{t_n})}{1 - F(t \,|\, \mathcal{H}_{t_n})}, \; t > t_n \text{ and } F(t \,|\, \mathcal{H}_{t_n}) < 1,$$

where $F(t \,|\, \mathcal{H}_{t_n})$ denotes the cumulative distribution function corresponding to $f(t \,|\, \mathcal{H}_{t_n})$. Technically, it is a system of functions defined for any $n \in \mathbb{N}_0$ and for

any history $\mathcal{H}_{t_n}$ (that is, for any n-tuple $t_1, \ldots, t_n$) as a function from $\{t \in R : t > t_n, F(t \,|\, \mathcal{H}_{t_n}) < 1\}$ to $\mathbb{R}$.

Remark. Again, in the special case of $n = 0$ we understand that $t_0 = 0$ and that $\lambda^*(t)$ is defined through the unconditional density (and distribution function) of T_1 as $\lambda^*(t) = f(t)/(1 - F(t))$, $t > 0$.

The interpretation of the conditional intensity function $\lambda^*(t)$ is rather similar to that of the density $f(t \,|\, \mathcal{H}_{t_n})$. For some $t > t_n$ and some small neighbourhood of t denoted by $\mathrm{d}t$ it holds that

$$f(t \,|\, \mathcal{H}_{t_n})\,|\mathrm{d}t| \approx \int_{\mathrm{d}t} f(s \,|\, \mathcal{H}_{t_n})\,\mathrm{d}s = \mathsf{P}(T_{n+1} \in \mathrm{d}t \,|\, \mathcal{H}_{t_n}).$$

Analogously, for $\lambda^*(t)$, where $t > t_n$, and some small punctured right neighbourhood of t denoted by $\mathrm{d}t$ we have that:

$$\lambda^*(t)\,|\mathrm{d}t| = \frac{f(t \,|\, \mathcal{H}_{t_n})\,|\mathrm{d}t|}{1 - F(t \,|\, \mathcal{H}_{t_n})} \approx \frac{\mathsf{P}(T_{n+1} \in \mathrm{d}t \,|\, \mathcal{H}_{t_n})}{\mathsf{P}(T_{n+1} > t \,|\, \mathcal{H}_{t_n})} = \mathsf{P}(T_{n+1} \in \mathrm{d}t \,|\, \mathcal{H}_{t_n}, T_{n+1} > t).$$

From a given conditional intensity function $\lambda^*(t)$ the cumulative distribution function $F(t \,|\, \mathcal{H}_{t_n})$ or the probability density function $f(t \,|\, \mathcal{H}_{t_n})$ can be obtained.

Theorem 1. *It holds that*

$$F(t \,|\, \mathcal{H}_{t_n}) = 1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right) \tag{1.1}$$

and

$$f(t \,|\, \mathcal{H}_{t_n}) = \lambda^*(t) \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right). \tag{1.2}$$

Proof. From the relation

$$\lambda^*(t) = \frac{f(t \,|\, \mathcal{H}_{t_n})}{1 - F(t \,|\, \mathcal{H}_{t_n})} = -\frac{\mathrm{d}}{\mathrm{d}t} \log(1 - F(t \,|\, \mathcal{H}_{t_n}))$$

and from the fact that $F(t_n \,|\, \mathcal{H}_{t_n}) = 0$ it follows that

$$\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s = -\log(1 - F(t \,|\, \mathcal{H}_{t_n})) + \log(1 - F(t_n \,|\, \mathcal{H}_{t_n})) = -\log(1 - F(t \,|\, \mathcal{H}_{t_n})),$$

which directly implies (1.1). Equation (1.2) then follows from differentiation of (1.1).

□

Theorem 1 is essential in supporting the aforementioned assertion that a conditional intensity function $\lambda^*(t)$ can be used to define the conditional distribution of T_{n+1} given $\mathcal{H}_{t_n}$. Indeed, once we are given $\lambda^*(t)$, thanks to equations (1.1) and (1.2) we know also $F(t_n \,|\, \mathcal{H}_{t_n})$ and $f(t_n \,|\, \mathcal{H}_{t_n})$. We still, however, have to make sure that $1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right)$ (as a function of t) is truly a cumulative distribution function or that $\lambda^*(t) \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right)$ is truly a probability density function.

Theorem 2. *If a system of functions $\lambda^*(t)$ satisfies for any $n \in \mathbb{N}_0$ and for any history $\mathcal{H}_{t_n}$ that*

- $\lambda^*(t)$ *is non-negative,*
- $\forall t > t_n : \int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s < \infty$ *and*
- $\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s \to \infty$ *as* $t \to \infty$*,*

then it uniquely defines a temporal point process $(T_n)_{n=1}^{\infty}$ *and it is its conditional intensity function.*

Proof. Uniqueness follows from Theorem 1. Thus, it is sufficient to realise that $F(t) := 1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right)$ is truly a cumulative distribution function. Continuity of $F(t)$, hence also right-continuity, follows from the fundamental theorem of calculus. Furthermore, when $t < u$, then $\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s \leq \int_{t_n}^{u} \lambda^*(s)\,\mathrm{d}s$, from which it follows that $1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right) \leq 1 - \exp\left(-\int_{t_n}^{u} \lambda^*(s)\,\mathrm{d}s\right)$. Hence, we have that $F(t)$ is non-decreasing. Finally, $F(t)$ has appropriate limits because $F(t_n) = 0$ and $\lim_{t\to\infty} F(t) = 1$.

□

When analysing a given realisation of a temporal point process, it is often a question of interest to find an appropriate model for the observed point pattern. If we have some parametric model (specified, for instance, by the conditional intensity function $\lambda_\theta^*(t)$, which depends on some parameter $\theta \in \Theta$) then in order to estimate the unknown parameter it is a popular choice to use the maximum likelihood inference. Let us define the integrated conditional intensity function as

$$\Lambda^*(t) = \int_0^t \lambda^*(s)\,\mathrm{d}s.$$

Remark. Here we understand that with the arrival of the next point t_{n+1} the definition of $\lambda^*(t)$ is updated – instead of the history $\mathcal{H}_{t_n}$ it now depends on the history $\mathcal{H}_{t_{n+1}}$. More precisely, if we suppose that in an observation interval $[0, T]$ we have observed a point pattern $t_1, \ldots, t_n$, then it is understood that $\lambda^*(t)$ is defined in a piecewise manner, i.e. on the interval $(t_i, t_{i+1}]$ it is defined through the history $\mathcal{H}_{t_i}$ as $\lambda^*(t) = \frac{f(t \mid \mathcal{H}_{t_i})}{1 - F(t \mid \mathcal{H}_{t_i})}$.

Theorem 3. *Suppose that in an observation interval* $[0, T]$ *we have observed a point pattern* $t_1, \ldots, t_n$*. Then the likelihood function is given by*

$$L(\theta) = \left(\prod_{i=1}^{n} \lambda_\theta^*(t_i)\right) \exp\left(-\Lambda_\theta^*(T)\right).$$

Proof. The joint density $f_\theta(t_1, \ldots, t_n)$ does not contain all the information available from the observed point pattern. It does not take into account that the time of the $(n+1)$-th event is greater than T, although this information is contained in the observed realisation as well. Therefore the likelihood is of the form

$$\begin{aligned} L(\theta) &= f_\theta(t_1, \ldots, t_n)(1 - F_\theta(T \mid \mathcal{H}_{t_n})) = \left(\prod_{i=1}^{n} f_\theta(t_i \mid \mathcal{H}_{t_{i-1}})\right)(1 - F_\theta(T \mid \mathcal{H}_{t_n})) \\ &= \left(\prod_{i=1}^{n} \lambda_\theta^*(t_i) \exp\left(-\int_{t_{i-1}}^{t_i} \lambda_\theta^*(s)\,\mathrm{d}s\right)\right) \exp\left(-\int_{t_n}^{T} \lambda_\theta^*(s)\,\mathrm{d}s\right) \\ &= \left(\prod_{i=1}^{n} \lambda_\theta^*(t_i)\right) \exp\left(-\int_0^T \lambda_\theta^*(s)\,\mathrm{d}s\right). \end{aligned}$$

In the third equation we used the relations (1.1) and (1.2) and our agreement that $t_0 = 0$.

□

2. Spatial-temporal point processes

Once we have acquainted ourselves with purely temporal point processes, we can shift our attention to point processes which are spatial-temporal. In this chapter we adapt definitions and theorems regarding marked temporal point processes from Rasmussen [2018] to our specific setting where marks contain both spatial locations as well as lifetimes of objects.

In contrast with a temporal point process, which is just a collection of times of events, a spatial-temporal point process contains not only information about times of events but also about their positions in space. That is, it can be modelled as a stochastic process $(T_n, X_n)_{n=1}^{\infty}$, where T_n represents time of the n-th event while X_n represents its position in space.

We want to study point processes in which every event has an associated random mark, which determines for how long this event influences its surroundings (or "for how long it lives"). It is natural to model a point process of this type by a stochastic process $(T_n, X_n, M_n)_{n=1}^{\infty}$, where T_n represents time, X_n position in space and M_n mark of the n-th event. So far we have been speaking about events and their positions in space, in this new context of point processes with marks (M_n), however, it will be more appropriate to introduce a slightly different terminology. Instead of events we will speak about objects and their times of births T_n, their times of deaths $T_n + M_n$, their lifetimes M_n and their positions in space X_n.

So far we have not specified in what space we observe positions of objects. From now on it will be always assumed that we observe them in some space $W \subset \mathbb{R}^d$ such that W is a bounded Borel set with positive Lebesgue measure, i.e. $\exists r > 0 : W \subset \mathcal{B}(0, r)$, $W \in \mathcal{B}(\mathbb{R}^d)$ and $|W| > 0$.

To summarize, we are going to concentrate on stochastic processes of the form $(T_n, X_n, M_n)_{n=1}^{\infty}$, where we assume that T_n and M_n are non-negative random variables and that X_n is a random variable with values in W. In addition we assume, just as we did in case of temporal point processes, that times of births $(T_n)_{n=1}^{\infty}$ constitute a simple point process, i.e. $T_1 < T_2 < T_3 < \ldots$. Once again, we understand that T_n represents time of birth, M_n lifetime and X_n position in space of the n-th object.

Remark. When integrating over the whole space W, we will usually omit the symbol W from under the integral sign, i.e. instead of $\int_W$ we will write only $\int$.

Remark. To shorten the notation, we will use the concept of history of the stochastic process $(T_n, X_n, M_n)_{n=1}^{\infty}$. By $\mathcal{H}_{t_n}$ we will again understand the whole history of the process up to time of birth of the n-th object (t_n), that is, times of births, positions in space and lifetimes of the first n objects. Once again, $\mathcal{H}_{t_n}$ will stand for $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$. It will be again useful to understand that $t_0 = 0$ and that $\mathcal{H}_{t_0}$ stands for no condition at all.

Analogously as in the case of temporal point processes, we can define stochastic process $(T_n, X_n, M_n)_{n=1}^{\infty}$ by specifying conditional densities

$$f(t_{n+1} \,|\, \mathcal{H}_{t_n}),\ f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1}),\ f(m_{n+1} \,|\, \mathcal{H}_{t_n},\ t_{n+1}, x_{n+1}),\ n \in \mathbb{N}_0, \tag{2.1}$$

because once these densities are specified, the system of all finite-dimensional distributions of the process $(T_n, X_n, M_n)_{n=1}^{\infty}$ is specified as well, since

$$\prod_{i=1}^{n} f(t_i \,|\, \mathcal{H}_{t_{i-1}}) f(x_i \,|\, \mathcal{H}_{t_{i-1}}, t_i) f(m_i \,|\, \mathcal{H}_{t_{i-1}}, t_i, x_i) = f(t_1, x_1, m_1, \ldots, t_n, x_n, m_n).$$

Remark. Similarly as in the previous chapter we prefer abbreviated notation of densities in which we try to suppress names of random variables when they are clear from the arguments, e.g. $f(m_{n+1} \,|\, \mathcal{H}_{t_n},\, t_{n+1}, x_{n+1})$ is a shorter notation of $f_{M_{n+1} \,|\, T_1, X_1, M_1, \ldots, T_{n+1}, X_{n+1}}(m_{n+1} \,|\, \mathcal{H}_{t_n},\, t_{n+1}, x_{n+1})$.

2.1 Conditional intensity function

Again, in order to define the process $(T_n, X_n, M_n)_{n=1}^{\infty}$, instead of prescribing the densities in (2.1) we will rather prescribe a conditional intensity function.

First, we will define the ground intensity as

$$\lambda^*(t) = \frac{f(t \,|\, \mathcal{H}_{t_n})}{1 - F(t \,|\, \mathcal{H}_{t_n})},\ t > t_n \text{ and } F(t \,|\, \mathcal{H}_{t_n}) < 1.$$

The conditional intensity function is then defined as

$$\lambda^*(t, x) = \lambda^*(t) f(x \,|\, \mathcal{H}_{t_n}, t),\ t > t_n, F(t \,|\, \mathcal{H}_{t_n}) < 1 \text{ and } x \in W.$$

Remark. Note that the ground intensity is defined exactly as the conditional intensity function for temporal point processes with the exception that now the history $\mathcal{H}_{t_n}$ contains not only times of births $t_1, \ldots, t_n$ but also positions in space and lifetimes $x_1, m_1, \ldots, x_n, m_n$.

Since the definition of the ground intensity $\lambda^*(t)$ is the same as was the definition of the conditional intensity function for purely temporal point processes, their interpretations are the same as well. For some $t > t_n$ and some small punctured right neighbourhood of t denoted by $\mathrm{d}t$ it holds that

$$\lambda^*(t)\,|\mathrm{d}t| = \frac{f(t \,|\, \mathcal{H}_{t_n})\,|\mathrm{d}t|}{1 - F(t \,|\, \mathcal{H}_{t_n})} \approx \frac{\mathsf{P}(T_{n+1} \in \mathrm{d}t \,|\, \mathcal{H}_{t_n})}{\mathsf{P}(T_{n+1} > t \,|\, \mathcal{H}_{t_n})} = \mathsf{P}(T_{n+1} \in \mathrm{d}t \,|\, \mathcal{H}_{t_n}, T_{n+1} > t).$$

Furthermore, the interpretation of the conditional intensity function $\lambda^*(t, x)$ is given by the following. For some $t > t_n$, some small punctured right neighbourhood of t denoted by $\mathrm{d}t$ and some small neighbourhood of x denoted by $\mathrm{d}x$ it holds that

$$\begin{aligned}
\lambda^*(t, x)\,|\mathrm{d}t|\,|\mathrm{d}x| &= \frac{f(t \,|\, \mathcal{H}_{t_n}) f(x \,|\, \mathcal{H}_{t_n}, t)\,|\mathrm{d}t|\,|\mathrm{d}x|}{1 - F(t \,|\, \mathcal{H}_{t_n})} = \frac{f(t, x \,|\, \mathcal{H}_{t_n})\,|\mathrm{d}t|\,|\mathrm{d}x|}{1 - F(t \,|\, \mathcal{H}_{t_n})} \\
&\approx \frac{\int_{\mathrm{d}t}\int_{\mathrm{d}x} f(s, y \,|\, \mathcal{H}_{t_n})\,\mathrm{d}y\,\mathrm{d}s}{1 - F(t \,|\, \mathcal{H}_{t_n})} = \frac{\mathsf{P}(T_{n+1} \in \mathrm{d}t, X_{n+1} \in \mathrm{d}x \,|\, \mathcal{H}_{t_n})}{\mathsf{P}(T_{n+1} > t \,|\, \mathcal{H}_{t_n})} \\
&= \mathsf{P}(T_{n+1} \in \mathrm{d}t, X_{n+1} \in \mathrm{d}x \,|\, \mathcal{H}_{t_n}, T_{n+1} > t).
\end{aligned}$$

In analogy to Theorem 1, it holds that from a given conditional intensity function $\lambda^*(t, x)$ densities $f(t \,|\, \mathcal{H}_{t_n})$ and $f(x \,|\, \mathcal{H}_{t_n}, t)$ can be obtained.

Theorem 4. *It holds that*

$$\int \lambda^*(t,x)\,\mathrm{d}x = \lambda^*(t),$$

$$F(t \mid \mathcal{H}_{t_n}) = 1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right), \tag{2.2}$$

and

$$f(t \mid \mathcal{H}_{t_n}) = \lambda^*(t) \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right). \tag{2.3}$$

Proof. The first point follows from

$$\int \lambda^*(t,x)\,\mathrm{d}x = \int \lambda^*(t) f(x \mid \mathcal{H}_{t_n}, t)\,\mathrm{d}x = \lambda^*(t).$$

In order to prove relations (2.2) and (2.3) it is sufficient to replicate the proof of Theorem 1 with the exception that now $\mathcal{H}_{t_n}$ stands for times of births, positions in space as well as lifetimes of the first n objects, whereas in Theorem 1 it stood only for times of the first n events.

□

Similarly as it was in case of purely temporal point processes, when we want to define a spatial-temporal point process by specifying the conditional intensity function $\lambda^*(t,x)$ we cannot expect to get a valid model by specifying this function arbitrarily, e.g. it cannot be negative.

Theorem 5. *If a system of functions $\lambda^*(t,x)$ satisfies for any $n \in \mathbb{N}_0$ and for any history $\mathcal{H}_{t_n}$ that*

- *$\lambda^*(t,x)$ is non-negative,*
- *$\forall t > t_n : 0 < \int \lambda^*(t,x)\,\mathrm{d}x < \infty$,*
- *$\forall t > t_n : \int_{t_n}^{t} \int \lambda^*(s,x)\,\mathrm{d}x\,\mathrm{d}s < \infty$ and*
- *$\int_{t_n}^{t} \int \lambda^*(s,x)\,\mathrm{d}x\,\mathrm{d}s \to \infty$ as $t \to \infty$,*

then it, together with a model for lifetimes of objects (i.e. with the densities $f(m_{n+1} \mid \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$), uniquely defines a spatial-temporal point process with random lifetimes of objects $(T_n, X_n, M_n)_{n=1}^{\infty}$ and it is its conditional intensity function.

Proof. Let us put $\lambda^*(t) = \int \lambda^*(t,x)\,\mathrm{d}x$. Then, since $0 < \lambda^*(t) < \infty$ we can put $f(x \mid \mathcal{H}_{t_n}, t) = \lambda^*(t,x)/\lambda^*(t)$. Function $f(x \mid \mathcal{H}_{t_n}, t)$ is a probability density function because it is non-negative and it integrates to 1. It remains to show that $F(t \mid \mathcal{H}_{t_n})$ obtained from $\lambda^*(t)$ by relation (2.2) is truly a cumulative distribution function. Continuity of $F(t \mid \mathcal{H}_{t_n})$, hence also right-continuity, follows from the fundamental theorem of calculus. Furthermore, when $t < u$, then $\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s \leq \int_{t_n}^{u} \lambda^*(s)\,\mathrm{d}s$, from which it follows that $1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\,\mathrm{d}s\right) \leq 1 - \exp\left(-\int_{t_n}^{u} \lambda^*(s)\,\mathrm{d}s\right)$. Hence, we have that $F(t \mid \mathcal{H}_{t_n})$ is non-decreasing. Finally, $F(t \mid \mathcal{H}_{t_n})$ has appropriate limits because $F(t_n \mid \mathcal{H}_{t_n}) = 0$ and $\lim_{t\to\infty} F(t \mid \mathcal{H}_{t_n}) =$

1. Therefore, we have all the densities from (2.1), which, together, specify all finite-dimensional distributions of the process $(T_n, X_n, M_n)_{n=1}^{\infty}$.

