Spatio-temporal point process models for interval-censored data

Robin Luca Markwitz

SPATIO-TEMPORAL POINT PROCESS MODELS FOR INTERVAL-CENSORED DATA

Robin Luca Markwitz

SPATIO-TEMPORAL POINT PROCESS MODELS FOR INTERVAL-CENSORED DATA

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr.ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Friday 26 September 2025 at 14.45 hours

by

Robin Luca Markwitz

born on the 28th of June, 1998
in Bad Soden am Taunus, Germany

This dissertation has been approved by:

Promotors
prof.dr. M.N.M. van Lieshout
prof.dr. R.J. Boucherie

Funding: NWO, grant number OCENW.KLEIN.068

Cover design: R.L. Markwitz & D.J. Rijkens
Printed by: Ipskamp Printing
Lay-out: LaTeX.
ISBN (print): 978-90-365-6811-1
ISBN (digital): 978-90-365-6812-8
URL: https://doi.org/10.3990/1.9789036568128

Graduation committee

Chair / secretary:	prof.dr.ir. B.R.H.M. Haverkort
Promotors:	prof.dr. M.N.M. van Lieshout Universiteit Twente, EEMCS Mathematics of Operations Research; Centrum Wiskunde & Informatica Stochastics Research Group
	prof.dr. R.J. Boucherie Universiteit Twente, EEMCS Mathematics of Operations Research
Committee Members:	prof.dr. W.M. Koolen Universiteit Twente, EEMCS Mathematics of Operations Research
	dr. C. Stegehuis Universiteit Twente, EEMCS Mathematics of Operations Research
	prof.dr. T. Müller Rijksuniversiteit Groningen Bernoulli Instituut
	prof.dr.ir. F.H. van der Meulen VU Amsterdam Faculteit der Bètawetenschappen
	prof.dr. C. Redenbach RPTU Kaiserslautern-Landau Fachbereich Mathematik

Dear reader,

This PhD thesis is the culmination of approximately five years' work. As is typical of such undertakings, this is not something that I could have achieved alone. If you are reading this, it is exceedingly likely that you have contributed in some way, however small, to this process. For that I would like to thank you.

I will proceed by first thanking my daily supervisor, Marie-Colette van Lieshout. You put your trust in me and helped me grow as a scholar. I hope I have met the expectations that you set for me. I would also like to thank my second supervisor Richard Boucherie, whose advice throughout this journey has been very valuable to me.

I would like to thank my PhD committee for reading my thesis and suggesting improvements. I would also like to thank the support staff at both CWI and UT for their assistance during these last 5 years, as well as Ipskamp Printing for printing this thesis.

To everyone I have directly worked with, shared an office with, discussed mathematics with, played ping-pong or volleyball with, talked with during coffee breaks and had fun outside of work with: thank you. I am forever grateful for your camaraderie, kindness, humour and passion. I was blessed with a truly massive cohort of fellow PhD and Master's students and I would like to thank each and every one of you.

To my friends and family, wherever in the world you may be: *thank you*, *danke* and *dankjewel* for supporting me during my PhD journey, whether directly or indirectly. You have helped make this possible.

Finally, to Dorien: thank you so much for being there for me. I love you.

Robin

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Problem description

In this thesis, we develop statistical methods for a specific subset of problems where intervals around points are observed, instead of the location of the individual points themselves. In applied statistics, and especially in the field of medical statistics, this kind of data is known as *interval-censored data*. The goal is to estimate the exact position of these points within their intervals. To achieve this, comprehensive models must be developed that take into account how these intervals are generated, the relationship between the intervals and the points, and how the exact position of the latent point within each interval changes as a result of the generation scheme and prior assumptions.

We turn to a motivating example to introduce the problem. Suppose a working person leaves his place of residence early in the morning, and returns late in the afternoon to discover that he has been burgled. Alternatively, a person may leave for a two-week vacation, only to find out that a property crime has been committed at her place of residence. The location of the crime is fixed, however the exact burglary time is uncertain. In the field of criminology, interval-censored data of this nature are known as *aoristic crime data* [120]. Similar data can be collected when investigating car arson fires. When a victim returns to their car and sees it alight or burned out, he or she is most likely unable to deduce at what time the incident occurred. Estimates of the most likely occurrence time of a crime along with its uncertainty are invaluable to law enforcement institutions such as police or other investigative officers, since they can update patrolling strategies or begin public awareness campaigns based on this information.

Our approach to solving this problem makes use of point process models. A *point process* scatters a number of points over some space. In this thesis, this refers to the union of intervals within which points fall, and in certain cases also the physical location of the point in space. We assume that a point from the point process model resides in each recorded interval. Parameters of the model, as well as the interval geometry, determine the relationship that the points have to one another. For example, if two time intervals overlap, are their corresponding points more likely to be close to each other, or far away? In later chapters, we also allow the spatial location of the point to have an effect on the temporal location. If a point is far away from another point in space, it might not be reasonable to assume that one affects the other. The full point process model will allow this behaviour to vary depending on the nature of the data.