For the uniqueness it suffices to realise that all the densities from (2.1) are determined uniquely (almost everywhere). We assume that the conditional densities of lifetimes $f(m_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$ are given, thus it suffices to show the uniqueness of $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$ and $f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1})$. Suppose that the conditional intensity function $\lambda^*(t,x)$ factorises into $\lambda^*(t,x) = \lambda_1^*(t) f_1(x \,|\, \mathcal{H}_{t_n}, t) = \lambda_2^*(t) f_2(x \,|\, \mathcal{H}_{t_n}, t)$. Then, by integrating $\lambda^*(t,x)$ with respect to x we get that $\lambda_1^*(t) = \lambda_2^*(t)$, which also implies that $f_1(x \,|\, \mathcal{H}_{t_n}, t) = f_2(x \,|\, \mathcal{H}_{t_n}, t)$. Hence, we have shown the uniqueness of the densities $f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1})$. The uniqueness of the densities $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$ then follows from the relation (2.3) because $\lambda_1^*(t) = \lambda_2^*(t)$.

□

2.2 Maximum likelihood estimation

Suppose that based on an observed realisation $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$ observed in some time window $[0, T]$ we need to choose the most appropriate model for a spatial-temporal point process $(T_n, X_n, M_n)_{n=1}^{\infty}$ from a given parametric family of models (specified, for instance, by the conditional intensity function $\lambda_\theta^*(t,x)$ and by the conditional densities of lifetimes $f_\gamma(m_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$, which depend on some parameters $\theta \in \Theta$ and $\gamma \in \Gamma$). In the rest of the thesis we will indeed assume that the conditional densities of lifetimes and the conditional intensity function are parametrised by different vectors of model parameters $\gamma \in \Gamma$ and $\theta \in \Theta$, i.e. no component of θ influences the distribution of lifetimes and no component of γ influences the conditional intensity function. Similarly as in the case of temporal point processes, the maximum likelihood inference can be used to estimate the unknown parameters.

Remark. The assumption that we observed exactly $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$ is a little unrealistic because the chances are that at time T some objects are still alive. A more realistic assumption is that instead of $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$ we actually observed $t_1, x_1, \min\{m_1, T - t_1\}, \ldots, t_n, x_n, \min\{m_n, T - t_n\}$ together with indicators $\mathit{1}[m_i \leq T - t_i]$. This assumption of censored lifetimes will be discussed later.

Let us define the integrated ground intensity as

$$\Lambda^*(t) = \int_0^t \lambda^*(s)\,\mathrm{d}s.$$

Remark. Here we understand that with the arrival of the next ($(n+1)$-th) object the definition of $\lambda^*(t,x)$ is updated – instead of the history $\mathcal{H}_{t_n}$ it now depends on the history $\mathcal{H}_{t_{n+1}}$.

Theorem 6. *Suppose that in a time window $[0,T]$ we have observed a point pattern $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$. Then the likelihood function is given by*

$$L(\theta,\gamma) = \left(\prod_{i=1}^{n} f_\gamma(m_i \,|\, \mathcal{H}_{t_{i-1}}, t_i, x_i)\right)\left(\prod_{i=1}^{n} \lambda_\theta^*(t_i, x_i)\right)\exp\left(-\Lambda_\theta^*(T)\right).$$

Proof.

$$
\begin{aligned}
L(\theta,\gamma) &= f_{\theta,\gamma}(t_1,x_1,m_1,\dots,t_n,x_n,m_n)(1-F_\theta(T\,|\,\mathcal{H}_{t_n})) \\
&= \left(\prod_{i=1}^{n} f_{\theta,\gamma}(t_i,x_i,m_i\,|\,\mathcal{H}_{t_{i-1}})\right)(1-F_\theta(T\,|\,\mathcal{H}_{t_n})) \\
&= \left(\prod_{i=1}^{n} f_\theta(t_i\,|\,\mathcal{H}_{t_{i-1}}) f_\theta(x_i\,|\,\mathcal{H}_{t_{i-1}},t_i) f_\gamma(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i)\right)(1-F_\theta(T\,|\,\mathcal{H}_{t_n})) \\
&= \left(\prod_{i=1}^{n} \lambda^*_\theta(t_i)\exp\left(-\int_{t_{i-1}}^{t_i}\lambda^*_\theta(s)\,\mathrm{d}s\right) f_\theta(x_i\,|\,\mathcal{H}_{t_{i-1}},t_i)\right) \\
&\quad \left(\prod_{i=1}^{n} f_\gamma(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i)\right)\exp\left(-\int_{t_n}^{T}\lambda^*_\theta(s)\,\mathrm{d}s\right) \\
&= \left(\prod_{i=1}^{n} f_\gamma(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i)\right)\left(\prod_{i=1}^{n}\lambda^*_\theta(t_i,x_i)\right)\exp\left(-\int_0^T\lambda^*_\theta(s)\,\mathrm{d}s\right).
\end{aligned}
\tag{2.4}
$$

In the fourth equation we used the relations (2.2) and (2.3) and our agreement that $t_0 = 0$.

□

Note that the likelihood function L factorises into two terms, out of which one depends on γ and the other depends on θ. Therefore, when maximising the likelihood it is sufficient to maximise these two terms separately. Indeed, when we have two non-negative functions $f(x)$ and $g(y)$ attaining their maximums in $\hat{x}$ and $\hat{y}$ respectively, then $f(x)g(y) \le f(\hat{x})g(y) \le f(\hat{x})g(\hat{y})$.

If we assume i.i.d. lifetimes $(M_n)_{n=1}^\infty$ independent of both times of births as well as positions in space or lifetimes given by geostatistical marking using some random field $Z(t,x)$, then the term $\prod_{i=1}^n f_\gamma(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i)$ further simplifies. First, suppose that the lifetimes are i.i.d. and independent of $(T_n,X_n)_{n=1}^\infty$ with distribution given by a density $f_\gamma(m)$. Then $\prod_{i=1}^n f_\gamma(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i) = \prod_{i=1}^n f_\gamma(m_i)$. In case of geostatistical marking, that is when we assume that $M_n = Z(T_n,X_n)$ where $Z(t,x)$ is some random field independent of $(T_n,X_n)_{n=1}^\infty$, it holds that

$$
\begin{aligned}
&f_{M_i\,|\,T_1,X_1,M_1,\dots,T_{i-1},X_{i-1},M_{i-1},T_i,X_i}(m_i\,|\,t_1,x_1,m_1,\dots,t_{i-1},x_{i-1},m_{i-1},t_i,x_i) \\
&\quad = f_{Z(T_i,X_i)\,|\,T_1,X_1,Z(T_1,X_1),\dots,T_{i-1},X_{i-1},Z(T_{i-1},X_{i-1}),T_i,X_i}(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i) \\
&\quad = \frac{f_{T_1,X_1,Z(t_1,x_1),\dots,T_i,X_i,Z(t_i,x_i)}(t_1,x_1,m_1,\dots,t_i,x_i,m_i)}{f_{T_1,X_1,Z(t_1,x_1),\dots,T_i,X_i}(t_1,x_1,m_1,\dots,t_i,x_i)} \\
&\quad = \frac{f_{Z(t_1,x_1),\dots,Z(t_i,x_i)}(m_1,\dots,m_i) f_{T_1,X_1,\dots,T_i,X_i}(t_1,x_1,\dots,t_i,x_i)}{f_{Z(t_1,x_1),\dots,Z(t_{i-1},x_{i-1})}(m_1,\dots,m_{i-1}) f_{T_1,X_1,\dots,T_i,X_i}(t_1,x_1,\dots,t_i,x_i)} \\
&\quad = f_{Z(t_i,x_i)\,|\,Z(t_1,x_1),\dots,Z(t_{i-1},x_{i-1})}(m_i\,|\,m_1,\dots,m_{i-1}).
\end{aligned}
$$

Therefore, the term $\prod_{i=1}^n f_\gamma(m_i\,|\,\mathcal{H}_{t_{i-1}},t_i,x_i)$ simplifies to the joint density $f_{Z(t_1,x_1),\dots,Z(t_n,x_n)}(m_1,\dots,m_n)$, which can again depend on some unknown parameter $\gamma \in \Gamma$.

Term $\exp\left(-\int_0^T \lambda^*_\theta(s)\,\mathrm{d}s\right)$ from the likelihood is a little problematic since $\int_0^T \lambda^*_\theta(s)\,\mathrm{d}s = \int_0^T \int \lambda^*_\theta(s,x)\,\mathrm{d}x\,\mathrm{d}s$ requires integration over space as well as over

time, which is computationally demanding. Since this term arose from the densities $f_\theta(t_i \mid \mathcal{H}_{t_{i-1}})$ present in (2.4), it is a promising idea to just omit them from the likelihood function. This leads us to a definition of the partial likelihood [Cox, 1975, Diggle et al., 2010]

$$L_p(\theta) = \prod_{i=1}^{n} f_\theta(x_i \mid \mathcal{H}_{t_{i-1}}, t_i). \tag{2.5}$$

By maximising the partial likelihood we, roughly speaking, maximise probability that conditionally on histories and times of births their positions would be exactly where we observed them. Since $\int \lambda^*_\theta(t_i) f_\theta(x \mid \mathcal{H}_{t_{i-1}}, t_i)\,\mathrm{d}x = \lambda^*_\theta(t_i)$, it holds that

$$f_\theta(x_i \mid \mathcal{H}_{t_{i-1}}, t_i) = \frac{\lambda^*_\theta(t_i) f_\theta(x_i \mid \mathcal{H}_{t_{i-1}}, t_i)}{\int \lambda^*_\theta(t_i) f_\theta(x \mid \mathcal{H}_{t_{i-1}}, t_i)\,\mathrm{d}x} = \frac{\lambda^*_\theta(t_i, x_i)}{\int \lambda^*_\theta(t_i, x)\,\mathrm{d}x}.$$

Hence, the partial likelihood can be expressed as

$$L_p(\theta) = \prod_{i=1}^{n} \frac{\lambda^*_\theta(t_i, x_i)}{\int \lambda^*_\theta(t_i, x)\,\mathrm{d}x}. \tag{2.6}$$

We see that the partial likelihood simplifies the integration over space-time to n integrations over space only. An unpleasant disadvantage of the partial likelihood is that it may not identify some parameters. This problem will be exemplified later.

2.3 Generation of realisations

When analysing a spatial-temporal point process it is often useful to be capable of simulating realisations from a particular model. Indeed, being able to simulate from a given model is helpful for model verification, prediction, Monte Carlo tests, estimation of distribution of summary statistics, etc.

Suppose that we are given a conditional intensity function $\lambda^*(t, x)$ together with the model for generating lifetimes, i.e. densities $f(m_{n+1} \mid \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$. Then a realisation of the point process determined by this system can be generated (in a time interval $[0, T]$) by the following algorithm [Diggle, 2013].

1. Put $t_0 = 0$, $n = 0$.
2. While (TRUE):
   (a) Generate $S \sim \mathrm{Exp}(1)$.
   (b) Find t such that $\int_{t_n}^{t} \lambda^*(u)\,\mathrm{d}u = S$.
   (c) If $(t > T)$: break.
   (d) Put $n = n + 1$.
   (e) Put $t_n = t$.
   (f) Generate x_n from the density proportional to $\lambda^*(t_n, \cdot)$.
   (g) Generate m_n from the density $f(m \mid \mathcal{H}_{t_{n-1}}, t_n, x_n)$.
3. Output $(t_1, x_1, m_1, \ldots, t_n, x_n, m_n)$.

Theorem 7. *Suppose that we are given a conditional intensity function $\lambda^*(t, x)$ together with a model for generating lifetimes, i.e. together with a collection of densities $f(m_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$. Then the aforementioned algorithm generates a realisation of the spatial-temporal point process with random lifetimes of objects that is uniquely determined by this system.*

Proof. The following proof is a little generalisation of the proof for purely temporal point processes, which is given in Rasmussen [2018]. The given conditional intensity function $\lambda^*(t, x)$ uniquely determines the ground intensity $\lambda^*(t)$ and thus also densities $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$ and $f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1})$. This uniqueness was discussed in Theorem 5. We need to show that conditional distributions of times of births, positions in space and lifetimes in the process generated by the algorithm coincide with the uniquely determined densities $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$, $f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1})$ and $f(m_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$. Let us start with the conditional distributions of times of births in the generated process. Let $F^*_{T_{n+1}}(t \,|\, \mathcal{H}_{t_n})$ denote the conditional cumulative distribution function of the time of birth of the $(n+1)$-th object (T_{n+1}) in the generated process. Then it holds for $t > t_n$ that

$$\begin{aligned}
F^*_{T_{n+1}}(t \,|\, \mathcal{H}_{t_n}) &= \mathsf{P}(T_{n+1} \le t \,|\, \mathcal{H}_{t_n}) = \mathsf{P}(\Lambda^*(T_{n+1}) \le \Lambda^*(t) \,|\, \mathcal{H}_{t_n}) \\
&= \mathsf{P}(\Lambda^*(T_n) + S \le \Lambda^*(t) \,|\, \mathcal{H}_{t_n}) = \mathsf{P}(S \le \Lambda^*(t) - \Lambda^*(t_n) \,|\, \mathcal{H}_{t_n}) \\
&= 1 - \exp\left(-(\Lambda^*(t) - \Lambda^*(t_n))\right) = 1 - \exp\left(-\int_{t_n}^{t} \lambda^*(u)\,\mathrm{d}u\right) \\
&= F(t \,|\, \mathcal{H}_{t_n}),
\end{aligned}$$

where $F(t \,|\, \mathcal{H}_{t_n})$ denotes the cumulative distribution function associated with the uniquely determined density $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$. In the last equality we used the relation (2.2). Hence, we showed that the conditional distributions of times of births in the process generated by the algorithm coincide with the densities $f(t_{n+1} \,|\, \mathcal{H}_{t_n})$ uniquely determined by the given conditional intensity function $\lambda^*(t, x)$. Since the conditional distributions of lifetimes in the generated process also coincide with the given densities $f(m_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1}, x_{n+1})$ (that is how we chose them in the algorithm), it remains to verify that the conditional distributions of positions of objects coincide with the densities $f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1})$. In the algorithm, position of the $(n+1)$-th object X_{n+1} was generated from the density proportional to $\lambda^*(t_{n+1}, \cdot)$, which is exactly $f(x_{n+1} \,|\, \mathcal{H}_{t_n}, t_{n+1})$.

□

3. Parametric models

3.1 Existence and examples

We will deal with models defined by the conditional intensity functions of the following form. Suppose that the number of objects which were born before time t is n_t. That means that $t_1 < \ldots < t_{n_t} < t$. Then, given the whole history of the process before time t, that is, given times of births, positions in space and lifetimes of the first n_t objects $t_1, x_1, m_1, \ldots, t_{n_t}, x_{n_t}, m_{n_t}$, the conditional intensity function will be specified as

$$\lambda^*(t,x) = \prod_{i=1}^{n_t} \Big(h(\|x - x_i\|) g(|t - t_i|) \Big)^{1[|t-t_i| \le m_i]}, \tag{3.1}$$

where $t > t_{n_t}$ and $x \in W$. Here, h and g are some interaction functions from $[0,\infty)$ to $[0,\infty)$.

This choice of the conditional intensity function has a relatively clear interpretation. At time t and position in space x it is affected by only those objects which were born before and are still alive at this time t. Furthermore, we suppose that their effects act multiplicatively and have a specific form. Namely, the effect of the i-th object on $\lambda^*(t,x)$ depends only on the distance from its position in space x_i to x through some function h and on the distance from its time of birth t_i to t through some function g.

In (3.1) we have introduced a model for the conditional intensity function $\lambda^*(t,x)$. So far we have not, however, addressed the question of validity of the resulting model. When speaking about point processes, it is required that their supports are random locally finite sets, i.e. their supports intersect an arbitrary compact set in just finitely many points. Similarly, we would like the stochastic process defined by the conditional intensity function from (3.1) to be non-explosive in the sense that almost surely it holds that

(A) positions in space of objects which are alive at the same time intersect an arbitrary compact subset of W in finitely many points,

(B) times of births intersect an arbitrary compact subset of $\mathbb{R}$ in finitely many points.

Theorem 8. *Consider a stochastic process with the conditional intensity function from (3.1). Suppose that the interaction functions h and g are bounded and that $\exists \delta_0 > 0$, $\forall x \in [0,\delta_0]$, $h(x) = 0$. Then this stochastic process almost surely satisfies conditions of non-explosiveness* (A) *and* (B).

Proof. Condition (A) follows directly from the hard-core property of the interaction function h that $\exists \delta_0 > 0$, $\forall x \in [0,\delta_0]$, $h(x) = 0$. Indeed, the conditional density of the position of the next ($(n+1)$-th) object (given the history $\mathcal{H}_{t_n}$ and the time of birth of this $(n+1)$-th object t_{n+1}) $f(x \,|\, \mathcal{H}_{t_n}, t_{n+1})$ is proportional to the conditional intensity function at time t_{n+1}, i.e. it is proportional to $\lambda^*(t_{n+1}, \cdot)$. Hence, it is equal to zero on δ_0-neighbourhoods of positions of those objects that

are alive at time t_{n+1}. Therefore, the distance between positions of arbitrary two objects that are alive at the same time is greater than δ_0.

It remains to show condition (B). First, we can realise that the conditional intensity function $\lambda^*(t,x)$ is under our assumptions bounded by some constant C. Indeed, thanks to boundedness of W and the hard-core property of h there exists an upper bound on the number of objects that can be alive at the same time , i.e. $\exists N_{Max} \in \mathbb{N}$ such that at an arbitrary time t the number of objects that are at this time alive is less than or equal to N_{Max}. Therefore, the number of factors in the product from (3.1) is less than or equal to N_{Max}, which together with boundedness of h and g implies also boundedness of $\lambda^*(t,x)$. Consequently, the ground intensity $\lambda^*(t)$ is bounded by $K := C\,|W|$.

Let us now concentrate on the period between two consecutive times of births, e.g. $T_n - T_{n-1}$, $n \in \mathbb{N}$ (we will understand that $T_0 = 0$). From the representation of the studied stochastic process in the algorithm above Theorem 7 it holds that $T_n - T_{n-1} \geq \frac{1}{K} Z_n$, where Z_n is a random variable with Exp(1) distribution. Hence, the time of explosion

$$\xi = \sum_{n=1}^{\infty} T_n - T_{n-1} \geq \frac{1}{K} \sum_{n=1}^{\infty} Z_n = \infty \;\; a.s.,$$

where $\{Z_n\}_{n=1}^{\infty}$ are i.i.d. random variables with Exp(1) distribution. The last equality follows from the strong law of large numbers:

$$\lim_{N\to\infty} \sum_{n=1}^{N} Z_n = \Big(\lim_{N\to\infty} N \Big) \Big(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} Z_n \Big) = \infty \;\; a.s.$$

□

Remark. The validity of Theorem 8 is not restricted only to models whose conditional intensity function is given by (3.1). From the proof of Theorem 8 it is obvious that for the almost sure validity of conditions (A) and (B) it is sufficient to assume that $\exists \delta_0 > 0$, $\forall t > 0$, $\lambda^*(t,\cdot)$ is zero on δ_0-neighbourhoods of positions of those objects which are alive at time t and that $\lambda^*(t,x)$ is bounded by some constant C.

Let us present some examples of models whose conditional intensity functions are given by (3.1).

Example. Let us define the interaction functions g and h as $g(t) \equiv 1.15$ and

$$h(x) = \begin{cases} 0, & \text{if} \quad x \leq 0.05, \\ 1 - 0.6 \exp\left(-\frac{(x-0.05)^2}{0.03}\right), & \text{if} \quad x > 0.05. \end{cases} \tag{3.2}$$

Plot of the interaction function h defined in this way is in Figure 3.1. Furthermore, we will put $W = [0,1]^2$ and we will assume that the lifetimes of objects $(M_n)_{n=1}^{\infty}$ are i.i.d. random variables independent of both times of births as well as positions in space, i.e. independent of $(T_n, X_n)_{n=1}^{\infty}$, with gamma distribution with the shape parameter equal to 10 and the scale parameter equal to 0.5.

Stochastic process $(T_n, X_n, M_n)_{n=1}^{\infty}$ defined in this way models a spatially repulsive behaviour of objects with random lifetimes in a unit square. Effects of

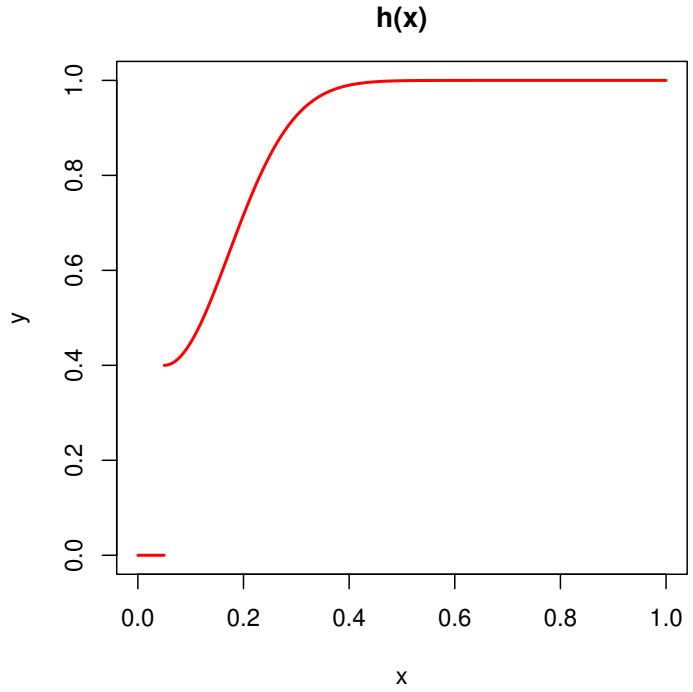


Figure 3.1: Plot of h function from (3.2).

the function h on the spatial interactions of positions of objects are rather clear. Consider an object O_0 whose position in W is x_0. Then the function h causes that when a new object is born while this object O_0 is still alive, the position at which it is born is with probability 1 further than 0.05 from x_0, i.e. positions in W whose distance from x_0 is less than or equal to 0.05 are forbidden. Furthermore, positions in space whose distance from x_0 is greater than 0.05 but still relatively small are quite strongly disfavoured. This penalisation decreases as the distance from x_0 increases.

In a simulated realisation of this model one object was born at time $t = 70.22$. The position of this object was generated from the density proportional to $\lambda^*(t, \cdot)$ (Figure 3.2). Plot of $\lambda^*(t, \cdot)$ exemplifies the spatially repulsive behaviour of the model: $\lambda^*(t, \cdot)$, and hence also the proportional density, attains largest values in regions which are unoccupied by already existing objects (namely, in the central part of W, in the upper right corner and in the lower right corner). Configurations of positions of objects which were alive at some selected times in this simulated realisation can be seen in Figure 3.3.

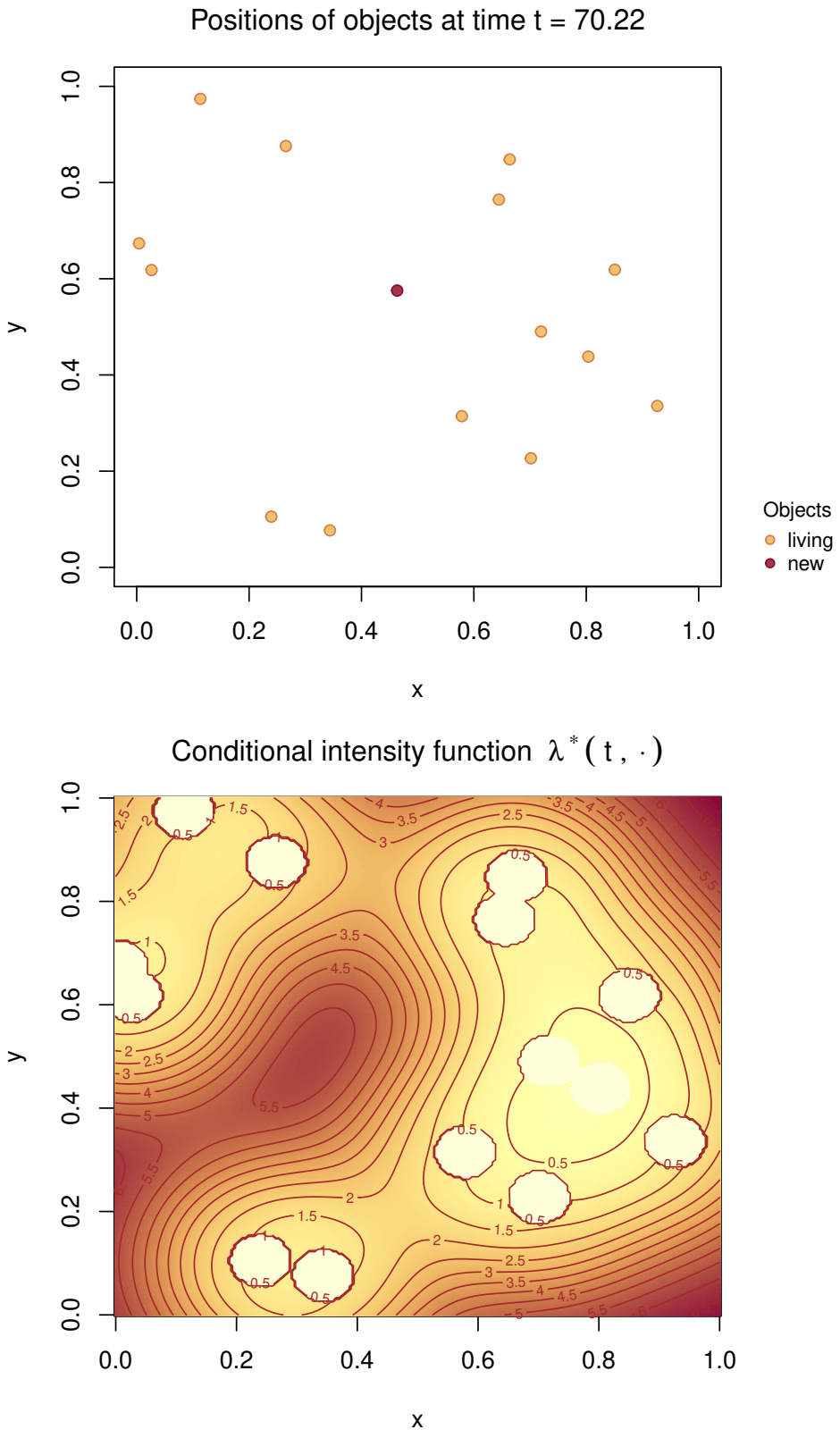


Figure 3.2: Configuration of positions of objects in a simulated realisation of a repulsive model at time $t = 70.22$ and the plot of $\lambda^*(t, \cdot)$.

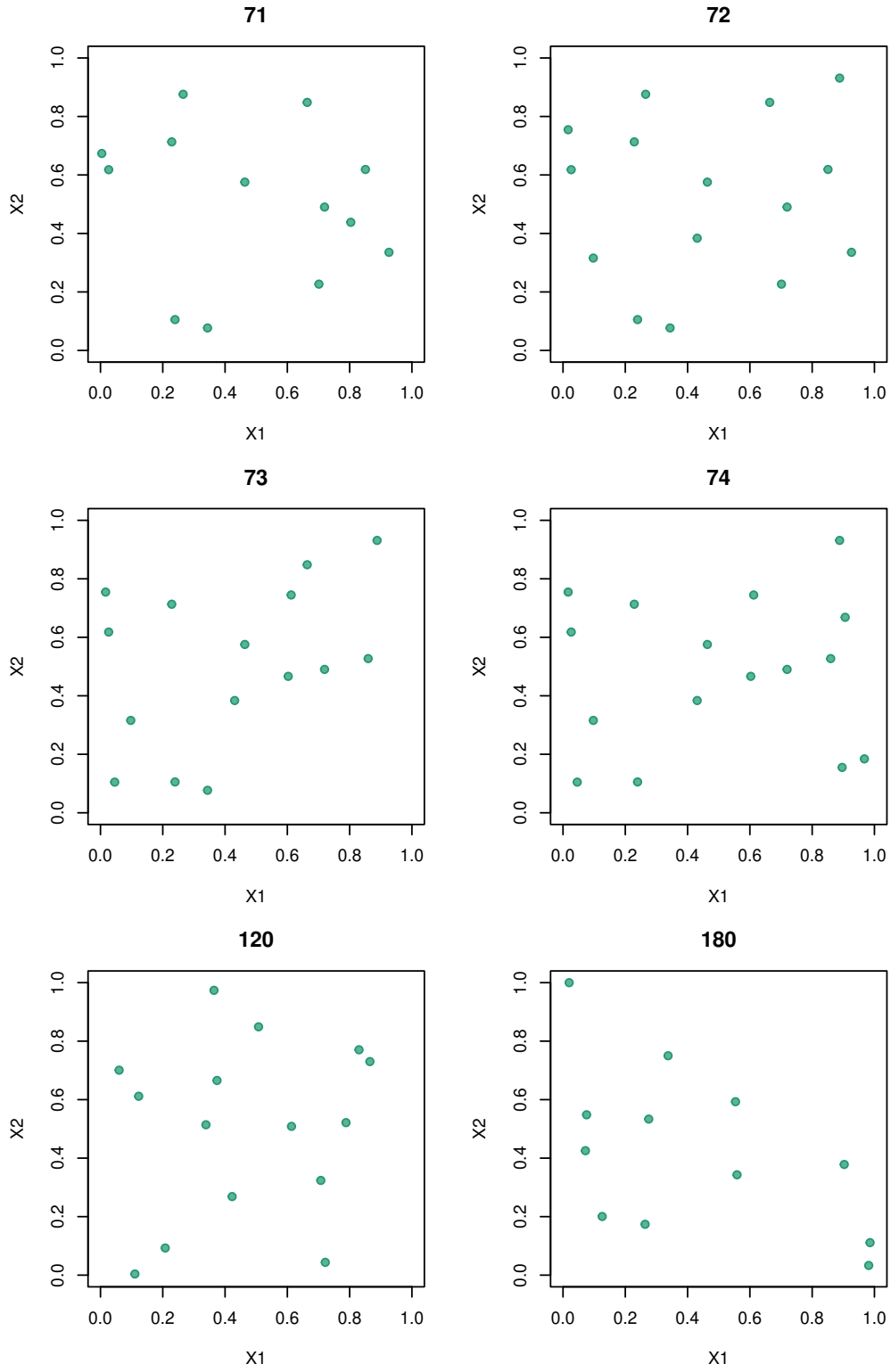


Figure 3.3: Configurations of positions of objects in a simulated realisation of a repulsive model at times 71, 72, 73, 74, 120 and 180.

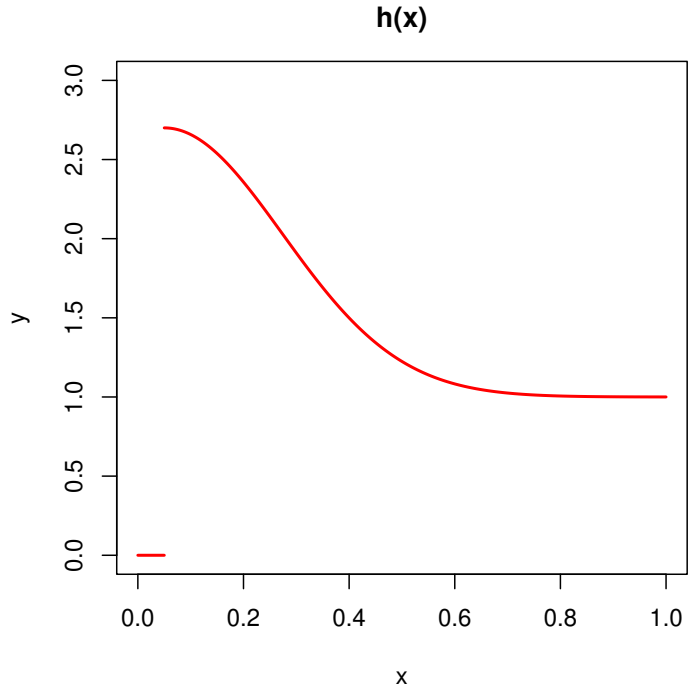


Figure 3.4: Plot of h function from (3.3).

Example. Let us define the interaction functions g and h as $g(t) \equiv 0.55$ and

$$h(x) = \begin{cases} 0, & \text{if} \quad x \leq 0.05, \\ 1 + 1.7 \exp\left(-\frac{(x-0.05)^2}{0.1}\right), & \text{if} \quad x > 0.05. \end{cases} \tag{3.3}$$

Plot of the interaction function h defined in this way is in Figure 3.4. Again, we will put $W = [0,1]^2$ and we will assume that the lifetimes of objects $(M_n)_{n=1}^{\infty}$ are i.i.d. random variables independent of both times of births as well as positions in space, i.e. independent of $(T_n, X_n)_{n=1}^{\infty}$, with uniform distribution on the interval $[3.5, 6.5]$.

Stochastic process $(T_n, X_n, M_n)_{n=1}^{\infty}$ defined in this way models a spatially attractive behaviour of objects with random lifetimes in a unit square. Effects of the function h on the spatial interactions of positions of objects are again rather clear. Consider an object O_0 whose position in W is x_0. Then the function h causes that when a new object is born while this object O_0 is still alive, the position at which it is born is with probability 1 further than 0.05 from x_0, i.e. positions in W whose distance from x_0 is less than or equal to 0.05 are forbidden. On the contrary, positions in space whose distance from x_0 is greater than 0.05 but still relatively small are quite strongly favoured. This attractive effect gradually decreases as the distance from x_0 increases.

In a simulated realisation of this model one object was born at time $t = 68.81$. The position of this object was generated from the density proportional to $\lambda^*(t, \cdot)$ (Figure 3.5). Plot of $\lambda^*(t, \cdot)$ exemplifies the spatially attractive behaviour of the model: $\lambda^*(t, \cdot)$, and hence also the proportional density, attains largest values in the vicinity of already existing objects (namely in the upper part of W). Configurations of positions of objects which were alive at some selected times in this simulated realisation can be seen in Figure 3.6.

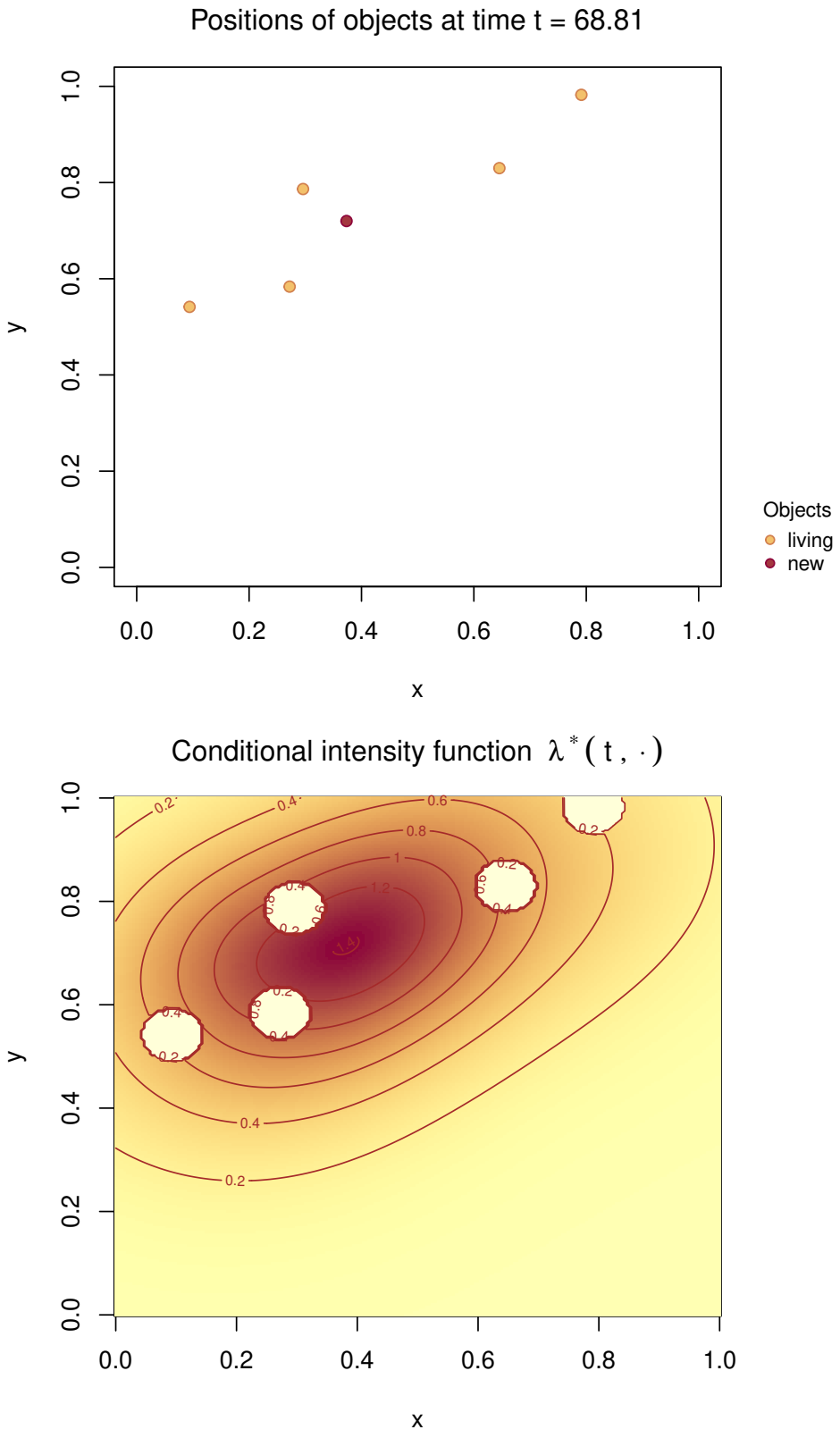


Figure 3.5: Configuration of positions of objects in a simulated realisation of an attractive model at time $t = 68.81$ and the plot of $\lambda^*(t, \cdot)$.

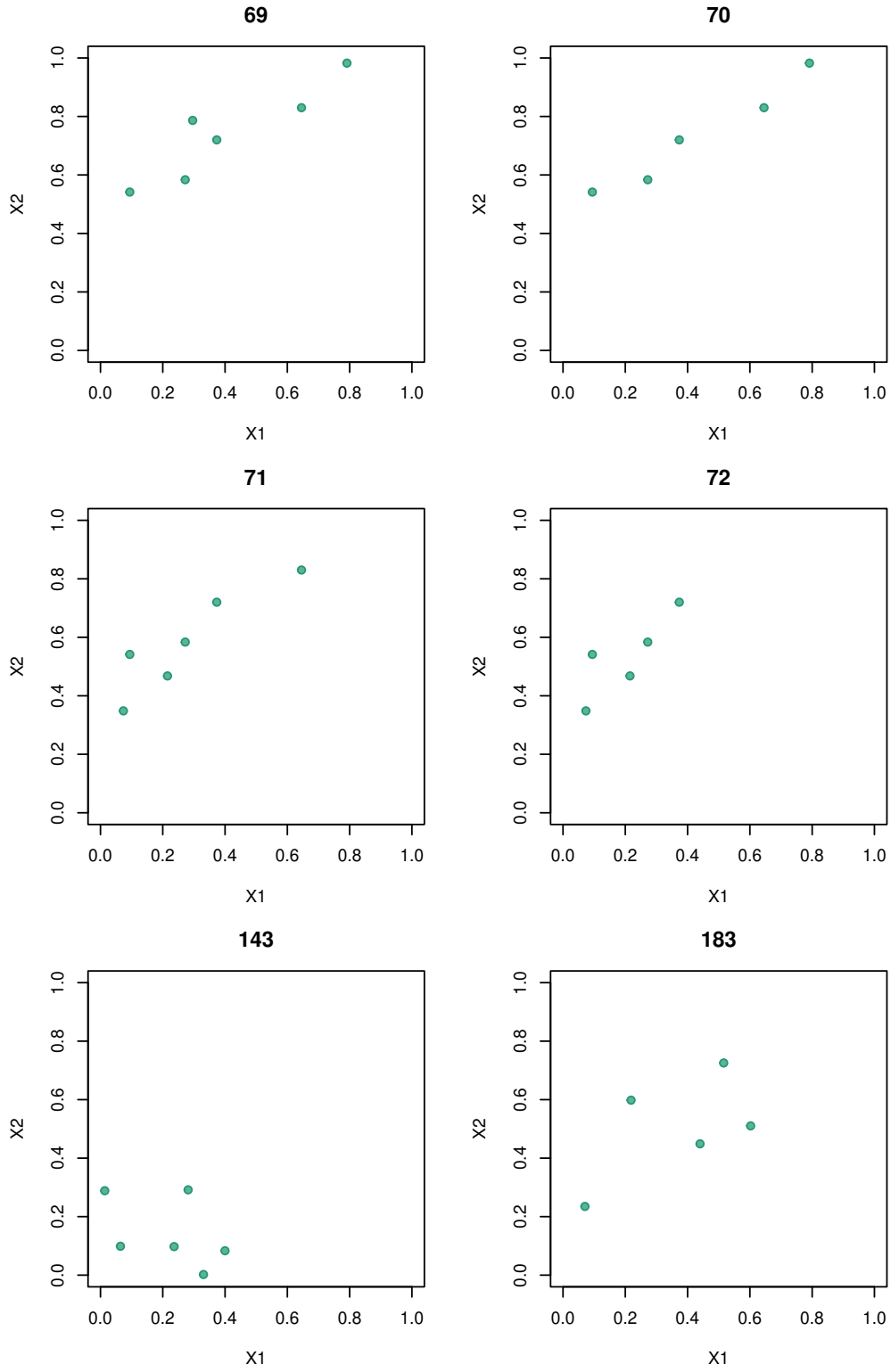


Figure 3.6: Configurations of positions of objects in a simulated realisation of an attractive model at times 69, 70, 71, 72, 143 and 183.