Statistical inference allows one to estimate missing or unobservable properties of a model. In this thesis, we will discuss, develop and apply methods of statistical inference to point process models for interval-censored or aoristic data. When presented with data that is amenable to point process analysis, the underlying process that generated the data is typically unknown. Some inference methods

assume a certain prior structure, after which inference methods based on the likelihood (given previously occurring points) are used to recover missing or unknown parameters [88]. Another class of inference methods deals with the reconstruction of partially observable parts of a point process [77]. The central problem that we are trying to solve necessitates a combination of these approaches. This is what is referred to as *state estimation* in this thesis.

We will use a framework based on Bayesian inference to perform state estimation. This means that we first assume that a prior point process encapsulates the latent behaviour of points, and that points are therefore distributed with respect to the chosen model. We assume that intervals are distributed according to a stochastic process, which will be referred to as the *censoring mechanism* throughout this thesis. This interval model also contains a number of parameters that must be estimated separately from any parameters present in the prior point process model, and it provides the likelihood of the intervals given the location of the points. The full model is then given by the combination of the initial point process model and the censoring mechanism. Simulation and parameter estimation will be performed using a variety of Markov chain Monte Carlo methods, after which these models are applied to both simulated and real data sets.

1.2 A gentle introduction to point process theory

Point processes are stochastic (random) processes that consist of a set of points that are distributed over some region. They can be seen as a natural extension of a number of similar but related concepts in different branches of mathematics. When modelling the lifetime of particular components of machines, or the behaviour of gamblers at a casino, or the waiting time between two events, mathematicians discovered something quite interesting. Take the example of an engineer studying the time until failure of a mechanical component. After much investigation, this engineer discovers that the remaining lifetime of this component tends to follow one particular distribution independently of how much time had passed at the time of observation. This is known as memorylessness, a property captured formally by the exponential distribution. The probability of lifetime x is

$$p(x;\lambda) = \lambda e^{-\lambda x} \qquad x \geq 0,$$

where p denotes the *probability density function* (pdf) and λ is the *rate parameter*. The value of x is typically assumed to range over $[0, \infty)$, the *support* of the distribution. This refers to the values of x which map to strictly positive p. The value of λ determines the shape of the distribution, and is directly related to the mean, which is $1/\lambda$.

This example allows us to introduce some fundamental concepts in probability, as well as the basics of point process theory. Let X be a random variable.

Informally, this can be seen as a function that gives each possible outcome of an experiment a value. Formally, it is a measurable function $X : \Omega \to E$, where Ω is the sample space and E is known as the state space. In a broader sense, we can describe an experimental setup by using a probability space. This typically takes the form $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is as before, $\mathcal{F}$ is a collection of events known as the event space and $\mathbb{P}$ is the probability function. The probability function takes values over sets of $\mathcal{F}$. For example, imagine that the engineer from before wishes to know what the probability is of the mechanical component lasting between 2 and 3 years. We write $F = (2, 3] \in \mathcal{F}$. This probability can then be written as $\mathbb{P}(X \in F)$, where X is the random variable from earlier.

Having introduced some terminology, we return to the motivating example. Assume that the engineer replaces a mechanical component immediately upon failure. Due to the memoryless property of the exponential distribution, the time to failure of a newly installed mechanical component follows the exact same distribution as the time to failure of any components installed previously. The engineer now wishes to investigate the times at which components have been replaced in the past. Clearly, this is a stochastic process with independent increments. We make a few assumptions:

1. At time $x = 0$, no component has failed.

2. Over any interval $(a, b]$, $\lambda|b-a|$ components fail on average (which is evident based on exponential increments).

Let N denote this stochastic (counting) process. Assumption 2 implies that $\mathbb{E}N(a, b] = \lambda|b - a|$. Using the other assumptions, the engineer is able to derive that the law describing the number of failures within the interval $(a, b]$ is governed by a Poisson distribution. In other words,

$$\mathbb{P}(N(a, b] = n) = \frac{(\lambda|b - a|)^n}{n!} e^{-\lambda|b-a|}.$$

This counting process is known as a Poisson point process. The engineer therefore has a way of modelling the failures and subsequent replacements of this mechanical component.

A Poisson point process may have a rate function $\lambda(\cdot)$ dependent on where a measurement is taken in the event space, instead of a constant rate. This is known as a non-homogeneous Poisson process. In point process theory, $\lambda(\cdot)$ is often referred to as the *intensity function*. We now introduce the intensity measure, which is defined as

$$\Lambda(B) = \mathbb{E}N(B) = \int_B \lambda(x)\, dx$$

over a set B. In the homogeneous Poisson case, clearly $\Lambda(B) = \lambda|B|$, where $|B|$ is the area (or equivalent) of B. This intensity measure, and therefore the intensity function or field itself, can take all kinds of different functional forms. A number of different point process models more complicated than the general Poisson process can be constructed by letting the form of the intensity function vary, or by introducing additional stochasticity. Common types of point processes are the Cox process [30], the Markov point process [86], the self-exciting point process [61], among many others.