3.2 Maximum likelihood estimation

Thanks to Theorem 6 we know that the likelihood function of a given realisation $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$ observed in a time window $[0, T]$ is of the form

$$L(\theta, \gamma) = \left(\prod_{i=1}^{n} f_\gamma(m_i \,|\, \mathcal{H}_{t_{i-1}}, t_i, x_i) \right) \left(\prod_{i=1}^{n} \lambda^*_\theta(t_i, x_i) \right) \exp\left(- \int_0^T \int \lambda^*_\theta(t, x) \, \mathrm{d}x \, \mathrm{d}t \right).$$

We have explained that since the densities $f_\gamma(m_i \,|\, \mathcal{H}_{t_{i-1}}, t_i, x_i)$ depend on a different parameter than the conditional intensity function $\lambda^*_\theta(t, x)$, in order to maximise $L(\theta, \gamma)$ it is sufficient to maximise the terms

$$L_1(\gamma) = \left(\prod_{i=1}^{n} f_\gamma(m_i \,|\, \mathcal{H}_{t_{i-1}}, t_i, x_i) \right)$$

and

$$L_2(\theta) = \left(\prod_{i=1}^{n} \lambda^*_\theta(t_i, x_i) \right) \exp\left(- \int_0^T \int \lambda^*_\theta(t, x) \, \mathrm{d}x \, \mathrm{d}t \right)$$

separately.

For the conditional intensity function given by (3.1), assuming that the interaction functions h and g depend on some unknown parameter $\theta \in \Theta$, is the logarithm of $L_2(\theta)$ of the form

$$\begin{aligned} \log\left(L_2(\theta)\right) = & \sum_{i=1}^{n} \log \left(\prod_{j=1}^{n_{t_i}} \Big(h_\theta(\|x_i - x_j\|) g_\theta(|t_i - t_j|) \Big)^{1[|t_i - t_j| \le m_j]} \right) \\ & - \int_0^T \int \prod_{i=1}^{n_t} \Big(h_\theta(\|x - x_i\|) g_\theta(|t - t_i|) \Big)^{1[|t - t_i| \le m_i]} \mathrm{d}x \, \mathrm{d}t. \end{aligned}$$

Similarly, from (2.6) we know that the partial likelihood is of the form

$$L_p(\theta) = \prod_{i=1}^{n} \frac{\lambda^*_\theta(t_i, x_i)}{\int \lambda^*_\theta(t_i, x) \, \mathrm{d}x},$$

which means that for the conditional intensity function given by (3.1) it holds that

$$\begin{aligned} \log\left(L_p(\theta)\right) = & \sum_{i=1}^{n} \log \left(\prod_{j=1}^{n_{t_i}} \Big(h_\theta(\|x_i - x_j\|) g_\theta(|t_i - t_j|) \Big)^{1[|t_i - t_j| \le m_j]} \right) \\ & - \sum_{i=1}^{n} \log \left(\int \prod_{j=1}^{n_{t_i}} \Big(h_\theta(\|x - x_j\|) g_\theta(|t_i - t_j|) \Big)^{1[|t_i - t_j| \le m_j]} \mathrm{d}x \right). \end{aligned}$$

Example. Perhaps the simplest example arises by choosing $g(t) \equiv 1$ and $h(x) = 1[x \ge \delta]$, where $\delta > 0$ is an unknown parameter, which needs to be estimated. This choice of interaction functions g and h leads to a hard-core model, in which objects influence the conditional intensity function only through spatial interactions given by h. The main effect of this function is that any two objects which live at the same time have to satisfy that the distance between their positions in space is not smaller than δ. Beyond this hard-core property, no further spatial interactions, neither attractive nor repulsive, are present.

First, let us concentrate on maximising $\log(L_2(\delta))$. For the current choice of g and h it holds that

$$\log(L_2(\delta)) = \sum_{i=1}^{n} \log\left(\prod_{j=1}^{n_{t_i}} \mathit{1}\left[\|x_i - x_j\| \geq \delta\right]^{\mathit{1}\left[|t_i - t_j| \leq m_j\right]}\right) - \int_0^T \int \prod_{i=1}^{n_t} \mathit{1}\left[\|x - x_i\| \geq \delta\right]^{\mathit{1}\left[|t - t_i| \leq m_i\right]} \mathrm{d}x \, \mathrm{d}t.$$

Once we denote

$$\delta_0 = \min\left\{\|x_i - x_j\| : i, j \in \mathbb{N};\, i \leq n;\, j < i;\, |t_i - t_j| \leq m_j\right\}$$

it holds that if $\delta \leq \delta_0$, then

$$\sum_{i=1}^{n} \log\left(\prod_{j=1}^{n_{t_i}} \mathit{1}\left[\|x_i - x_j\| \geq \delta\right]^{\mathit{1}\left[|t_i - t_j| \leq m_j\right]}\right) = 0,$$

whereas if $\delta > \delta_0$, then this term is equal to $-\infty$ (with the convention that $\log(0) = -\infty$). Furthermore, the term

$$\int_0^T \int \prod_{i=1}^{n_t} \mathit{1}\left[\|x - x_i\| \geq \delta\right]^{\mathit{1}\left[|t - t_i| \leq m_i\right]} \mathrm{d}x \, \mathrm{d}t = \int_0^T |W| - \Big| \bigcup_{\substack{t_i < t \\ |t - t_i| \leq m_i}} \mathcal{B}(x_i, \delta) \cap W \Big| \, \mathrm{d}t$$

is monotonic in δ. Indeed, when $\delta_1 < \delta_2$, then

$$\Big| \bigcup_{\substack{t_i < t \\ |t - t_i| \leq m_i}} \mathcal{B}(x_i, \delta_1) \cap W \Big| \leq \Big| \bigcup_{\substack{t_i < t \\ |t - t_i| \leq m_i}} \mathcal{B}(x_i, \delta_2) \cap W \Big|$$

and thus

$$\int_0^T |W| - \Big| \bigcup_{\substack{t_i < t \\ |t - t_i| \leq m_i}} \mathcal{B}(x_i, \delta_1) \cap W \Big| \, \mathrm{d}t \geq \int_0^T |W| - \Big| \bigcup_{\substack{t_i < t \\ |t - t_i| \leq m_i}} \mathcal{B}(x_i, \delta_2) \cap W \Big| \, \mathrm{d}t.$$

Therefore, $\log(L_2(\delta))$ attains its maximum in δ_0.

Analogously, we can try to maximise

$$\log(L_p(\delta)) = \sum_{i=1}^{n} \log\left(\prod_{j=1}^{n_{t_i}} \mathit{1}\left[\|x_i - x_j\| \geq \delta\right]^{\mathit{1}\left[|t_i - t_j| \leq m_j\right]}\right) - \sum_{i=1}^{n} \log\left(\int \prod_{j=1}^{n_{t_i}} \mathit{1}\left[\|x - x_j\| \geq \delta\right]^{\mathit{1}\left[|t_i - t_j| \leq m_j\right]} \mathrm{d}x\right).$$

Behaviour of the first term was already discussed – if $\delta \leq \delta_0$, then it is equal to 0 and if $\delta > \delta_0$, then it is equal to $-\infty$. The second term is again monotonic in δ. Hence, $\log(L_p(\delta))$ attains its maximum in δ_0 as well.

In this example we have been able to maximise both the full likelihood as well as the partial likelihood analytically. In more complicated models numerical optimisation will be necessary.

4. Censoring of lifetimes

As it was already stated (before Theorem 6), the assumption of observing exactly $t_1, x_1, m_1, \ldots, t_n, x_n, m_n$ in a given time window $[0, T]$ is not entirely realistic because at time T some objects might be still alive. If we suppose that the time of death of the i-th object is greater than T, i.e. $t_i + m_i > T$, then instead of observing the exact value of its lifetime m_i we only have the information that it is greater than $T - t_i$. Basic introduction to the analysis of censored data can be found in Kalbfleisch and Prentice [2002].

Let us introduce a stochastic process $(T_n, X_n, C_n, \Delta_n)_{n=1}^{\infty}$, where T_n and X_n still represent time of birth and position in space of the n-th object, $C_n = \min\{M_n, T - T_n\}$ stands for the censored lifetime and $\Delta_n = \mathit{1}[M_n \leq T - T_n]$ indicates whether the n-th lifetime M_n is actually censored ($\Delta_n = 0$) or not ($\Delta_n = 1$).

In contrast with random variable M_n, random variables C_n and Δ_n are actually observable (provided that $T_n \leq T$). Let us suppose that $T_n \leq T$. Then there are two options for the time of death $T_n + M_n$. It can happen that $T_n + M_n \leq T$ or that $T_n + M_n > T$. In the first case we observe even the exact lifetime M_n and therefore also C_n and Δ_n. It holds that $C_n = M_n$ and $\Delta_n = 1$. In the second case we do not observe the exact lifetime M_n, however, we do observe that the n-th object is at time T still alive. Thus, we do have the information that $T_n + M_n > T$, which is sufficient to determine the values of C_n and Δ_n. It holds that $C_n = T - T_n$ and $\Delta_n = 0$.

Remark. We need to be able to distinguish that for some objects we do not know the exact values of their lifetimes, but only know that their times of deaths are greater than T. Therefore, it will be useful to denote $t_1, x_1, c_1, \delta_1, \ldots, t_n, x_n, c_n, \delta_n$ by $\mathcal{H}'_{t_n}$. That is, while $\mathcal{H}_{t_n}$ represents times of births, positions in space and exact lifetimes of the first n objects, $\mathcal{H}'_{t_n}$ tolerates that for some objects information about their exact times of deaths is not available and it is only known that they are greater than T.

Let us now introduce a condition for the conditional densities of times of births and the conditional densities of positions in space, which will be referred to in the following theorem.

Condition 9. *In the considered model it is satisfied that* $\forall i \in \mathbb{N} \setminus \{1\}$, $\forall \mathcal{H}'_{\tilde{t}_{i-1}} = (\tilde{t}_1, \tilde{x}_1, \tilde{c}_1, \tilde{\delta}_1, \ldots, \tilde{t}_{i-1}, \tilde{x}_{i-1}, \tilde{c}_{i-1}, \tilde{\delta}_{i-1})$, $\forall \tilde{t}_i \in (\tilde{t}_{i-1}, T]$ *and* $\forall \tilde{x}_i \in W$*:*

$$f_\theta(\tilde{t}_i \,|\, \mathcal{H}'_{\tilde{t}_{i-1}}) = f_{T_i \,|\, T_1, X_1, M_1, \ldots, T_{i-1}, X_{i-1}, M_{i-1}; \theta}(\tilde{t}_i \,|\, \tilde{t}_1, \tilde{x}_1, \tilde{c}_1, \ldots, \tilde{t}_{i-1}, \tilde{x}_{i-1}, \tilde{c}_{i-1})$$

and

$$f_\theta(\tilde{x}_i \,|\, \mathcal{H}'_{\tilde{t}_{i-1}}, \tilde{t}_i) = f_{X_i \,|\, T_1, X_1, M_1, \ldots, T_{i-1}, X_{i-1}, M_{i-1}, T_i; \theta}(\tilde{x}_i \,|\, \tilde{t}_1, \tilde{x}_1, \tilde{c}_1, \ldots, \tilde{t}_{i-1}, \tilde{x}_{i-1}, \tilde{c}_{i-1}, \tilde{t}_i).$$

Condition 9, which we just introduced, deserves a short comment. Vaguely speaking, this condition states that the probability that the next object is born at some particular time $t \leq T$ is not altered by changing the condition in a way that instead of assuming that some times of deaths are greater than T we just assume that they are equal to T. Similarly for the position in space of the next

object. Roughly speaking, the condition states that the probability that the next object occurs at some particular position $x \in W$ is not altered by changing the condition in a way that instead of assuming that some times of deaths are greater than T we just assume that they are equal to T.

Theorem 10. *Let us suppose that in a given time window $[0,T]$ we observed a realisation $t_1, x_1, c_1, \delta_1, \ldots, t_n, x_n, c_n, \delta_n$. In addition, suppose that Condition 9 holds. Then, the likelihood function is of the form*

$$L(\theta,\gamma) = \left(\prod_{i=1}^{n} \left(f_{M_i;\gamma}(c_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)^{\delta_i} \left(1 - F_{M_i;\gamma}(T - t_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)^{1-\delta_i}\right) \left(\prod_{i=1}^{n} \lambda^*_\theta(t_i, x_i)\right) \exp\left(-\Lambda^*_\theta(T)\right),$$

where in the terms from the first product subscript M_i was added in order to emphasize that we mean conditional density or distribution function of M_i.

Proof. Similarly as in the proof of Theorem 6 the likelihood function can be written as

$$\begin{aligned} L(\theta,\gamma) &= f_{\theta,\gamma}(t_1, x_1, c_1, \delta_1, \ldots, t_n, x_n, c_n, \delta_n)(1 - F_\theta(T \,|\, \mathcal{H}'_{t_n})) \\ &= \left(\prod_{i=1}^{n} f_\theta(t_i \,|\, \mathcal{H}'_{t_{i-1}}) f_\theta(x_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i) f_\gamma(c_i, \delta_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)(1 - F_\theta(T \,|\, \mathcal{H}'_{t_n})). \end{aligned}$$

Let us examine the term $f_\gamma(c_i, \delta_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)$. If $\delta_i = 0$, then

$$f_\gamma(c_i, \delta_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i) = 1 - F_{M_i;\gamma}(T - t_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i),$$

whereas if $\delta_i = 1$, then

$$f_\gamma(c_i, \delta_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i) = f_{M_i;\gamma}(c_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i).$$

Indeed, the first relation follows from (4.1), which is implied by the following sequence of implications

$$\begin{aligned} &[\Delta_i = 0] = [M_i > T - T_i] \subset [C_i = T - T_i] \Rightarrow \\ &[\Delta_i = 0, C_i = T - T_i] = [\Delta_i = 0] = [M_i > T - T_i] \Rightarrow \\ &[\Delta_i = 0, C_i = T - T_i, T_i = t_i] = [M_i > T - T_i, T_i = t_i] \Rightarrow \\ &[\Delta_i = 0, C_i = T - t_i, T_i = t_i] = [M_i > T - t_i, T_i = t_i]. \end{aligned} \tag{4.1}$$

Similarly, the second relation follows from (4.2):

$$\begin{aligned} &[\Delta_i = 1] = [M_i = C_i] \Rightarrow \\ &[\Delta_i = 1, C_i = c_i, T_i = t_i] = [M_i = c_i, C_i = c_i, T_i = t_i] = [M_i = c_i, T_i = t_i], \end{aligned} \tag{4.2}$$

where the last equality is implied by the inclusion

$$[M_i = c_i, T_i = t_i] \subset [C_i = c_i],$$

which holds as long as $c_i \leq T - t_i$. Hence,

$$f_\gamma(c_i, \delta_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i) = \left(f_{M_i;\gamma}(c_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)^{\delta_i} \left(1 - F_{M_i;\gamma}(T - t_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)^{1-\delta_i}.$$

The remaining terms, thanks to the assumptions imposed on the considered model, can be treated in the same way as in Theorem 6. Again, they transform into

$$\left(\prod_{i=1}^{n} \lambda^*_\theta(t_i, x_i)\right) \exp\left(-\int_0^T \lambda^*_\theta(s)\,\mathrm{d}s\right).$$

□

Similarly as in the discussion after Theorem 6, if we assume i.i.d. lifetimes $(M_n)_{n=1}^\infty$ independent of both times of births as well as positions in space, i.e. independent of $(T_n, X_n)_{n=1}^\infty$, with distribution given by a density $f_\gamma(m)$ (whose associated cumulative distribution function is $F_\gamma(m)$) then the term

$$\prod_{i=1}^{n} \left(f_{M_i;\gamma}(c_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)^{\delta_i} \left(1 - F_{M_i;\gamma}(T - t_i \,|\, \mathcal{H}'_{t_{i-1}}, t_i, x_i)\right)^{1-\delta_i}$$

simplifies to

$$\prod_{i=1}^{n} \left(f_\gamma(c_i)\right)^{\delta_i} \left(1 - F_\gamma(T - t_i)\right)^{1-\delta_i}.$$

Furthermore, let us comment on the feasibility of Condition 9, which we assumed in Theorem 10. Models whose conditional intensity function has a causal character in a sense that it does not depend on the future (e.g. the one given by (3.1)) satisfy it. Indeed, let us consider two histories:

$$\begin{aligned}
\mathcal{H}^1_{t_n} &= t_1, x_1, m_1, \ldots, t_n, x_n, m_n, \\
\mathcal{H}^2_{t_n} &= t_1, x_1, \min\{m_1, T - t_1\}, \ldots, t_n, x_n, \min\{m_n, T - t_n\},
\end{aligned}$$

where $t_n < T$. Then, in models with causal conditional intensity functions, these two histories induce conditional intensity functions, which coincide in the time window $(t_n, T]$. That, however, implies that $\forall t_{n+1} \in (t_n, T]$, $\forall x_{n+1} \in W$:

$$\begin{aligned}
f(t_{n+1} \,|\, \mathcal{H}^1_{t_n}) &= f(t_{n+1} \,|\, \mathcal{H}^2_{t_n}), \\
f(x_{n+1} \,|\, \mathcal{H}^1_{t_n}, t_{n+1}) &= f(x_{n+1} \,|\, \mathcal{H}^2_{t_n}, t_{n+1}).
\end{aligned}$$

5. Implementation details

In order to be able to simulate a realisation of a spatial-temporal point process by the algorithm described above Theorem 7 or in order to calculate the (partial) likelihood function we need to be capable of integrating the conditional intensity function $\lambda^*(t,x)$.

5.1 Integration over space

First, let us discuss the numerical integration of the conditional intensity function $\lambda^*(t,x)$ for t fixed as a function of x, i.e. the numerical integration of $\int \lambda^*(t,x)\,\mathrm{d}x$. We use Monte Carlo integration [Caflisch, 1998, Kroese et al., 2011]. Consider a sequence of i.i.d. random variables $\{X_n\}_{n=1}^{\infty}$ with uniform distribution on our space W. From the following relation, where P_{X_1} denotes the distribution of X_1, it follows that $\int \lambda^*(t,x)\,\mathrm{d}x$ can be approximated by $|W|\,\frac{1}{N}\sum_{i=1}^{N}\lambda^*(t,X_i)$. Indeed,

$$\begin{aligned}\int \lambda^*(t,x)\,\mathrm{d}x &= \int \lambda^*(t,x)\frac{|W|}{|W|}\,\mathrm{d}x = |W|\int \lambda^*(t,x)\,\mathrm{d}\,\mathsf{P}_{X_1}(x)\\ &= |W|\,\mathsf{E}\,\lambda^*(t,X_1) \approx |W|\,\frac{1}{N}\sum_{i=1}^{N}\lambda^*(t,X_i),\end{aligned}$$

where the last step is justified by the strong law of large numbers (which holds if $\mathsf{E}\,\lambda^*(t,X_1) = \frac{1}{|W|}\int \lambda^*(t,x)\,\mathrm{d}x < \infty$). The question of practical significance is what value of N to choose. We choose this value with the aid of the central limit theorem and Slutsky's theorem. If we in addition have that $0 < \operatorname{var}\lambda^*(t,X_1) < \infty$, then it holds that

$$\frac{\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\lambda^*(t,X_i) - \frac{1}{|W|}\int \lambda^*(t,x)\,\mathrm{d}x\right)}{\sqrt{S_N^2}} \xrightarrow[N\to\infty]{D} N(0,1),$$

where S_N^2 denotes the sample variance of $\{\lambda^*(t,X_i)\}_{i=1}^{N}$. This convergence implies that

$$\mathsf{P}\left(\left|\,|W|\,\frac{1}{N}\sum_{i=1}^{N}\lambda^*(t,X_i) - \int \lambda^*(t,x)\,\mathrm{d}x\,\right| < \frac{\sqrt{S_N^2}}{\sqrt{N}}\,|W|\,u_{1-\frac{\alpha}{2}}\right) \xrightarrow[N\to\infty]{} 1-\alpha,$$

where $u_{1-\frac{\alpha}{2}}$ denotes the $(1-\frac{\alpha}{2})$-quantile of $N(0,1)$ distribution. This means that for sufficiently large N, probability that the error of our numerical integration is less than $\frac{\sqrt{S_N^2}}{\sqrt{N}}\,|W|\,u_{1-\frac{\alpha}{2}}$ is approximately $1-\alpha$. Hence, we would like to choose the value of N so that $\frac{\sqrt{S_N^2}}{\sqrt{N}}\,|W|\,u_{1-\frac{\alpha}{2}} \approx 0.01\int \lambda^*(t,x)\,\mathrm{d}x$ because then with probability of approximately $1-\alpha$ the relative error of the numerical integration is less than 1%. Unfortunately, we do not know the value of $\int \lambda^*(t,x)\,\mathrm{d}x$, hence we have to approximate it by $|W|\,\frac{1}{N}\sum_{i=1}^{N}\lambda^*(t,X_i)$. That means that during the numerical integration we will keep increasing the value of N until $\frac{\sqrt{S_N^2}}{\sqrt{N}}\,|W|\,u_{1-\frac{\alpha}{2}} \le 0.01\,|W|\,\frac{1}{N}\sum_{i=1}^{N}\lambda^*(t,X_i)$ or until some maximal admissible value for N is reached. In our implementation we choose $\alpha = 0.05$ and the maximal admissible value for N as 50 000.

5.2 Integration over space-time

As it was already mentioned, when generating a realisation of a spatial-temporal point process by the previously described algorithm or when calculating the likelihood function, we need to be able to integrate the conditional intensity function $\lambda^*(t,x)$ not only over space but also over space-time, i.e. we need to be able to calculate integrals of the form $\int_{T_1}^{T_2}\int \lambda^*(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_{T_1}^{T_2} \lambda^*(t)\,\mathrm{d}t$. In order to approximate such an integral we can take a partition $\{t_i\}_{i=0}^n$ of the interval $[T_1,T_2]$ such that $T_1 = t_0 < t_1 < \ldots < t_{n-1} < t_n = T_2$ and use the trapezoidal rule:

$$\int_{T_1}^{T_2}\int \lambda^*(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_{T_1}^{T_2} \lambda^*(t)\,\mathrm{d}t \approx \sum_{i=1}^{n} \frac{\hat{\lambda}^*(t_{i-1}) + \hat{\lambda}^*(t_i)}{2}(t_i - t_{i-1}),$$

where $\hat{\lambda}^*(t_i)$ denotes the numerical approximation of $\lambda^*(t_i) = \int \lambda^*(t_i,x)\,\mathrm{d}x$ as described above. A little inconveniently, the ground intensity $\lambda^*(t)$ might have discontinuity points, e.g. when an object is born or when it dies. To avoid making unnecessary integration errors which might arise by not including these discontinuity points in the partition $\{t_i\}_{i=0}^n$, we choose this partition not only as a regular grid of points in $[T_1,T_2]$ but as a union $D \cup G$, where D denotes the set of discontinuity points of the ground intensity $\lambda^*(t)$ which lie in the interval $[T_1,T_2]$ and G denotes a regular dense grid in $[T_1,T_2]$. Furthermore, we modify our numerical approximation by using one-sided limits:

$$\int_{T_1}^{T_2}\int \lambda^*(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_{T_1}^{T_2} \lambda^*(t)\,\mathrm{d}t \approx \sum_{i=1}^{n} \frac{\hat{\lambda}^*(t_{i-1}+) + \hat{\lambda}^*(t_i-)}{2}(t_i - t_{i-1}),$$

where $\hat{\lambda}^*(t_i-)$ denotes the approximation of $\lim_{t\to t_i-} \lambda^*(t)$ (similarly for the $+$ sign).

Let us remark that in our implementation we choose the regular grid of points G so that the distance between two consecutive points is 0.01 when generating realisations of a spatial-temporal point process and (due to computational reasons) 0.1 when calculating the likelihood function.

5.3 Optimisation

Let us now discuss problems associated with maximisation of the (partial) likelihood function.

Remark. In the following we will speak about the partial likelihood L_p, however, virtually everything applies also to the maximisation of the likelihood function L (or more precisely to its factor L_2).

The partial likelihood L_p is generally too complicated to be maximised analytically, hence it is necessary to use the numerical optimisation. We use the Nelder–Mead method.

The Nelder-Mead method is a direct search optimisation method based on function comparisons. The main idea of the algorithm is to construct a simplex (a convex hull of $d+1$ affinely independent vertices in $\mathbb{R}^d$) and in each iteration of the algorithm try to improve the worst vertex (in case of minimisation, the one

with the largest value of the objective function). More precisely, if we denote the vertices $x_1, \ldots, x_{d+1}$ so that $f(x_1) \leq \ldots \leq f(x_{d+1})$, then in each iteration of the algorithm a new point x_r is proposed (a reflection of x_{n+1} around the centroid of the remaining vertices) and if $f(x_r) < f(x_n)$ then a new simplex without x_{n+1} is constructed. If, however, $f(x_r) \geq f(x_n)$, then to construct the new simplex some kind of contraction or shrink around the best vertex is used. More extensive description of the algorithm can be found in Nelder and Mead [1965] or Lagarias et al. [1998]. Let us remark that we use this algorithm as implemented in function `optim` in R [R Core Team, 2022].

A little inconveniently, for a given parameter θ we cannot calculate the exact theoretical value of $L_p(\theta)$ (or of $\log(L_p(\theta))$, which we actually try to maximise) but rather some approximation of this true value affected by the random integration error. From the relation (2.6) we have that

$$\log(L_p(\theta)) = \sum_{i=1}^{n} \log\Big(\lambda^*_\theta(t_i, x_i)\Big) - \sum_{i=1}^{n} \log\Big(\int \lambda^*_\theta(t_i, x)\,\mathrm{d}x\Big).$$

Moreover, from section 5.1 we know how we approximate $\int \lambda^*_\theta(t_i, x)\,\mathrm{d}x$. Therefore, we approximate $\log(L_p(\theta))$ by

$$\hat{l}_p(\theta) = \sum_{i=1}^{n} \log\Big(\lambda^*_\theta(t_i, x_i)\Big) - \sum_{i=1}^{n} \log\Big(|W|\,\frac{1}{N_i}\sum_{j=1}^{N_i} \lambda^*_\theta(t_i, X^i_j)\Big),$$

where $\{X^i_j : i, j \in \mathbb{N}, j \leq N_i\}$ are i.i.d. random variables with uniform distribution on the space W.

What typically happened when we ran the Nelder-Mead algorithm with the default (strict) stopping criterion is that in relatively few function evaluations the algorithm converged to the vicinity of the point which was eventually declared as the achieved maximum, however, it took many more function evaluations for the algorithm to actually stop. This phenomenon is caused by the fact that we are unable to compute the exact theoretical value of $\log(L_p(\theta))$ but only this value plus some stochastic noise, which means that when the Nelder-Mead simplex converges to the vicinity of the optimal solution, it is difficult for the calculated functional values in the vertices of the simplex to be sufficiently close for the algorithm to terminate. A sample run of the algorithm demonstrating the problem can be seen in the appendix A.1. Therefore, in order to reduce time consumption of a single optimisation, we would like to set the stopping criterion less strictly. We would like to estimate with what variance we calculate the objective function on the neighbourhood of the optimal solution and use this information to set the stopping criterion. This variance could be estimated by calculating many values of the objective function (in the true value of the parameter or in its neighbourhood) and then calculating their sample variance. This approach, however, would be too time consuming. Fortunately, we can estimate this variance of interest in only one function evaluation.

Indeed,

$$\begin{aligned}
\operatorname{var} \hat{l}_p(\theta) &= \operatorname{var}\left[\sum_{i=1}^{n} \log\left(|W|\,\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i)\right)\right] \\
&= \sum_{i=1}^{n} \operatorname{var}\left[\log\left(|W|\,\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i)\right)\right] \\
&\approx \sum_{i=1}^{n} \operatorname{var}\Bigg[\log\Big(\int \lambda_\theta^*(t_i, x)\,\mathrm{d}x\Big) \\
&\qquad + \log'\Big(\int \lambda_\theta^*(t_i, x)\,\mathrm{d}x\Big)\Big(|W|\,\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i) - \int \lambda_\theta^*(t_i, x)\,\mathrm{d}x\Big)\Bigg] \\
&= \sum_{i=1}^{n} \operatorname{var}\left[\log'\Big(\int \lambda_\theta^*(t_i, x)\,\mathrm{d}x\Big)|W|\,\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i)\right] \\
&= \sum_{i=1}^{n} \left(\frac{1}{\int \lambda_\theta^*(t_i, x)\,\mathrm{d}x}\right)^2 \frac{|W|^2}{N_i} \operatorname{var}\lambda_\theta^*(t_i, X_1^i) \\
&\approx \sum_{i=1}^{n} \left(\frac{1}{|W|\,\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i)}\right)^2 \frac{|W|^2}{N_i} \operatorname{var}\lambda_\theta^*(t_i, X_1^i) \\
&\approx \sum_{i=1}^{n} \frac{1}{\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i)\right)^2} \frac{1}{N_i} S_{N_i,i}^2,
\end{aligned}$$

where $S_{N_i,i}^2$ denotes the sample variance of $\{\lambda_\theta^*(t_i, X_j^i)\}_{j=1}^{N_i}$. In the first approximation we approximated by the first-order Taylor polynomial of logarithm expanded about point $\int \lambda_\theta^*(t_i, x)\,\mathrm{d}x$. Furthermore, all equalities follow from the formula $\operatorname{var}(aX + b) = a^2 \operatorname{var}(X)$ and from the fact that $\{X_j^i\}$ are i.i.d.

Hence, we can set the stopping criterion so that the algorithm terminates if the calculated functional values in the vertices of the simplex satisfy that the difference of the largest and the smallest value is comparable to the estimate of the standard deviation with which we calculate the objective function on the neighbourhood of the optimal solution, i.e. when it is comparable to

$$\sqrt{\sum_{i=1}^{n} \frac{1}{\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\lambda_\theta^*(t_i, X_j^i)\right)^2} \frac{1}{N_i} S_{N_i,i}^2}. \tag{5.1}$$

Let us remark that in our implementation we set the parameter `reltol` in the function `optim` as a quotient of the estimate from (5.1) calculated for the true value of the unknown parameter and the absolute value of the objective function calculated at the starting point of the optimisation. This choice causes that the algorithm terminates if the calculated functional values in the vertices of the simplex satisfy that the difference of the largest and the smallest value is less than or equal to the estimate from (5.1) calculated for the true value of the unknown parameter.

Remark. Described stopping criterion is a viable option mainly in simulation studies, when the true value of the parameter is known. In practice, other stopping criteria are more appropriate. For example an iterative approach - first run the

algorithm with a benevolent tolerance for the difference of functional values in the worst and the best vertex, once the algorithm converges, estimate the variance with which we calculate the objective function on the neighbourhood of the point to which we converged, use this estimate to update the tolerance and run the algorithm again. Furthermore, a suitable stopping criterion is to stop the algorithm once the vertices of the simplex are sufficiently close or once the algorithm does not improve the value in the best vertex during a specified number of iterations.

6. Simulation study

In this chapter we would like to demonstrate usage of estimation methods described in previous chapters, namely the likelihood function (see Theorem 6 or Theorem 10) and the partial likelihood (see (2.6)).

First, let us introduce a parametric family of models we will be interested in. We will consider models in which the conditional intensity function is again given by (3.1) and in which lifetimes of objects $(M_n)_{n=1}^{\infty}$ are i.i.d. random variables independent of both times of births as well as positions in space, i.e. independent of $(T_n, X_n)_{n=1}^{\infty}$, with distribution given by a density $f_\gamma(m)$. Furthermore, we will suppose that the interaction functions g and h satisfy that $g(t) \equiv c$ and that

$$h(x) = \begin{cases} 0, & \text{if} \quad x \leq \delta, \\ 1 + \alpha \exp\left(-\frac{(x-\delta)^2}{\beta}\right), & \text{if} \quad x > \delta, \end{cases}$$

where $c > 0$, $\delta > 0$, $\alpha \geq -1$ and $\beta > 0$. We will consider models in which positions of objects lie in a unit square, i.e. we will put $W = [0,1]^2$.

In our simulation we will assume that parameters c and δ are known and that parameters α and β need to be estimated. We would like to assess the accuracy with which we estimate these parameters by particular methods in different time windows. That is, for every considered choice of parameters we will simulate $N = 100$ realisations from the model determined by this combination of parameters in some time window $[0, T]$. Then, we will estimate α and β from every one of those 100 simulated realisations (either by the likelihood function or by the partial likelihood), after which we will estimate the relative bias and the relative mean squared error of our estimators.

Let us denote by $\hat{\alpha}_i$ the estimate of α calculated from the i-th simulated realisation. Then we estimate the relative bias of the estimator of α (when $\alpha \neq 0$) by

$$\text{rbias}(\alpha) = \frac{\frac{1}{N}\sum_{i=1}^{N} \hat{\alpha}_i - \alpha}{|\alpha|}$$

and the relative mean squared error of the estimator of α by

$$\text{rMSE}(\alpha) = \frac{\frac{1}{N}\sum_{i=1}^{N} (\hat{\alpha}_i - \alpha)^2}{\alpha^2}.$$

Similarly for β.

Let us remark that for a given combination of parameters we did not simulate N realisations for every considered time window separately. We simulated N realisations only in the longest considered time window and subsequently restricted them to shorter time windows as required. In addition, let us remark that in our simulations we were choosing the starting point for the optimisation algorithm so that its coordinates were approximately three times smaller or greater than the corresponding coordinates of the true parameter. This corresponds to the incorporation of some expert knowledge regarding the relevant scale of the parameter values.

We simulated from 6 models – 4 spatially repulsive (Table 6.1, Table 6.2, Table 6.3, Table 6.4) and 2 spatially attractive (Table 6.5, Table 6.6).

The first thing we can notice from all the tables is that as the considered time window extends, our estimates (by the likelihood as well as the partial likelihood) get more accurate. This is completely natural, because when the time window extends, more objects are observed, and therefore more information is available to estimate the unknown parameters.

Similarly, we would expect estimates by the likelihood function to be more accurate than the corresponding estimates by the partial likelihood. This is because while in the partial likelihood we only consider conditional densities of positions of observed objects (see (2.5)), in the likelihood function we consider also conditional densities of times of births of observed objects (can be seen in (2.4)). Thus, when we estimate unknown parameters by the likelihood function, we use more information than when we estimate them by the partial likelihood. Indeed, when we compare estimates of relative MSEs for likelihood and partial likelihood between rows corresponding to same time windows, we can notice that estimates by the likelihood function are generally more accurate. However, there are some exceptions. For instance, in Table 6.4 we can see that rMSE(β) corresponding to estimates by the partial likelihood from simulated realisations in time window $[0, 100]$ is 0.857, whereas the value corresponding to estimates by the likelihood is 2.607. Curves corresponding to the estimates of unknown parameters from the simulated realisations are shown in Figure 6.1. In both plots in Figure 6.1 there are two curves which are extremely flat. They correspond to two largest estimates of β – values 0.397 and 0.481 in case of the partial likelihood and values 0.678 and 0.925 in case of the likelihood. When we calculate rMSE(β) without these two largest estimates, we obtain value 0.301 instead of 0.857 in case of the partial likelihood and value 0.368 instead of 2.607 in case of the likelihood. Value 0.368 is still larger than 0.301 but not as much as it was before. Other situations, where rMSE(β) for the likelihood was larger than the corresponding rMSE(β) for the partial likelihood can be seen in Table 6.3, Table 6.4 and Table 6.5, always in rows corresponding to time window $[0, 50]$. This situation is always caused by a couple of extreme estimates of β by the likelihood function, which considerably increased the value of rMSE(β). From our simulations it seems that estimation by the likelihood function might be more susceptible to this occasional extreme estimate of β.

Moreover, we can observe that rbias(α) is in case of estimation by the likelihood function very small in virtually all simulations. A similar thing can be said about rbias(α) in case of estimation by the partial likelihood but only in strongly repulsive models. Parameter β is a scaling parameter and is more difficult to estimate, because we need to see pairs of points at different distances from each other to assess the gradual changes in the interaction function h. Hence, rbias(β) and rMSE(β) is for short time windows typically large (because the number of observed points is small) but decreases with increasing length of the observation window.

Let us add a short observation about qualitative behaviour of the estimates by the likelihood function and by the partial likelihood. From plots of h functions corresponding to the estimates of α and β (such as the ones in Figure 6.1) it seems that the envelopes around the theoretical h function are of slightly different nature in case of estimation by the partial likelihood than in case of estimation by the likelihood. In case of the likelihood, envelopes tend to get narrower somewhere

around the point at which the theoretical h function increases (or decreases) most sharply. In Figure 6.2 we can see plots of h functions corresponding to the estimates of $\alpha = -0.6$ and $\beta = 0.03$ by the likelihood (realisations from which unknown parameters were estimated were simulated in time window $[0, 150]$) separated into two plots according to the value of the estimate of α (h functions corresponding to the estimates for which $\hat{\alpha}_i > -0.6$ were included in the upper plot, whereas h functions corresponding to the estimates for which $\hat{\alpha}_i \leq -0.6$ were included in the lower plot). Similarly for Figure 6.3, in which h functions corresponding to the estimates by the partial likelihood are displayed. We can notice that in Figure 6.2, most of the simulated h functions "cross" the true h function (most of the simulated curves corresponding to $\hat{\alpha}_i > -0.6$ satisfy that at $x = 0.3$ they are already below the true h function, whereas most of the simulated curves corresponding to $\hat{\alpha}_i \leq -0.6$ satisfy that at $x = 0.3$ they are already above the true h function). On the contrary, simulated curves in Figure 6.3 do not have such a strong tendency to "cross" the true h function. Described behaviour might, to some extent, be explained by the fact that the likelihood function considers not only conditional densities of positions but also conditional densities of times of births of observed objects. The time at which a new object is born depends on the integral of the conditional intensity function (see the algorithm above Theorem 7), which might explain why the simulated h functions corresponding to the estimates of α and β by the likelihood tend to "cross" the true h function. Underestimation of the values of the true h function for smaller values of x results in smaller values of the integrals of the conditional intensity function, an effect which needs to be corrected by overestimation of the values of the true h function for larger values of x.