Point processes can be defined over many different spaces, such as the Cartesian plane, a network, or intervals on the real line. Some examples of possible sets of points in practice are arrival times of a bus, times of break-ins, and the locations of trees in a forest. Point processes that take values over a spatial region are often known as spatial point processes, whereas those which take values over time are known as temporal point processes [40]. Given a set of points and a spatial and/or temporal region in which they occurred, point process theory has investigated structural features of the distribution of points such as clustering or regularity (see [39] for a detailed overview), has developed work on inference [54, 102, 103], and has successfully applied statistical models to real-life data. In addition, concepts in theoretical statistics such as convergence, stationarity and ergodicity have been rigorously extended to point process theory [34]. Point process theory thus allows for accurate modelling of a large number of phenomena with an appropriate level of mathematical rigour.

1.3 Bayesian statistics and point processes

In statistics, there are two common schools of thought on how probability should be interpreted. These two philosophies are *Bayesian* and *frequentist* statistics. The frequentist sees probability as the long-term limit of the frequency of an event in an experiment, whereas the Bayesian interprets probability as a measure of the belief that they have of a certain event occurring based on prior knowledge. After seeing many thousands of mechanical components fail and noting the lifetime of each component, a frequentist engineer determines by a simple calculation that the failure rate within three years of a given component is 65%. A Bayesian engineer, on the other hand, based on his many years of prior knowledge, assumes that component lifetime follows some lifetime distribution with rate parameter λ. The value of the parameter λ is then determined or updated after taking the data into account, or may even be set in advance based on experience. The frequentist sees the parameters present in a given experiment as constants depending purely on the data, whereas the Bayesian sees them as variables influenced by prior beliefs. While this example may be somewhat simplistic in nature, it describes a clear distinction in philosophy.

The history of Bayesian point process models spans back more than 40 years. Many problems that point process models aim to solve deal with imperfect observations or require the estimation of intensities or summary statistics. Diggle extended the Rosenblatt-Parzen kernel density estimation method to point processes [37]. This method assumes an underlying Cox point process model and uses parametric methods to retrieve the smoothing value by way of an estimator. Though it is never explicitly stated, this is certainly a Bayesian method. Ogata & Katsura later wrote about performing spatial intensity estimation using an objective Bayesian regime [110]. Møller et al. introduced a variant of the Cox process known as the log-Gaussian Cox process, and described Bayesian methods for predicting the Gaussian surface that is not directly observed [101]. In all of these cases, we have an underlying distribution that is then manipulated in some way before estimation of the entire process is performed.

In Bayesian language, the underlying process is usually referred to as the *prior* distribution. The prior encodes our initial beliefs about the state, shape and overall nature of the imperfectly observed process. For example, if we assume an underlying point process structure, a Poisson point process prior X with rate λ is a possible choice. Point process distributions are typically described by way of a density function $p(\mathbf{x})$, where $\mathbf{x}$ is a realisation of the point process X. This density thus becomes the prior in this case. Let $\mathbf{u}$ be a data set. In Bayesian analysis, we want to know how likely this data configuration is given our prior assumptions. This can be calculated directly by evaluating $p(\mathbf{u} \mid \mathbf{x})$, which represents the probability of observing the configuration $\mathbf{u}$, given the configuration $\mathbf{x}$. This is often referred to as the *likelihood*, *forward model* or *forward term*, since it describes the probability of the data $\mathbf{u}$ given our prior $\mathbf{x}$. To determine the conditional or *posterior* distribution, which is the probability of observing $\mathbf{x}$ given the data $\mathbf{u}$, we use Bayes' theorem:

$$p(\mathbf{x} \mid \mathbf{u}) = \frac{p(\mathbf{u} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{u})}.$$

In this formula, $p(\mathbf{u})$ represents the probability of the data. While it is possible to calculate this in theory by using the law of total probability, in practice this is not feasible due to very large sample spaces typically being used. Therefore, the form $p(\mathbf{x} \mid \mathbf{u}) \propto p(\mathbf{u} \mid \mathbf{x})p(\mathbf{x})$ is usually used. In Chapter 2, a subsection is dedicated to the specific Bayesian framework used throughout this thesis.

1.4 Interval-censored data and previous approaches

Interval-censored data occur very commonly in the fields of medical statistics and applied criminology, the latter of which refers to this type of data as aoristic crime data. For example, when medical professionals monitor patients with

chronic illnesses, they might request a blood sample every six months to determine whether a patient's condition has worsened [144]. Traditionally, estimation for interval-censored data relied on methods from survival analysis. Kaplan & Meier discussed a number of actuarial methods in survival analysis and developed non-parametric methods for incomplete observations [76]. Cox then extended these methods to regression models [31], developing the proportional hazards approach, which became one of the most used statistical models of the 20th century, seeing application in a large number of fields. Huang developed estimation methods for the proportional hazards model when the data are interval-censored, and when the form of the underlying hazard function is unspecified [67, 68]. While these methods are certainly very powerful and can be applied in many contexts, they assume that the contribution of the survival function for a given observation is entirely dependent on the interval itself. The implicit assumption is that the intervals and the events within them are governed by the same process.