This observation also explains why the estimation by the likelihood function appeared to be more susceptible to the occasional extreme estimate of β. Bond between α and β in the integral of the conditional intensity function in the likelihood function could have caused these extreme estimates. Similarly, it explains why in attractive models in estimation of α the likelihood function performed so much better than the partial likelihood. It prevented estimates of α from being too large as this would also extremely increase the integral of the conditional intensity function.

Estimates of relative bias and relative MSE					
Partial likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
30	0.032	0.003	0.190	0.319	46.65
50	0.020	0.002	0.099	0.123	80.34
100	0.012	0.000	0.075	0.046	165.71
150	0.010	0.000	0.078	0.035	248.65
Likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
30	0.033	0.004	0.238	0.252	46.65
50	0.021	0.002	0.142	0.073	80.34
100	0.010	0.000	0.070	0.021	165.71

Table 6.1: Estimates of relative biases and relative MSEs in particular time windows when $\alpha = -1$ and $\beta = 0.02$ in addition to average numbers of observed objects in these time windows. Lifetimes from $U(3.5, 6.5)$ distribution, $\delta = 0.025$ and $c = 1.15$.

Estimates of relative bias and relative MSE					
Partial likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
30	-0.018	0.014	0.044	0.328	43.03
50	-0.018	0.010	0.005	0.115	73.55
100	-0.006	0.004	0.007	0.070	147.52
150	-0.001	0.002	0.006	0.046	221.14
Likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
30	-0.002	0.013	0.154	0.161	43.03
50	-0.010	0.009	0.062	0.054	73.55
100	-0.002	0.004	0.045	0.024	147.52

Table 6.2: Estimates of relative biases and relative MSEs in particular time windows when $\alpha = -0.85$ and $\beta = 0.05$ in addition to average numbers of observed objects in these time windows. Lifetimes from $U(3, 6)$ distribution, $\delta = 0.05$ and $c = 1.25$.

Estimates of relative bias and relative MSE					
Partial likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	0.018	0.077	0.335	0.750	81.57
100	-0.012	0.032	0.092	0.254	168.48
150	-0.003	0.022	0.066	0.141	254.89
200	-0.008	0.013	0.027	0.090	342.19
Likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	0.016	0.059	0.362	0.986	81.57
100	-0.004	0.028	0.115	0.171	168.48
150	0.002	0.018	0.092	0.092	254.89

Table 6.3: Estimates of relative biases and relative MSEs in particular time windows when $\alpha = -0.6$ and $\beta = 0.03$ in addition to average numbers of observed objects in these time windows. Lifetimes from $\Gamma(\text{shape} = 10, \text{scale} = 0.5)$ distribution, $\delta = 0.05$ and $c = 1.15$.

Estimates of relative bias and relative MSE					
Partial likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	-0.152	0.228	0.597	7.426	46.51
100	-0.043	0.069	0.197	0.857	93.20
150	-0.049	0.051	0.143	0.695	140.20
200	-0.048	0.037	0.074	0.221	186.95
Likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	-0.078	0.144	0.591	13.951	46.51
100	0.001	0.054	0.412	2.607	93.20
150	-0.004	0.042	0.190	0.530	140.20

Table 6.4: Estimates of relative biases and relative MSEs in particular time windows when $\alpha = -0.4$ and $\beta = 0.07$ in addition to average numbers of observed objects in these time windows. Lifetimes from $\Gamma(\text{shape} = 8, \text{scale} = 1)$ distribution, $\delta = 0.1$ and $c = 1.135$.

Estimates of relative bias and relative MSE					
Partial likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	0.489	2.331	0.209	1.278	30.54
100	0.292	0.954	0.271	1.012	60.46
150	0.147	0.319	0.185	0.501	90.36
200	0.105	0.164	0.139	0.375	120.79
Likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	-0.041	0.183	0.720	9.520	30.54
100	-0.034	0.040	0.102	0.195	60.46
150	-0.031	0.028	0.062	0.108	90.36

Table 6.5: Estimates of relative biases and relative MSEs in particular time windows when $\alpha = 1.7$ and $\beta = 0.1$ in addition to average numbers of observed objects in these time windows. Lifetimes from $U(3.5, 6.5)$ distribution, $\delta = 0.05$ and $c = 0.55$.

Estimates of relative bias and relative MSE					
Partial likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	0.230	0.490	0.118	0.409	45.78
100	0.066	0.075	-0.014	0.057	108.93
150	0.015	0.036	-0.009	0.034	170.20
200	0.029	0.020	0.016	0.027	235.59
Likelihood					
T	rbias(α)	rMSE(α)	rbias(β)	rMSE(β)	Avg. # obs.
50	0.011	0.021	-0.065	0.081	45.78
100	0.015	0.009	-0.066	0.035	108.93

Table 6.6: Estimates of relative biases and relative MSEs in particular time windows when $\alpha = 2.5$ and $\beta = 0.05$ in addition to average numbers of observed objects in these time windows. Lifetimes from $\Gamma(\text{shape} = 6.5, \text{scale} = 1)$ distribution, $\delta = 0.075$ and $g(t) = 0.65\, \mathit{1}[t > 0.5]$.

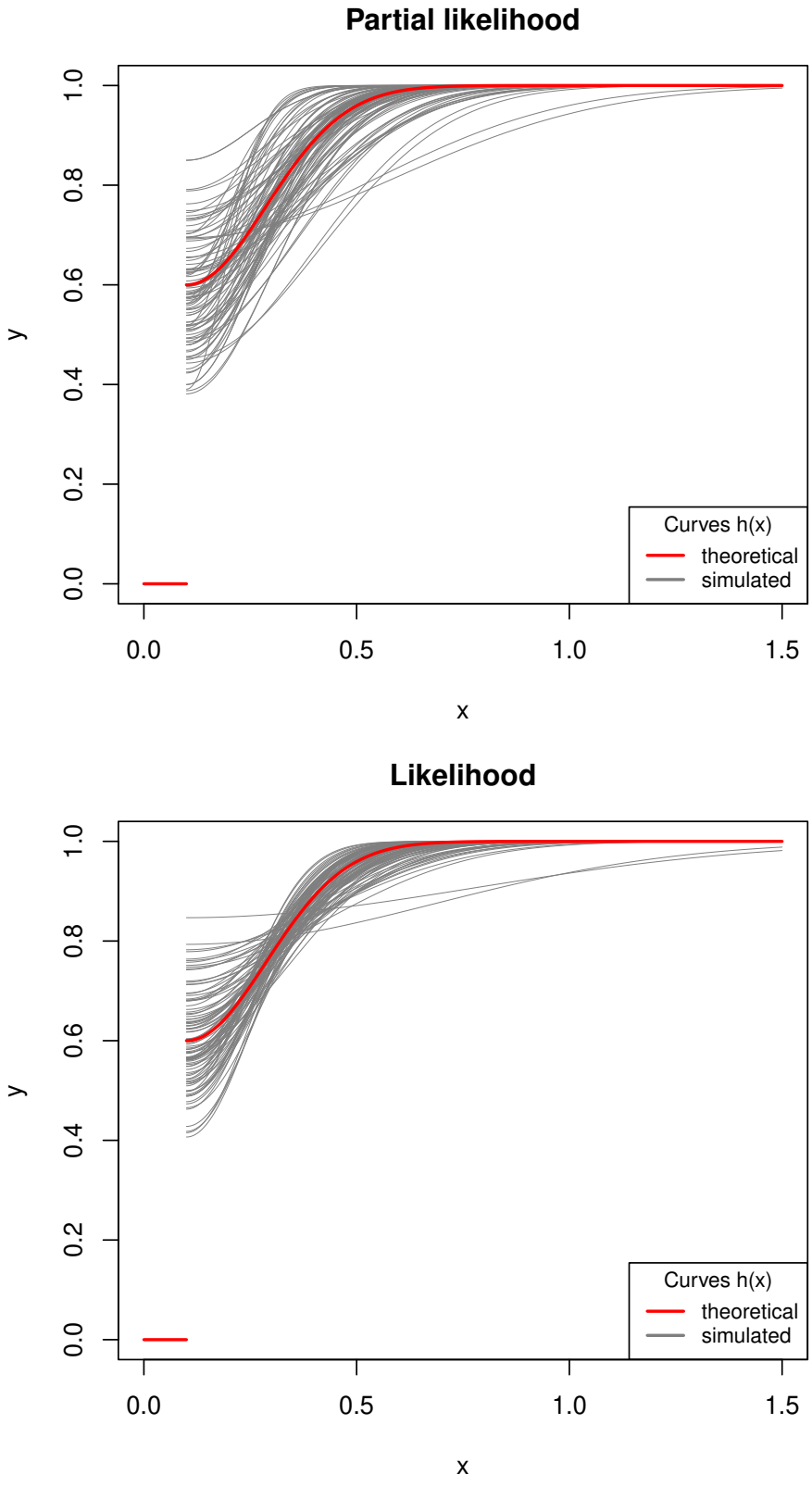
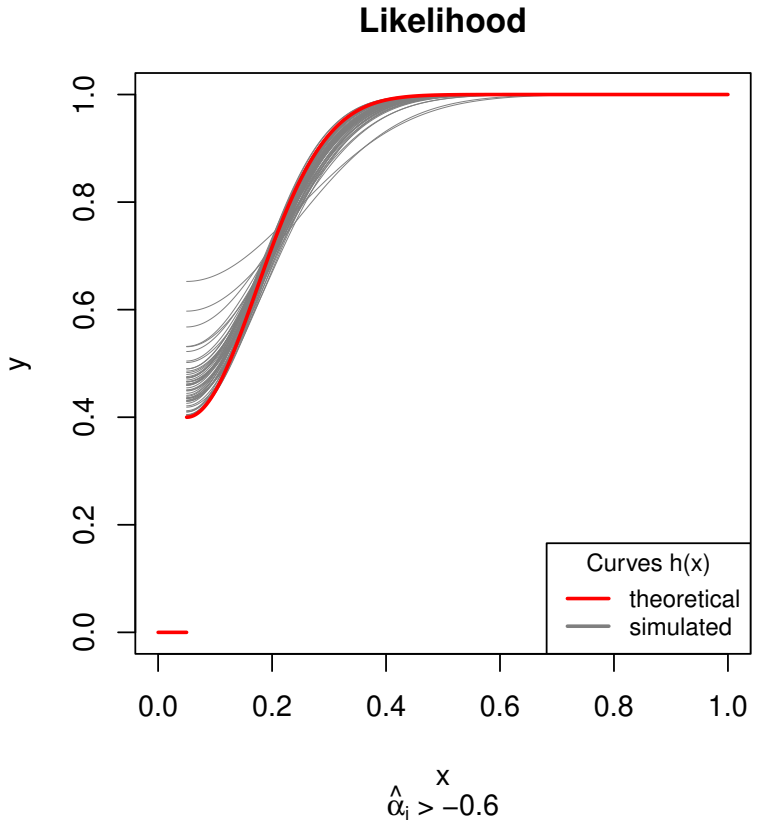


Figure 6.1: Plots of h functions corresponding to estimates of $\alpha = -0.4$ and $\beta = 0.07$ by both the partial likelihood as well as the likelihood, realisations from time window $[0, 100]$.



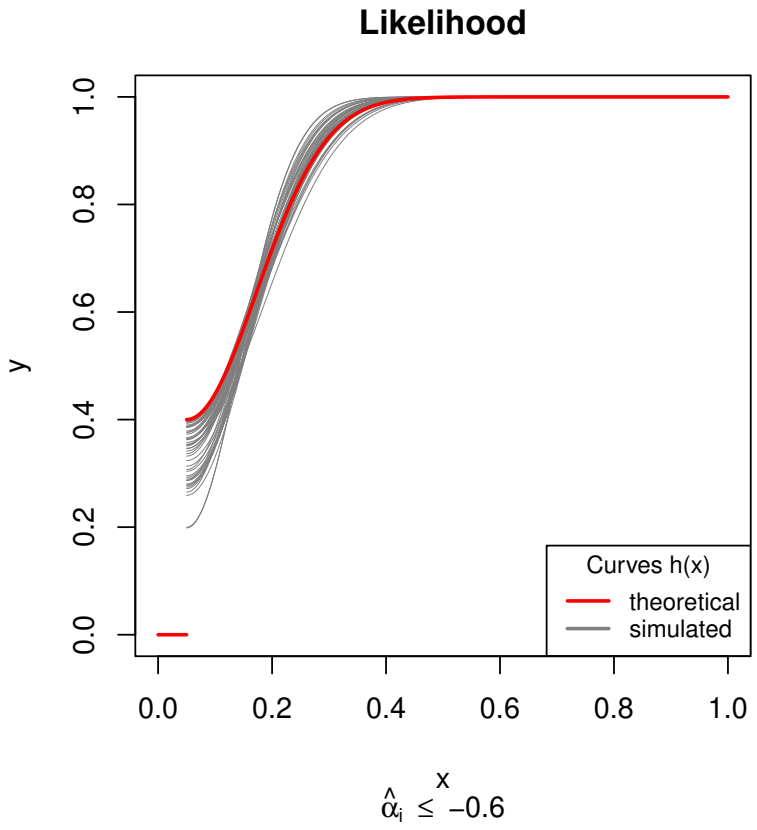
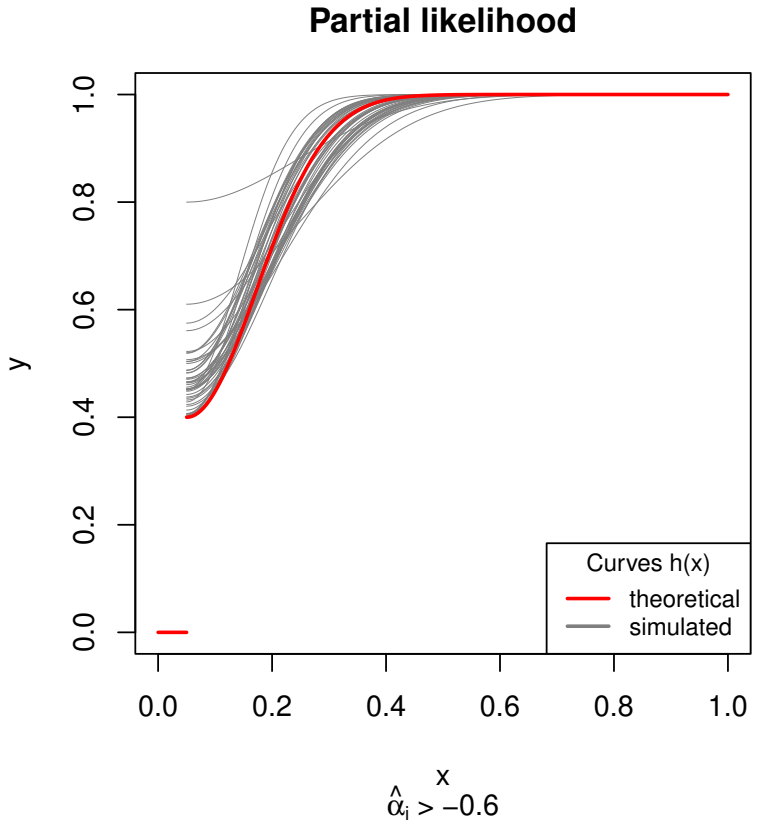


Figure 6.2: Plots of h functions corresponding to estimates of $\alpha = -0.6$ and $\beta = 0.03$ by the likelihood, realisations from time window $[0, 150]$.



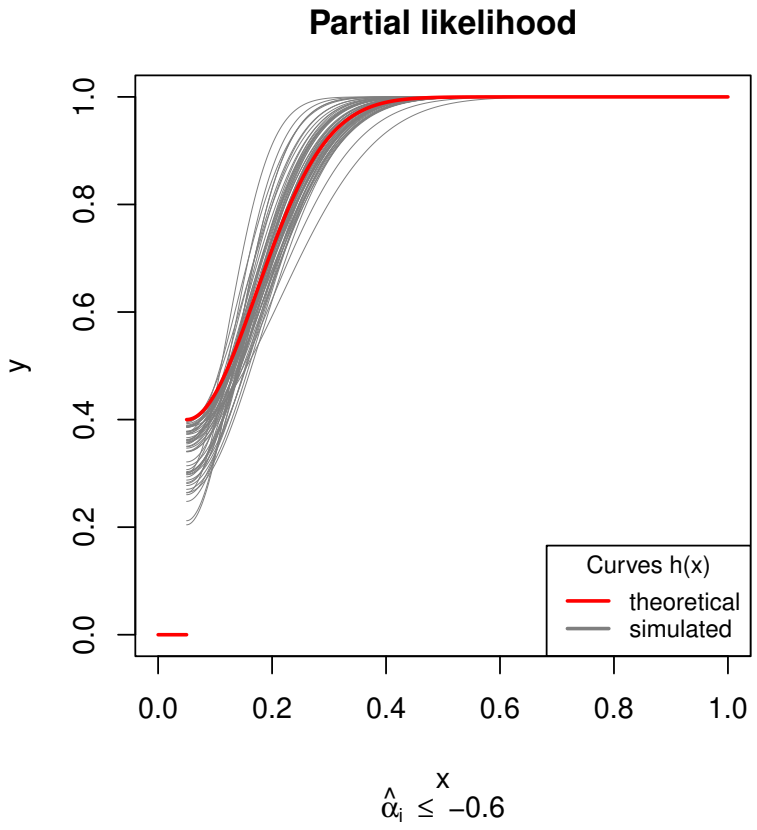


Figure 6.3: Plots of h functions corresponding to estimates of $\alpha = -0.6$ and $\beta = 0.03$ by the partial likelihood, realisations from time window $[0, 150]$.

In Chapter 4 we have discussed how to estimate parameters of the distribution of lifetimes in the presence of censoring. We will now describe a short simulation demonstrating estimation of distribution of lifetimes in extending time windows $[0, T]$ when (A) using lifetimes of all objects observed in this time window (included in the likelihood as in Theorem 10) as well as when (B) using lifetimes of only those objects whose exact lifetimes were observed, i.e. of objects whose times of death are less than or equal to T. In the first part of this chapter when we generated $N = 100$ realisations from the model in which $\alpha = -0.4$ and $\beta = 0.07$, lifetimes of objects were generated from the gamma distribution with the shape parameter $k = 8$ and the scale parameter $\theta = 1$, i.e. from the distribution given by the density

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} 1_{(0,\infty)}(x),\ x \in \mathbb{R},$$

where $k = 8$ and $\theta = 1$. We will use these 100 simulated realisations (restricted to different time windows $[0, T]$) to estimate k and θ in two different ways: (A) and (B). We will refer to (A) as having a correction for censored lifetimes, and to (B) as having no correction for censored lifetimes. Subsequently, we will use the calculated estimates of k and θ to estimate the relative bias and the relative mean squared error of the estimators of k and θ (as in the first part of this chapter with α and β). These estimates, together with the average number of observed objects and the average proportion of censored (time of death is greater than T) observations can be seen in Table 6.7. In the upper part of the table we can see that the estimates by (A) get more accurate as the length of the time window increases. The same conclusion can be reached for estimates by (B) from the lower part of the table. This is completely natural because as the length of the time window increases more observations are available (average number of observations in individual time windows can be seen in sixth column of the table). Comparing the corresponding values in the upper and in the lower part of the table we can observe the following. As the length of the time window increases, estimates by (A) and (B) become more and more similar. This is caused by the fact that with extending time window proportion of censored observations decreases, which means that the effect of censoring weakens. Furthermore, we can observe that estimates by (B) are generally more biased than estimates by (A).

Estimates of relative bias and relative MSE						
With correction for censored lifetimes						
T	rbias(k)	rMSE(k)	rbias(θ)	rMSE(θ)	Avg. # obs.	Avg. prop. cens.
25	0.204	0.211	-0.068	0.137	23.18	0.316
35	0.128	0.107	-0.050	0.074	32.42	0.233
50	0.060	0.054	-0.009	0.055	46.51	0.155
100	0.037	0.025	-0.017	0.023	93.20	0.080
200	0.007	0.012	0.003	0.011	186.95	0.042
Without correction for censored lifetimes						
T	rbias(k)	rMSE(k)	rbias(θ)	rMSE(θ)	Avg. # obs.	Avg. prop. cens.
25	0.277	0.400	-0.159	0.119	23.18	0.316
35	0.154	0.135	-0.096	0.082	32.42	0.233
50	0.067	0.056	-0.039	0.054	46.51	0.155
100	0.036	0.025	-0.026	0.023	93.20	0.080
200	0.007	0.012	-0.003	0.010	186.95	0.042

Table 6.7: Estimates of relative biases and relative MSEs, average numbers of observations and average proportions of censored observations in particular time windows when $k = 8$ and $\theta = 1$.