We wish to develop methods where the interval censoring mechanism is separate from the process governing event occurrences within these intervals. Occurrences might be constrained by their corresponding intervals, but behave according to an entirely different underlying distribution. Recent Bayesian approaches involve assigning priors to observations. Briz-Redón, analysing a data set very similar to the ones that will be considered in this thesis, used a Bayesian method which assigns uniform priors to observations within each interval [22]. Crema developed a Bayesian framework for the analysis of aoristic data in archaeology [32]. We also use a Bayesian approach, but assume that the prior takes the form of a point process. We therefore do not work on a per-interval basis, but assume that we observe realisations from an entire process which manifests itself within a set of intervals, which themselves are assumed to have been generated by a stochastic process.

Since this thesis views estimation methods for interval-censored data from a point process perspective, a literature review would not be complete without mentioning methods in point process theory that perform state estimation or missing data analysis. Diggle et al. [38] tackle the missing data problem by using a log-Gaussian Cox process [101] to infer the temporal distribution of calls to the NHS for gastroenterological complaints. Another approach by Tucker et al. uses the self-exciting process (also known as the Hawkes process) developed by Hawkes & Oates [61] as a prior process, augmenting this with the observed data [36]. As far as we are aware, no model has used point process methods to develop an inference scheme for the estimation of points within interval-censored data, while modelling interval-censored data as a stochastic process. This will be the main aim of this thesis.

1.5 Contributions and thesis layout

This thesis contains four chapters. The first three chapters will contain the mathematical theory from three research articles in probability and statistics. The final chapter, Chapter 5, will focus on estimation methods and applications to criminology and takes its content from all three previously mentioned research articles, as well as an applied criminology article.

In Chapter 2, we provide a brief introduction to the problem at hand from a mathematical perspective and introduce a temporal censoring mechanism based on homogeneous alternating renewal process theory. We also introduce the Bayesian framework that will be used throughout the thesis to build the full marked point process model. This complete model uses a Markov point process [86] prior, with the likelihood coming from the renewal process censoring mechanism [90]. We show that the conditional (or posterior) distribution exists by providing a measure-theoretic proof. We then derive the form of this posterior distribution, after which we provide a concrete statistical inference framework for this model. This chapter is based on:

- Lieshout, M. N. M. van, & Markwitz, R. L. (2023). State estimation for aoristic models. *Scandinavian Journal of Statistics, 50*(3), 1068–1089.

In Chapter 3, we outline the potential shortcomings of a purely homogeneous temporal censoring mechanism and offer an alternative. Specifically, the time at which an interval starts and the length of the interval cannot be assumed to be time-invariant in many applications. To rectify this, we introduce a censoring mechanism based on a two-state non-homogeneous semi-Markov process, which is effectively a non-homogeneous alternating renewal process [91]. We provide a detailed mathematical overview of this process and derive the same joint forward and backward recurrence distribution as in the homogeneous case, which shows its viability as a censoring mechanism. Since it is not possible to model this process directly due to the complexity of the non-homogeneous renewal function, we develop a scheme by which this renewal function can be modelled indirectly without loss of generality. We provide a number of auxiliary distributions and some details regarding state estimation methods. This chapter will take its content from the following paper:

- Lieshout, M. N. M. van, & Markwitz, R. L. (2025). A non-homogeneous semi-Markov model for interval censoring. *Journal of Applied Probability, 62*(2), 494-515.

In Chapter 4, we keep the likelihood structure from Chapter 3 the same, but assume that the underlying point processes also takes values in space. Since

applications of the theory developed in earlier chapters are likely to be within the field of criminology, we assume that points take values on a Euclidean graph, which models a typical street network. An overview of network properties and notation is provided. We then introduce a network area-interaction point process model, show that it is Markovian and otherwise satisfies the definition of an area-interaction process with the chosen interaction area. After verifying the validity of the model, we discuss perfect simulation methods, meaning a spatial birth-death process and coupling from the past algorithm. We derive the form of the final model, which is made up of a (spatial) network area-interaction point process prior and a semi-Markov inspired temporal censoring mechanism. Novel parameter estimation methods and uncertainty quantification based on an estimating equation approach is described and outlined in detail. The chapter's content is based on the following paper:

- Lieshout, M. N. M. van, & Markwitz, R. L. (n.d.). A network area-interaction point process model with temporal censoring applied to car arson fires. *Manuscript in preparation.*

Chapter 5 contains a detailed overview of criminological theory relevant to possible applications of our various models. Gaps in the criminological literature, especially as it relates to policing methods and the analysis of aoristic crime data, are identified and explored in detail. The models discussed in Chapters 2, 3 and 4 are re-introduced in a criminological context. We then test these models on two data sets - open source crime data of residential burglaries in Washington D.C., and a data set of car arson fires in Enschede. We discuss the addition of spatial covariates to the prior model, and implement a random forest method to single out relevant covariates for the car arson fire data set. Results, which include parameter values and visualisations of the estimated points, are then provided at the end of the chapter. This chapter's content is based on the three papers already mentioned, plus:

- Markwitz, R.L. (2024). A likelihood-based approach to developing effective proactive police methods. In *The UN Sustainable Development Goals and Provision of Security, Responses to Crime and Security Threats, and Fair Criminal Justice Systems*, pp. 285-304. University of Maribor Press.