7. Real data application

In this chapter we will demonstrate application of the partial likelihood to the real data. We are thankful to RNDr. Zdeněk Janovský, Ph.D., from the Faculty of Science, Charles University, for providing us with the dataset. We are interested in reproductive patterns of Succisa pratensis (a flowering plant also known as devil's-bit, which can be seen in Figure 7.1) in an observed meadow. Our data come from 6 observation squares in the meadow, each having dimensions 100cm $\times$ 100cm. All the squares were observed once a year. Furthermore, for all the squares, the year of the first observation is 2008, whereas the year of the last observation is 2019. For each observed plant, we know its location in a particular square and the years in which it was observed. Population of plants associated with one particular square can be seen in Figure 7.2 (positions of plants which were observed in years $2008 - 2013$) and Figure 7.3 (positions of plants which were observed in years $2014 - 2019$).

The number of plants which were born in a particular year is quite variable. This is caused by the fact that in some years conditions for plants were simply more favourable than in others, e.g. some years were unusually dry. In addition, the number of years from which we have available data is quite limited. Hence, we will only try to model positions in which new plants are born but not the time at which they are born (or rather, due to discretisation, the number of new plants which occur from one year to another). Thus, we will only use the partial likelihood instead of the full likelihood function (Theorem 6) in our analysis.

It is known that Succisa pratensis can reproduce either by seeds or by creating clones. The seeds are not adapted to being spread by the wind. Hence, we would expect that for a given plant its daughter is most likely to occur up to some distance δ_1 from this plant (where δ_1 is the distance at which the ability of clonal reproduction fades away) because up to this distance effects of reproduction by seeds and clonal reproduction combine. Furthermore, we would expect that positions at distances greater than δ_1 but less than δ_2 (where δ_2 is the distance at which the ability of reproduction be seeds fades away) are also quite likely but not as likely as the ones at distances less than δ_1 because at these positions effects of clonal reproduction are already very small whereas effects of reproduction by seeds are still considerable. Finally, we would expect that positions at distances greater than δ_2 from the given plant are least likely or even forbidden.

In the observed meadow, however, an additional external effect is present – turning of the hay. This causes that the seeds get randomly dispersed. Hence, it is reasonable to assume that the seeds are equally likely to appear in all positions in the meadow (apart from positions occupied by already existing plants). This means that it might be reasonable to model reproductive patterns of these plants by a model similar to the ones presented in Chapter 3 or Chapter 6.

So far we have always assumed that times of events, or times of births of objects, constitute a simple point process. In the current situation we do not know the exact times at which the observed plants were born, we only know in which year they were observed for the first time. We will deal with this problem by calculating the partial likelihood (see (2.6)) in a way that in the term associated with every particular plant we will consider the conditional intensity function

Figure 7.1: Succisa pratensis (Wikimedia Commons, distributed under a CC-BY-SA-3.0 licence).

determined by those objects which were alive in the previous year. This approach is quite natural because when two plants are born in the same year they probably did not arise from each other but from the plants which were alive in the previous year. Furthermore, plants which were observed already in the first year for which data are available will not be included in the partial likelihood since we do not know how the observed squares looked in the previous year. In addition, we will aim at estimating unknown parameters in the proposed parametric models using the data from all 6 observed squares at the same time by maximising the product of partial likelihoods associated with every individual square. This step will be justified by the independence of these squares and by the fact that conditions in all of them are practically identical.

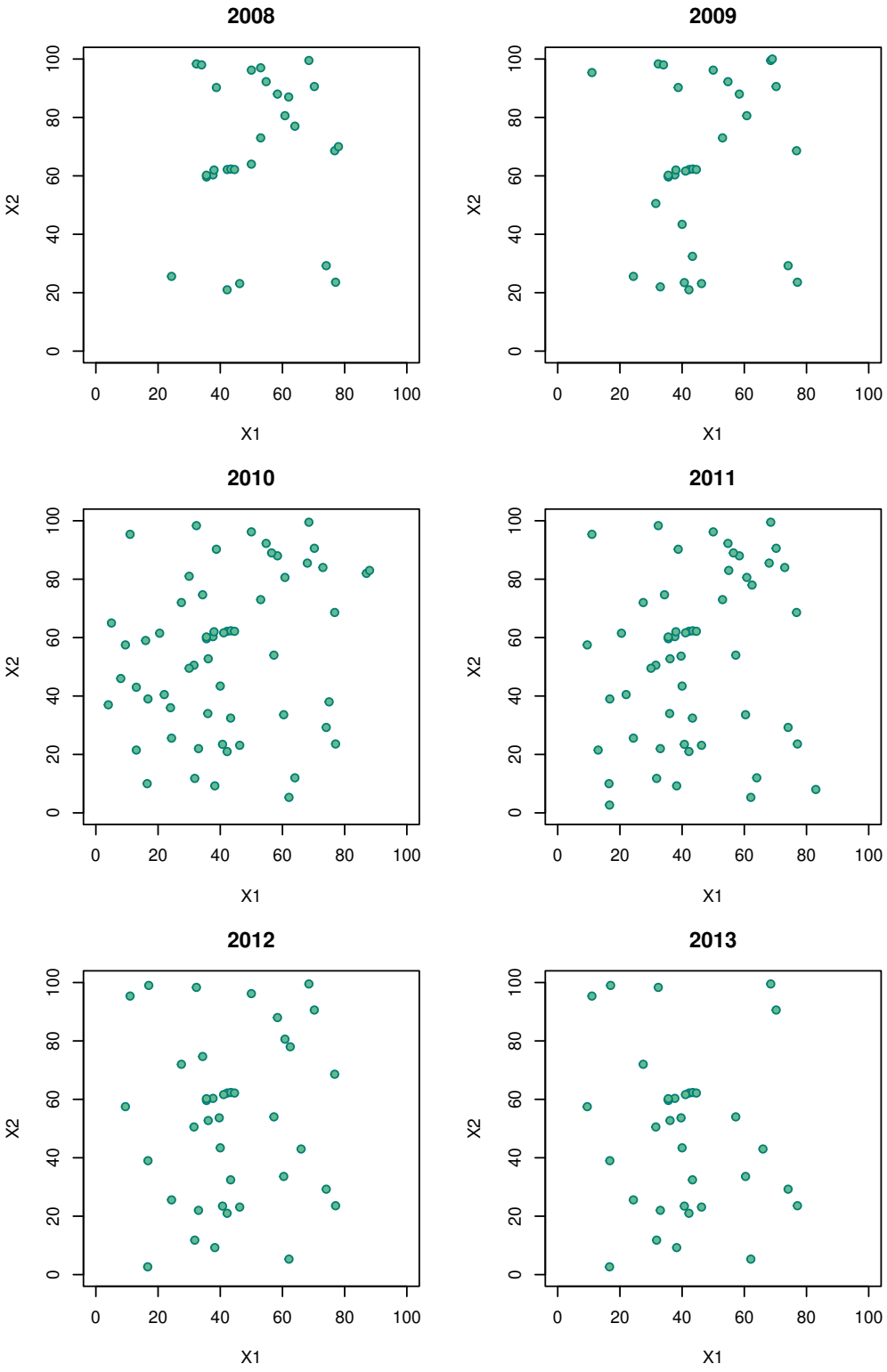


Figure 7.2: Population of plants observed in square 1 in years 2008 – 2013.

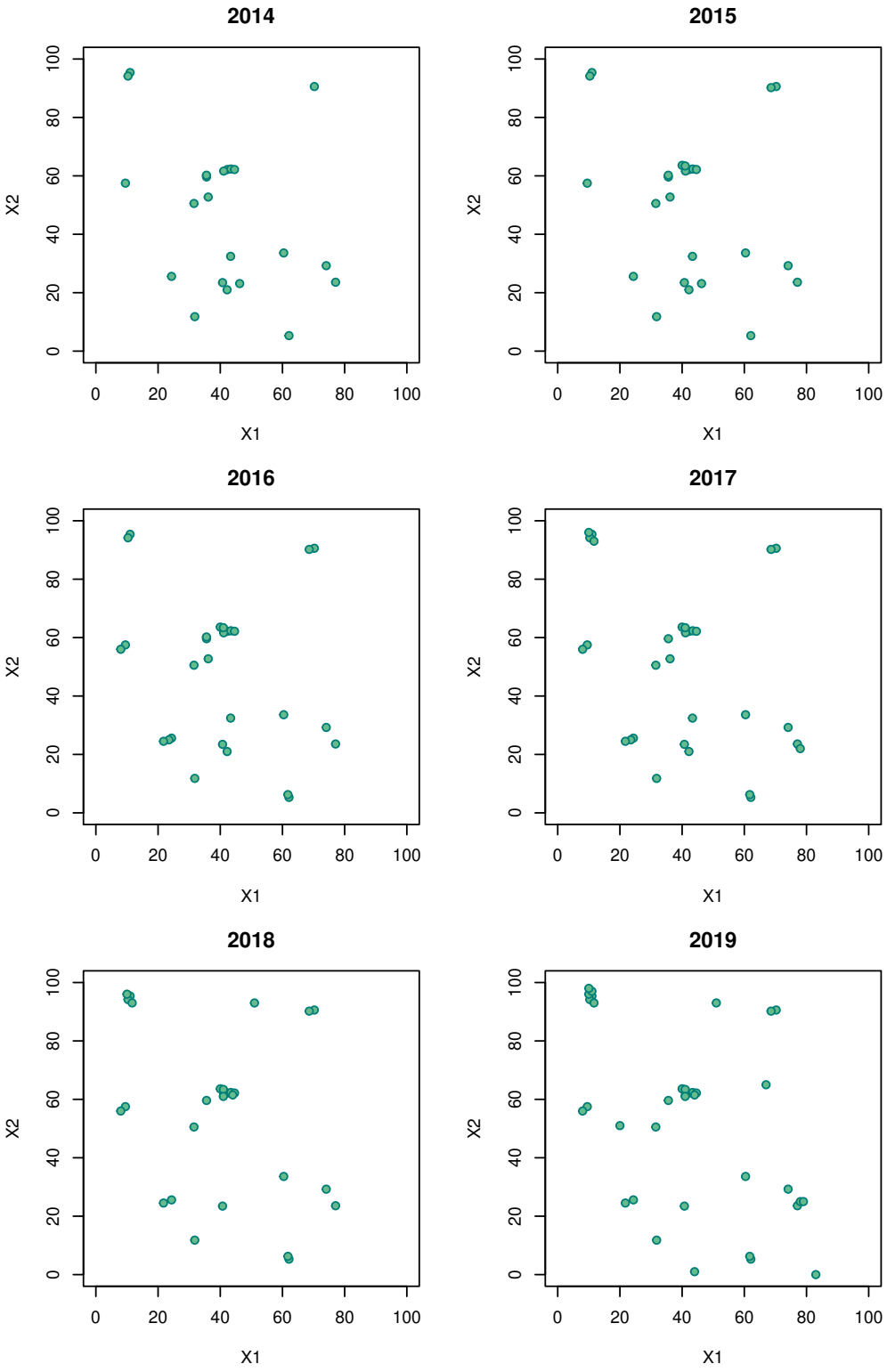


Figure 7.3: Population of plants observed in square 1 in years 2014 – 2019.

7.1 Multiplicative model

Let us start with the basic multiplicative model we are already acquainted with. We will slightly modify the specification of the conditional intensity function given by (3.1) so that it reflects the current situation discussed before. Suppose that the number of plants which were observed (in a particular square) in year $t-1$ is n_{t-1} and that they were observed in positions $x_1, \ldots, x_{n_{t-1}}$ then the conditional intensity function in year t and position x will be specified as

$$\lambda^*(t,x) = \prod_{i=1}^{n_{t-1}} h(\|x - x_i\|), \tag{7.1}$$

where h is again some interaction function from $[0,\infty)$ to $[0,\infty)$. Notice that the parameters of the interaction function g from (3.1) are not identifiable by the partial likelihood. Indeed, from the expression (2.6) it is obvious that the terms $g(|t-t_i|)$ get factored out. In the present situation we are interested only in spatial dynamics of the model and can thus omit these terms, which leads to (7.1). Furthermore, we will suppose that

$$h(x) = \begin{cases} 0, & \text{if} \quad x \leq \delta, \\ 1 + \alpha \exp\left(-\frac{(x-\delta)^2}{\beta}\right), & \text{if} \quad x > \delta, \end{cases}$$

where $\alpha \geq -1$ and $\beta > 0$ are unknown parameters and where $\delta = 0.6$ was chosen as the smallest observed distance in which two plants can coexist. We will refer to this model as the multiplicative model.

Let us first begin by estimating the unknown parameters from each observed square individually. Estimates from each individual square can be seen in Table 7.1. The corresponding h functions are plotted in grey in Figure 7.4. We can see that the estimated curves show some natural variability but all of them correspond to spatial attraction on similar scales. Next, we will try to accumulate the information included in individual squares and estimate the unknown parameters from all squares simultaneously (by maximising the product of partial likelihoods associated with every individual square). We obtain estimates $\hat{\alpha} = 2.5$ and $\hat{\beta} = 3.9$, the corresponding h function is plotted in red in Figure 7.4. We would like to evaluate how reliable our estimate is. We will estimate unknown parameters from each quintuple of observed squares (each time leaving the remaining square out). We would like these leave-one-out estimates to be relatively close to the cumulative estimate obtained from all 6 observed squares. These new estimates can be seen in Table 7.2, while the corresponding h functions are plotted in Figure 7.5. We can see that the h functions corresponding to the leave-one-out estimates reproduce the h function estimated from all 6 observed squares very closely, which means that behaviour of the plants in all the squares is very similar and there is no outlier. Let us remark that this procedure would not necessarily reveal an outlier if the given square contained extremely few points, which, however, is not our case.

Estimates in individual squares		
Estimate in square	$\hat{\alpha}$	$\hat{\beta}$
1	2.1	3.8
2	2.3	4.7
3	4.0	2.0
4	4.0	3.8
5	2.6	3.7
6	2.8	6.8
Cumulative estimate	2.5	3.9

Table 7.1: Estimates of α and β in the multiplicative model in individual observed squares.

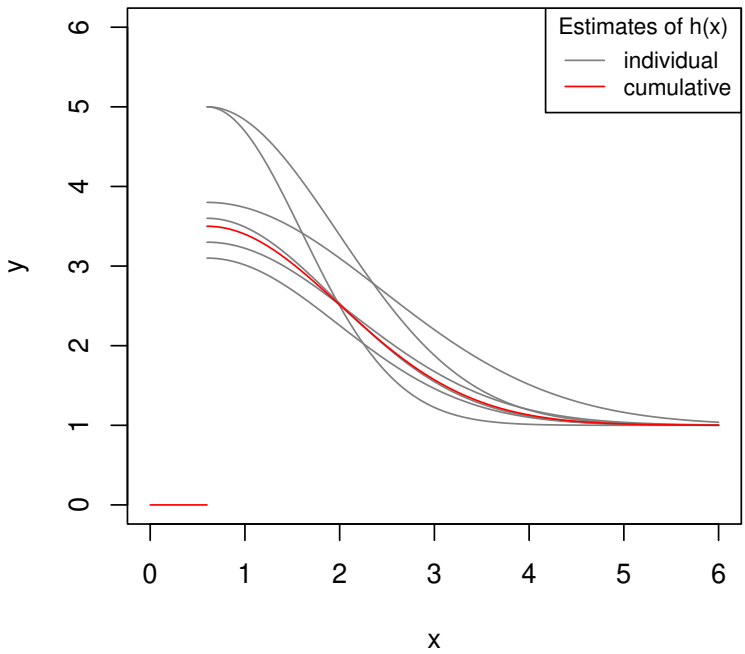


Figure 7.4: Estimated h functions in the multiplicative model in individual squares and h function estimated cumulatively from all observed squares.

Leave-one-out estimates		
Left out square	$\hat{\alpha}$	$\hat{\beta}$
1	2.8	3.9
2	2.5	3.8
3	2.4	4.1
4	2.3	3.9
5	2.4	4.0
6	2.6	3.3
Cumulative estimate	2.5	3.9

Table 7.2: Estimates of α and β in the multiplicative model when one of the observed squares is left out.

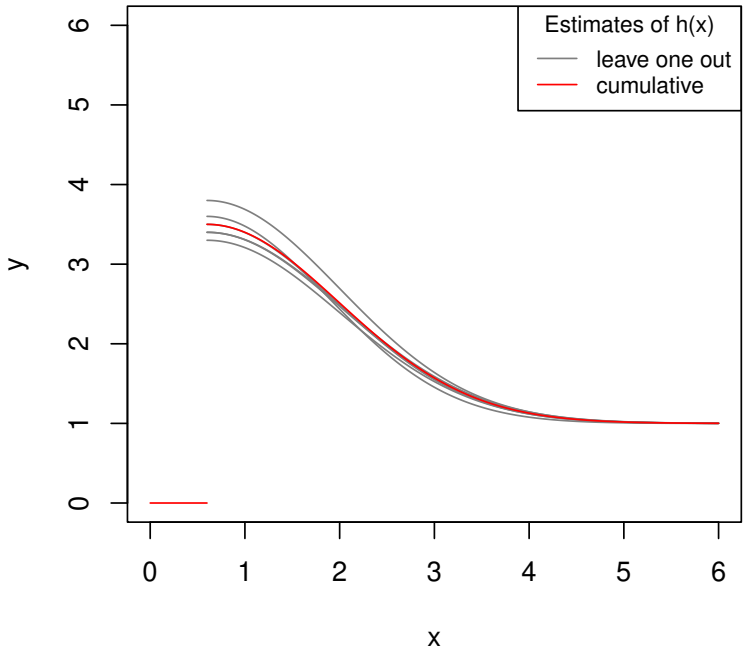


Figure 7.5: Estimated h functions in the multiplicative model when one of the observed squares is left out and h function estimated cumulatively from all observed squares.

We have estimated α and β, however, we do not know whether the resulting model is actually a suitable model for our data. In order to answer this question we would like to perform a Monte Carlo based goodness of fit test. First, we need to propose a suitable test statistic. One option is to consider some kind of nearest neighbour distances. In our observed data we will go through all objects whose year of first observation is greater than 2008 and for every such object we will calculate its distance from the configuration of points which were observed in the year preceding the year of first observation of this object. Similarly for the simulated realisations. For every object from the observed data whose year of first observation is greater than 2008 we will do the following: supposing that this object's year of first observation is t, we will generate a position from the density proportional to the conditional intensity function determined by those objects from our observed data which were alive in year $t-1$ (in the same square as the considered object), then we will calculate distance from the generated position to the configuration of points which determined the density from which this position was generated. Hence, in 1 simulated realisation we will generate exactly as many positions as is the number of objects in our data whose year of first observation is greater than 2008.

Remark. Let us emphasise that in simulated realisations it would be possible to simulate a new position in year t from the density induced by the simulated positions associated with year $t-1$. This approach, however, would increase variability of simulated realisations, leading to a test with smaller power. Therefore, we prefer generation of new positions from densities induced by the observed configurations of points.

Subsequently, we can perform a global envelope test [Myllymäki et al., 2017] for empirical cumulative distribution functions calculated from simulated distances and from distances observed in the data. Let us remark that we performed the test using function `global_envelope_test` from Myllymäki and Mrkvička [2020]. We simulated 500 realisations, which create an envelope, which can be seen (together with the empirical cumulative distribution function for the observed data) in Figure 7.6. The curve associated with the observed data is extreme (p-value of 0.004), compared to the simulated curves. That means that the proposed model does not describe observed data sufficiently well. Figure 7.6 also suggests how our estimated model differs from the observed data. In the observed data smaller distances (2cm $-$ 8cm) are more frequent.