CHAPTER 2

A renewal process censoring regime for aoristic crime data

The focus of this chapter is to develop a Bayesian inference framework for interval-censored data that is able to infer missing information and takes into account expert knowledge and interaction. An interval censoring mechanism, introduced in Chapter 1, constructs two states (observable and partially observable) within which events can occur. In the fully observed state, event locations are recorded perfectly. In the partially observed state, we record the entire interval within which the event occurred. This censoring mechanism takes the form of a stochastic process. In this chapter, we assume that the censoring mechanism can be modelled by an alternating renewal process. After observing a data set, we construct a partially observed marked point process via this alternating renewal process censoring mechanism. It is assumed that each observation occurs within its own alternating renewal process.

We use a Bayesian framework to perform inference on this marked point process. As is typical for Bayesian inference, a prior underlying form for the distribution of the exact locations of the points is assumed [102]. A posterior distribution form is generated using Markov chain Monte Carlo methods for a choice of prior forms after updating the prior with the interval data [50]. As Markov chain Monte Carlo methods simulate the behaviour of the posterior distribution provided the interval data, state estimation is performed with intervals to reconstruct the most likely distribution of occurrence times of partially observed events. The prior form encodes assumptions about the nature of the behaviour that one might want to include in one's model. For example, if one believes that events are more likely to occur near other events, the prior distribution might be set to a point process that exhibits clustering behaviour. If one however operates under the assumption that events are more likely to be spread out in time, a point process model that allows for regular behaviour may be chosen as the prior. In the following subsections, we introduce the mathematical concepts required to build this model.

2.1 Alternating renewal process theory

In terms of modelling, an *alternating renewal process* is often used when there is a strong belief that a phenomenon cycles between two different states. In this thesis, the first of these two states refers to the case when only partial observation is possible and an entire interval is recorded, which we call a Y-phase. At all other times, we assume that full information is available and the recorded interval is just the point in time. This is called a Z-phase in this thesis. In the case of the motivating example outlined in Chapter 1.1, the Y-phase would represent a time interval within which the occurrence time is not observed exactly, whereas the Z-phase represents perfect observation, e.g. by a CCTV camera or a bystander.

Let $C_1, C_2, \ldots$ be a sequence of random vectors such that $C_i = (Y_i, Z_i)$ represents an alternating renewal process [6, 126]. Let the kth realisation of Y_k be known

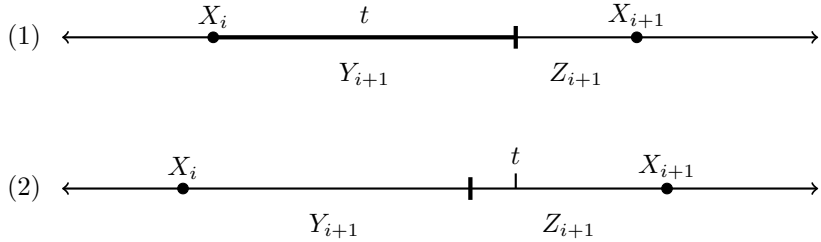


Figure 2.1 View of two alternating renewal processes showing a point falling in Y- and Z-phases. The short bold vertical line denotes the boundary between the two phases. In this application, when t falls in a Y-phase (1), an interval, denoted by the bold horizontal line, is recorded. When t falls in a Z-phase (2), the exact time is recorded.

as the kth Y-phase, and similarly for the kth Z-phase in the Z case. The Y_i are independent and identically distributed, as are the Z_i. Introduce $T_i = Y_i + Z_i$ as the length of phase i, and let X_i be the time of the ith renewal of the alternating renewal process, noting that no renewal occurs between Y_i and $Y_i + Z_i$. By convention, we set $X_0 = 0$. The distributions of the Y_i are absolutely continuous with PDF $f_Y(l)$ on $\mathbb{R}^+$. Assume that $0 < \mathbb{E}[T_i] < \infty$ for all i, and that a moment measure exists. By the strong law of large numbers, it can be shown that

$$N(t) = \sup\{n \in \mathbb{N}_0 \,:\, X_n \leq t\},$$

the number of renewal phases that have elapsed as of time t, is well-defined. By $\mathbb{N}_0$ we mean $\mathbb{N} \cup \{0\}$.