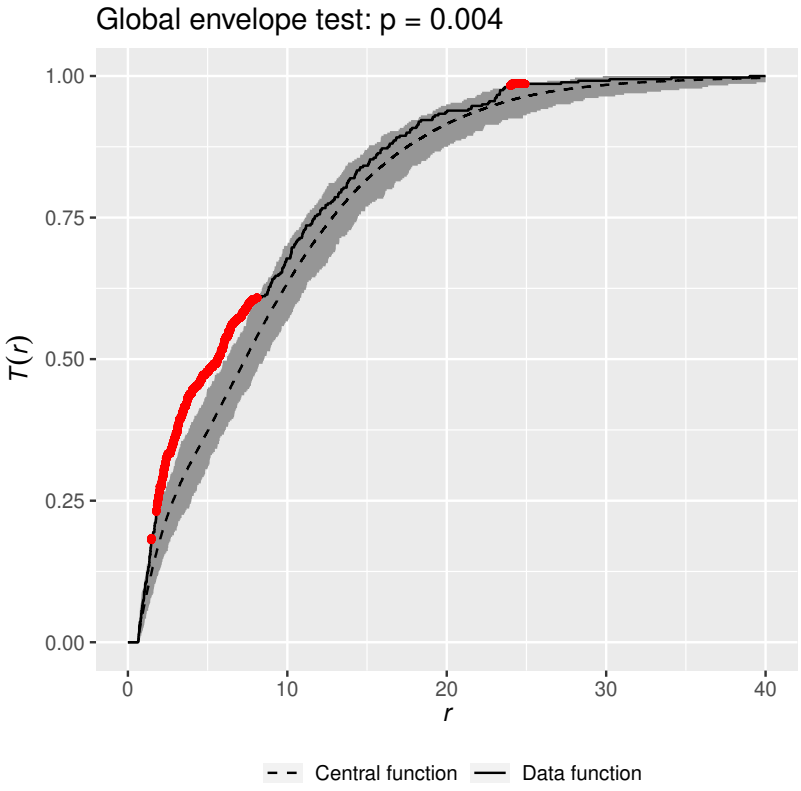


Figure 7.6: Goodness of fit test for the estimated multiplicative model.

7.2 Additive model

It seems that the multiplicative model (7.1) is not an appropriate model to describe in which positions new plants tend to occur. We believe that the reason why the multiplicative model is not really suitable is that in the observed data there are several relatively large clusters ($5-7$ plants in very close proximity), which do not produce a new plant in the following year. If the estimated h function had a much greater jump at 0.6, then the density proportional to the conditional intensity function would have extremely large values on the neighbourhoods of these clusters, hence, we would expect new plants to be born almost exclusively in these neighbourhoods, which is not always the case as we have said. Consequently, in our estimated model isolated plants might not have such a strong attractive power as they might have in reality. Some kind of additive model instead of the multiplicative one might be less susceptible to this effect.

Let us now define an additive model by specifying the conditional intensity function in the following way. Suppose that the number of plants which were observed (in a particular square) in year $t-1$ is n_{t-1} and that they were observed in positions $x_1, \ldots, x_{n_{t-1}}$ (let us denote $M = \{x_1, \ldots, x_{n_{t-1}}\}$) then the conditional intensity function in year t and position x will be specified as

$$\lambda^*(t,x) = \begin{cases} 0, & \text{if} \quad dist(x,M) \leq 0.6, \\ 1 + \sum_{i=1}^{n_{t-1}} h(\|x - x_i\|), & \text{if} \quad dist(x,M) > 0.6, \end{cases}$$

where

$$h(x) = \begin{cases} 0, & \text{if} \quad x \leq 0.6, \\ \alpha \exp\left(-\frac{(x-0.6)^2}{\beta}\right), & \text{if} \quad x > 0.6. \end{cases}$$

We will refer to this model as the additive model. Again, we will begin by estimating unknown parameters α and β from each observed square individually. Estimates from each individual square can be seen in Table 7.3. The corresponding h functions are plotted in grey in Figure 7.7. Immediately we can see that in the estimated models the attractive power of isolated plants is much larger than it was in the multiplicative model. Let us now combine information contained in all 6 observed squares and estimate the unknown parameters from all these squares simultaneously (by maximising the product of partial likelihoods associated with every individual square). We obtain cumulative estimates $\hat{\alpha} = 11.6$ and $\hat{\beta} = 2.7$, the corresponding h function is plotted in red in Figure 7.7. Again, we would like to evaluate how reliable our estimate is. We will estimate unknown parameters from each quintuple of observed squares (each time leaving the remaining square out). We would like these leave-one-out estimates to be relatively close to the cumulative estimate obtained from all 6 observed squares. These new estimates can be seen in Table 7.4, while the corresponding h functions are plotted in Figure 7.8. We can see that the h functions corresponding to the leave-one-out estimates reproduce the h function estimated from all 6 observed squares very closely, which means that behaviour of the plants in all the squares is very similar and there is no outlier.

Estimates in individual squares		
Estimate in square	$\hat{\alpha}$	$\hat{\beta}$
1	11.2	2.3
2	7.1	2.5
3	13.1	1.3
4	12.1	2.4
5	12.9	3.6
6	13.6	4.8
Cumulative estimate	11.6	2.7

Table 7.3: Estimates of α and β in the additive model in individual observed squares.

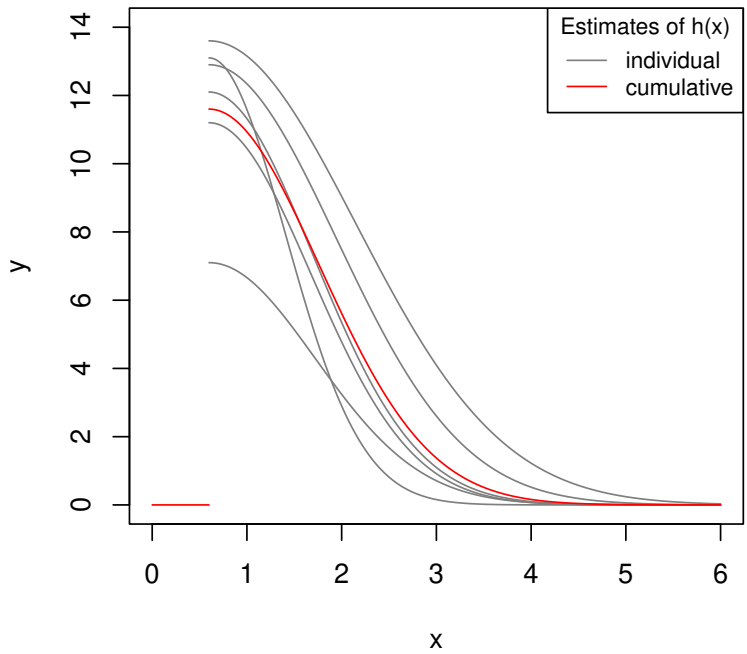


Figure 7.7: Estimated h functions in the additive model in individual squares and h function estimated cumulatively from all observed squares.

Leave-one-out estimates		
Left out square	$\hat{\alpha}$	$\hat{\beta}$
1	11.6	2.7
2	12.4	2.7
3	11.8	3.2
4	10.9	2.9
5	11.2	2.4
6	11.5	2.3
Cumulative estimate	11.6	2.7

Table 7.4: Estimates of α and β in the additive model when one of the observed squares is left out.

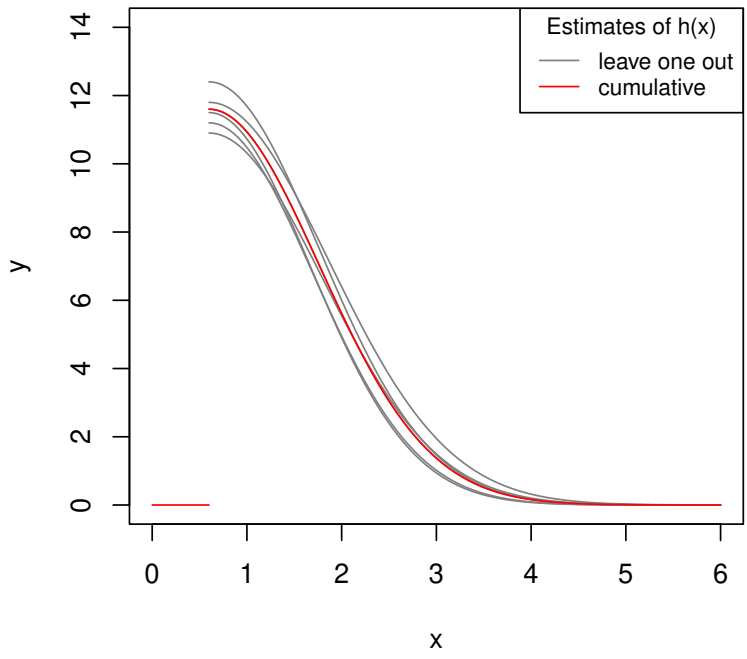


Figure 7.8: Estimated h functions in the additive model when one of the observed squares is left out and h function estimated cumulatively from all observed squares.

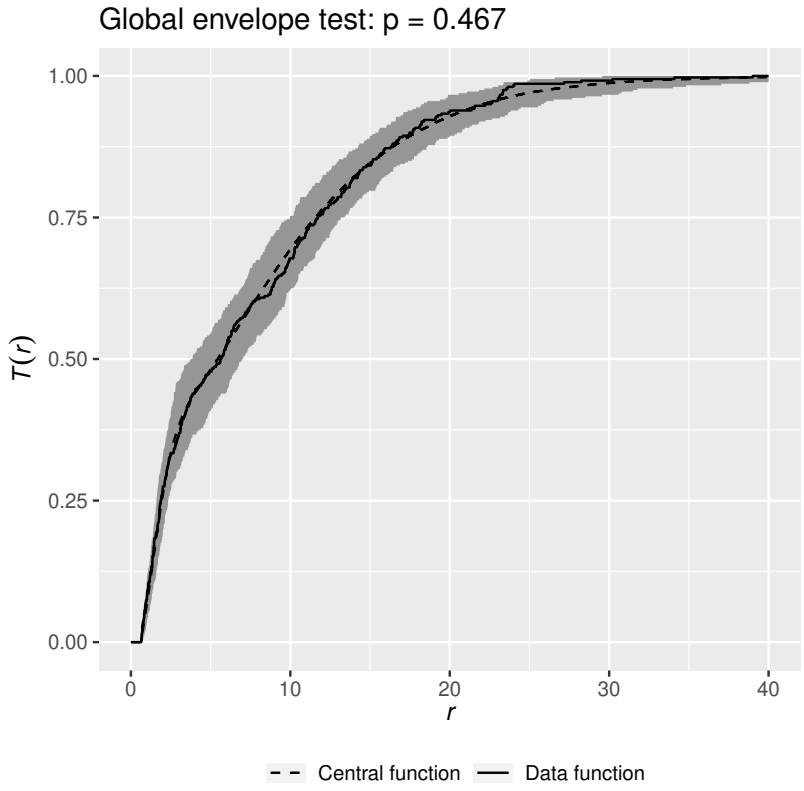


Figure 7.9: Goodness of fit test for the estimated additive model.

Similarly as in the case of multiplicative model, we would like to address the question whether the estimated model is actually a suitable model for the observed data. We performed a Monte Carlo based goodness of fit test exactly as in the case of multiplicative model, again simulating 500 realisations. An envelope created by the simulated empirical distribution functions of nearest neighbour distances to the last year's configuration of points in addition to the empirical distribution function associated with the observed data can be seen in Figure 7.9. We can see that the estimated additive model captures the behaviour of the empirical distribution function of the observed data much better (p-value of 0.467) than the estimated multiplicative model. Hence, this model seems to describe the spatial dynamics of propagation of the observed plants in an acceptable manner.

Conclusion

In this thesis, we studied point processes of objects with random lifetime. We started by presenting theory for purely temporal point processes, which we then generalised to spatial-temporal processes with random lifetimes by considering the corresponding spatial locations and lifetimes as marks. We derived sufficient conditions for the conditional intensity function to uniquely define (together with a model for lifetimes of objects) a spatial-temporal point process with random lifetimes. In addition, we derived the form of the likelihood function for spatial-temporal point patterns with observed lifetimes. Furthermore, we addressed the question of non-explosiveness of the proposed parametric models as well as the question of how the presence of censoring of lifetimes affects the form of the likelihood function. A part of the thesis was our own implementation of the generating algorithm and the (partial) likelihood function, which we used to assess behaviour of estimates of unknown parameters in considered parametric models in a simulation study. Finally, we demonstrated application of the partial likelihood to the real data. We proposed two types of models that describe the spatial dynamics of propagation of Succisa pratensis in an observed meadow, one with multiplicative structure and one with additive structure of interactions. Then, we used a Monte Carlo based goodness of fit test (based on construction of a global envelope for a suitable functional statistic) to assess how well these models describe observed data. The multiplicative model was rejected, whereas the additive one was not.

Bibliography

Russel E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7:1–49, 1998.

D. R. Cox. Partial Likelihood. *Biometrika*, 62(2):269–276, 1975.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. 2nd Edition. Springer, New York, 2003.

Peter J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. 3rd Edition. Chapmann and Hall/CRC, New York, 2013.

Peter J. Diggle, Irene Kaimi, and Rosa Abellana. Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data. *Biometrics*, 66(2):347–354, 2010.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. 2nd Edition. John Wiley and Sons, New York, 2002.

D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley and Sons, New York, 2011.

Jeffrey C. Lagarias, James A. Reeds, Margaret H. Wright, and Paul E. Wright. Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.

M. Myllymäki and T. Mrkvička. GET: Global Envelopes in R. *arXiv:1911.06583 [stat.ME]*, 2020. URL `https://arxiv.org/abs/1911.06583`. Accessed: 2022-04-23.

M. Myllymäki, T. Mrkvička, P. Grabarnik, H. Seijo, and U. Hahn. Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404, 2017.

J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 1965.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`. Accessed: 2022-04-23.

Jakob Gulddahl Rasmussen. Lecture Notes: Temporal Point Processes and the Conditional Intensity Function, 2018. URL `https://arxiv.org/abs/1806.00221`. Accessed: 2022-04-23.

List of Figures

List of Tables

A. Attachments

A.1 Sample run of the Nelder-Mead method

We attach a sample run of the Nelder-Mead method applied on the partial likelihood with the default stopping criterion. We show the progress of the algorithm between 1. and 17. function evaluation, between 61. and 75. function evaluation and between 292. and 307. function evaluation. In the attached run of the algorithm, rows starting with the word `Tracing` signify individual function evaluations. Details of these evaluations can be seen in the subsequent rows, i.e. rows starting with `[1]`. The first two numbers in these rows correspond to the coordinates of the point at which the objective function was evaluated, whereas the third number is the calculated functional value at this point (affected by the stochastic noise from the numerical integration). Remaining rows (the ones starting with the words `LO-REDUCTION`, `EXTENSION`, etc.) show which steps the algorithm took. We can see that after 70 function evaluations the algorithm was already very close to the point which was eventually declared as the achieved optimum. Nevertheless, it took 307 function evaluations for the algorithm to actually terminate.

```
> optim(c(0, 0.25), partial_likelihood, method="Nelder",
        control=list(fnscale=-1, parscale=c(-0.85, 0.05), trace=1))
  Nelder-Mead direct search function minimizer
Tracing fn(par, ...) on exit
[1] 0.00000 0.25000 3.71671
function value for initial parameters = -3.716710
  Scaled convergence tolerance is 5.53833e-08
Stepsize computed as 0.500000
Tracing fn(par, ...) on exit
[1] -0.42500  0.25000 14.84387
Tracing fn(par, ...) on exit
[1] 0.000000 0.275000 3.743526
BUILD              3 -3.716710 -14.843872
Tracing fn(par, ...) on exit
[1] -0.42500  0.27500 14.52651
Tracing fn(par, ...) on exit
[1] -0.31875  0.26875 12.52375
LO-REDUCTION       5 -3.743526 -14.843872
Tracing fn(par, ...) on exit
[1] -0.8500  0.2500 10.8101
Tracing fn(par, ...) on exit
[1] -0.63750  0.25625 16.41589
LO-REDUCTION       7 -14.526509 -16.415894
Tracing fn(par, ...) on exit
[1] -0.63750  0.23125 17.19978
Tracing fn(par, ...) on exit
[1] -0.743750  0.209375 17.370715
EXTENSION          9 -14.843872 -17.370715
Tracing fn(par, ...) on exit
[1] -0.956250  0.215625  4.121492
Tracing fn(par, ...) on exit
[1] -0.5578125  0.2414063 16.6428559
HI-REDUCTION      11 -16.415894 -17.370715
```

```
Tracing fn(par, ...) on exit
[1] -0.6640625  0.1945313 18.6536263
Tracing fn(par, ...) on exit
[1] -0.6773438  0.1636719 19.9761063
EXTENSION         13 -16.642856 -19.976106
Tracing fn(par, ...) on exit
[1] -0.8632812  0.1316406 21.1420068
Tracing fn(par, ...) on exit
[1] -1.016016e+00  7.675781e-02 -1.000000e+10
REFLECTION        15 -17.370715 -21.142007
Tracing fn(par, ...) on exit
[1] -0.7968750  0.0859375 24.3700604
Tracing fn(par, ...) on exit
[1] -0.82343750  0.02421875 21.38131210
.
.
.
.
.
Tracing fn(par, ...) on exit
[1] -0.87016888  0.06211344 25.39386972
LO-REDUCTION      61 -25.420880 -25.464197
Tracing fn(par, ...) on exit
[1] -0.86956965  0.06141885 25.39540081
Tracing fn(par, ...) on exit
[1] -0.87016888  0.06211344 25.38883918
Tracing fn(par, ...) on exit
[1] -0.86998292  0.06210081 25.39591302
Tracing fn(par, ...) on exit
[1] -0.86996914  0.06188191 25.33276593
SHRINK            65 -25.332766 -25.464197
Tracing fn(par, ...) on exit
[1] -0.86961100  0.06207553 25.33144553
Tracing fn(par, ...) on exit
[1] -0.86987960  0.06193032 25.38078106
HI-REDUCTION      67 -25.380781 -25.464197
Tracing fn(par, ...) on exit
[1] -0.86970054  0.06202713 25.41871120
Tracing fn(par, ...) on exit
[1] -0.86974530  0.06200293 25.38357195
LO-REDUCTION      69 -25.395913 -25.464197
Tracing fn(par, ...) on exit
[1] -0.86931484  0.06178296 25.34935477
Tracing fn(par, ...) on exit
[1] -0.86981590  0.06202135 25.35832014
Tracing fn(par, ...) on exit
[1] -0.86979007  0.06197872 25.37671101
Tracing fn(par, ...) on exit
[1] -0.86964888  0.06194188 25.39017520
SHRINK            73 -25.376711 -25.464197
Tracing fn(par, ...) on exit
[1] -0.8694560  0.0618198 25.3644456
Tracing fn(par, ...) on exit
[1] -0.86970656  0.06193899 25.38072773
.
.
.
.
```

```
.
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.38816056
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.38538736
LO-REDUCTION     293 -25.388161 -25.503503
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.26827178
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.43424875
HI-REDUCTION     295 -25.426385 -25.503503
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.33268128
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.35284744
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.30812259
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.33933608
SHRINK           299 -25.308123 -25.503503
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.40525427
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.42434863
LO-REDUCTION     301 -25.339336 -25.503503
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.35761833
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.34595514
LO-REDUCTION     303 -25.357618 -25.503503
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.37782773
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.35706658
LO-REDUCTION     305 -25.377828 -25.503503
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.29122457
Tracing fn(par, ...) on exit
[1] -0.86959718  0.06185662 25.37769351
Polytope size measure not decreased in shrink
Exiting from Nelder Mead minimizer
    307 function evaluations used
$par
[1] -0.86959718  0.06185662

$value
[1] 25.5035

$counts
function gradient
     307       NA

$convergence
[1] 10

$message
NULL
```