We now define the renewal function, which intuitively can be viewed as the expected number of cycles to have occurred by time t, as

$$M(t) = \mathbb{E}N(t) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \leq t), \quad t \geq 0,$$

which can be shown to be finite and absolutely continuous with respect to the Lebesgue measure [6, 126]. To see a visualisation of an alternating renewal process, see Figure 2.1.

This idea allows us to construct a model for the censoring mechanism in time. Depending on t, either partial or full observation is possible. Though the framework has been outlined, it is yet unclear how occurrence times and the censoring mechanism are formally linked. In the next chapter, we use distributions of age and excess to work towards a full model.

2.2 Age and excess distribution

Aoristic data generated by the censoring mechanism can be expressed formally in terms of the age and excess (also referred to as residual lifetime, or forward/backward recurrence times) with respect to the Y-process. For an alternating renewal process, with respect to the Y-process, the *age* is defined as:

$$\begin{aligned} A(t) &= (t - X_{N(t)})\mathbf{1}\{X_{N(t)} + Y_{N(t)+1} > t\} \\ &= \begin{cases} t - X_{N(t)}, & X_{N(t)} + Y_{N(t)+1} > t \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where $t - X_{N(t)}$ is the distance between the current value of t and the last renewal interval, and the indicator function $\mathbf{1}\{X_{N(t)} + Y_{N(t)+1} > t\}$ denotes whether or not t is contained in a Y-phase.

The *excess* is defined as:

$$\begin{aligned} B(t) &= (X_{N(t)+1} - Z_{N(t)+1} - t)\,\mathbf{1}\{X_{N(t)} + Y_{N(t)+1} > t\} \\ &= \begin{cases} X_{N(t)} + Y_{N(t)+1} - t, & X_{N(t)} + Y_{N(t)+1} > t \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where $X_{N(t)} + Y_{N(t)+1} - t$ is the distance between the end of the Y-phase and the current point t. Note that if $t > X_{N(t)} + Y_{N(t)+1}$, $B(t)$ is set to zero, since when t is in a Z-phase, a new renewal has not occurred yet. The indicator $\mathbf{1}\{X_{N(t)} + Y_{N(t)+1} > t\}$ restricts the support of $B(t)$.

Let the pair $(a, l) \in \mathbb{R} \times \mathbb{R}^+$ correspond to the closed interval $[a, a+l]$. In this way, an interval on the real line is parametrised by its left-most point and length. A recorded interval $[a, a+l]$ in which a given latent point $t \geq 0$ falls is then written as $t + [-A(t), B(t)]$, seeing that $a = t - A(t)$ and $a + l = t + B(t)$. The closed interval $[-A(t), B(t)]$ is called the mark attached to t and corresponds to the two-dimensional parameter vector $I(t) = (-A(t), A(t) + B(t)) \in \mathbb{R} \times \mathbb{R}^+$. Note that the mark always contains the origin 0. For a visualisation, see Figure 2.2.

We would like to derive the form and nature of the mark distribution. To achieve this, the joint distribution of age and excess with respect to the Y-process is first considered.

Proposition 2.1. *Let N be an alternating renewal process. Assume that T_1 is absolutely continuous with respect to Lebesgue measure and that $0 < \mathbb{E}[T_1] < \infty$. Then, for $t \geq 0$, the joint distribution of $(A(t), B(t))$ has an atom at $(0, 0)$ of size*

$$c(t) = F_Y(t) - \int_0^t [1 - F_Y(t - s)]\, dM(s)$$

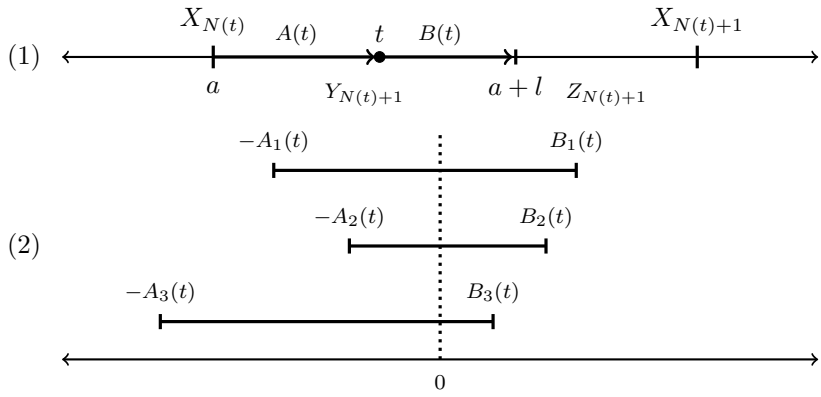


Figure 2.2 Diagram (1) shows an alternating renewal process where t falls in a Y-phase. The age $A(t)$ and excess $B(t)$ distributions describe the elapsed and remaining time, a and $a+l$ the two ends of the interval. Diagram (2) shows the parametrisation of three example intervals, with corresponding subscripts, after shifting to the origin.

and, for $0 \leq u \leq t, v \geq 0$,

$$\begin{aligned}\mathbb{P}(A(t) \leq u; B(t) \leq v) = c(t) &+ [F_Y(t+v) - F_Y(t)]\,\mathbf{1}\{u = t\} \\ &+ \int_{t-u}^{t} [F_Y(t+v-s) - F_Y(t-s)]\,dM(s).\end{aligned}$$

Here F_Y *denotes the cumulative distribution function of* Y_1 *and* m *the renewal function.*

Proof. Write F_n for the cumulative distribution function of X_n, $n \in \mathbb{N}$. By partitioning over the number of renewals up to time t and upon noting that $N(t) = n$ if and only if $X_n \leq t$ and $X_n + Y_{n+1} + Z_{n+1} > t$, one obtains that

$$\begin{aligned}\mathbb{P}(A(t) \leq u; B(t) \leq v) &= c(t) + \mathbb{P}(t - X_{N(t)} \leq u;\, t < X_{N(t)} + Y_{N(t)+1} \leq t + v) \\ &= c(t) + \sum_{n=0}^{\infty} \mathbb{P}(t - u \leq X_{N(t)};\, t < X_{N(t)} + Y_{N(t)+1} \leq t + v;\, N(t) = n) \\ &= c(t) + \mathbb{P}(t - u \leq X_{N(t)};\, t < X_{N(t)} + Y_{N(t)+1} \leq t + v;\, N(t) = 0) \\ &\quad + \sum_{n=1}^{\infty} \int_{t-u}^{t} \mathbb{P}(t - s < Y_{n+1} \leq t + v - s)\,dF_n(s).\end{aligned}$$

The claim follows by an application of Fubini's theorem for the last term, the observation that

$$\mathbb{P}(t - u \leq X_{N(t)};\, t < X_{N(t)} + Y_{N(t)+1} \leq t + v;\, N(t) = 0) = \mathbb{P}(t < Y_1 \leq t + v)$$

if $u = t$ and zero otherwise, and because

$$\begin{aligned} c(t) &= 1 - \mathbb{P}(X_{N(t)} + Y_{N(t)+1} > t) \\ &= 1 - \mathbb{P}(Y_1 > t) - \sum_{n=1}^{\infty} \int_0^t \mathbb{P}(Y_{n+1} > t - s)\, dF_n(s) \\ &= F_Y(t) - \int_0^t [1 - F_Y(t-s)]\, dM(s). \end{aligned}$$

□

The long-run behaviour as time goes to infinity can be obtained by appealing to the key renewal theorem. Recall that this theorem applies to functions $h : \mathbb{R}^+ \to \mathbb{R}$ such that $h(t) \geq 0$ for all $t \geq 0$, h is monotonically non-increasing and integrable. For such functions, it states that

$$\lim_{t \to \infty} \int_0^t h(t-s)\, dM(s) = \frac{1}{\mathbb{E}[T_1]} \int_0^{\infty} h(s)\, ds$$

[126, Theorem 3.4.2].

Specialising to the parameter vector $I(t)$, the following theorem holds.

Theorem 2.2. *Let N be an alternating renewal process. Assume that Y_1 and T_1 are absolutely continuous with respect to Lebesgue measure with probability density functions f_Y and f, and that $0 < \mathbb{E}[T_1] < \infty$. As $t \to \infty$, the joint probability distribution of $(-A(t), A(t) + B(t))$ tends in distribution to ν, the mixture of an atom at $(0, 0)$ and an absolutely continuous component that has probability density function $f_Y(l)/\mathbb{E}[Y_1]$ on $\{(a, l) \in \mathbb{R} \times \mathbb{R}^+ : a \leq 0 \leq a + l\}$. The mixture weights are, respectively, $\mathbb{E}[Z_1]/\mathbb{E}[T_1]$ and $\mathbb{E}[Y_1]/\mathbb{E}[T_1]$.*

Proof. First, let us consider the limit behaviour of the joint cumulative distribution function of $A(t)$ and $B(t)$ as $t \to \infty$. With the notation of Proposition 2.1, by Theorem 3.4.4 in [126] or the key renewal theorem applied to $1 - F_Y$, $c(t)$ converges to $\mathbb{E}[Z_1]/\mathbb{E}[T_1]$. Also, for $t > u$, the second term in the joint cumulative distribution function of $A(t)$ and $B(t)$ is zero. For the last term, note that for $v \geq 0$ the function $h_v : \mathbb{R}^+ \to \mathbb{R}$ defined by $h_v(s) = 1 - F_Y(v + s)$ is non-negative, monotonically non-increasing and integrable. Hence the key renewal theorem implies that

$$\lim_{t \to \infty} \int_0^t [1 - F_Y(t + v - s)]\, dM(s) = \frac{1}{\mathbb{E}[T_1]} \int_0^{\infty} [1 - F_Y(v + s)]\, ds.$$

Analogously, for fixed $u \geq 0$, $t - u \to \infty$ if and only if $t \to \infty$. Writing $s = t - u$,

$$\lim_{s \to \infty} \int_0^s [1 - F_Y(s + u + v - r)]\, dM(r) = \frac{1}{\mathbb{E}[T_1]} \int_0^{\infty} [1 - F_Y(u + v + r)]\, dr$$

$$= \lim_{t\to\infty} \int_0^{t-u} [1 - F_Y(t+v-r)]\, dM(r).$$

We conclude that $Q(u,v) = \lim_{t\to\infty} \mathbb{P}(A(t) \le u; B(t) \le v)$ exists and equals

$$Q(u,v) = \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]} + \frac{1}{\mathbb{E}[T_1]} \int_0^u [F_Y(v+s) - F_Y(s)]\, ds.$$

We continue by taking derivatives. We obtain

$$\begin{aligned} Q_u(u,v) &= \frac{d}{du}\left[\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]} + \frac{1}{\mathbb{E}[T_1]} \int_0^u (F_Y(y+v) - F_Y(y))\ dy\right] \\ &= \frac{1}{\mathbb{E}[T_1]}[F_Y(u+v) - F_Y(u)] = \frac{1}{\mathbb{E}[T_1]}\left[\int_0^{u+v} f_Y(y)\, dy - \int_0^u f_Y(y)\, dy\right] \\ &= \frac{1}{\mathbb{E}[T_1]} \int_u^{u+v} f_Y(y)\, dy = \frac{1}{\mathbb{E}[T_1]} \int_0^v f_Y(u+y)\, dy. \\ Q_{uv}(u,v) &= \frac{d}{dv}\left[\frac{1}{\mathbb{E}[T_1]} \int_0^v f_Y(u+y)\, dy\right] = \frac{f_Y(u+v)}{\mathbb{E}[T_1]} = \frac{f_Y(l)}{\mathbb{E}[T_1]}, \end{aligned}$$

using the fact that the length $l = u + v$.

Recall that the probability of being in a Y-phase is $f_Y(l)$. The length l has been taken into account by considering the joint age and excess distributions, of which $u + v$ denotes the length of an arbitrary interval within the alternating renewal process. To ensure that $Q_{uv}(u,v)$ is a probability density function, the density must be integrated over arbitrary starting points a, corresponding to all intervals $[a, a+l]$. So b must be found so that $b\int_a^{a+d} f_Y(l) = 1$ for d such that l falls in $[0,d]$. Therefore

$$\int_a^{a+d} f_Y(l)\, dl = F_Y(a+d) - F_Y(a) = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}, \tag{2.1}$$

as this is the total amount contributed by points in Y-phases. Therefore $b = \frac{\mathbb{E}[T_1]}{\mathbb{E}[Y_1]}$. Hence

$$b\, Q_{xz}(x,z) = b\, \frac{f_Y(l)}{\mathbb{E}[T_1]} = \frac{\mathbb{E}[T_1]}{\mathbb{E}[Y_1]}\, \frac{f_Y(l)}{\mathbb{E}[T_1]} = \frac{f_Y(l)}{\mathbb{E}[Y_1]}.$$

Since the Y_i are i.i.d., the density is $\frac{f_Y(l)}{\mathbb{E}[Y_1]}$ when $l > 0$, which is exactly ν.

For an alternative but simple explanation, let $h : (\mathbb{R}^+)^2 \to \mathbb{R}^- \times \mathbb{R}^+$ be the bijection defined by

$$h(u,v) = (-u, u+v).$$

Since h is differentiable, the absolutely continuous part has probability density function $f_Y(h^{-1}(a,l))\, |\det J_{h^{-1}}(a,l)|/\mathbb{E}[Y_1] = f_Y(-a+a+l)/\mathbb{E}[Y_1]$, where $J_{h^{-1}}$ is the Jacobian of h^{-1}. □

Corollary 2.3. *For the left-most point of an interval a and an interval length l, the joint distribution f is*

$$f(a,l) = \frac{f_Y(l)}{\mathbb{E}[Y_1]}\,\mathbf{1}\{a < 0 < a + l\}. \tag{2.2}$$

Definition 2.4. *The hazard rate λ of the random variable Y on $\mathbb{R}^+$ is*

$$\lambda(l) = \begin{cases} \frac{f_Y(l)}{1-F_Y(l)} & \text{if } F_Y(l) < 1, \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

for $l \in \mathbb{R}^+$.

2.3 Model formulation

We are now ready to formulate a model. Let $\mathcal{X}$ be an open subset of the positive half-line $\mathbb{R}^+$. We define the point process X on $\mathcal{X}$ with its corresponding state space $\mathcal{N}_\mathcal{X}$ consisting of finite sets $\{t_1, \ldots, t_n\} \subset \mathcal{X}$ where $n \in \mathbb{N}_0$. We equip this with the Borel σ-algebra of the weak topology [34]. We will assume that the distribution of X is specified in terms of a probability density function p_X with respect to the distribution of a unit rate Poisson process on $\mathcal{X}$ [86].

Recall that we want to model two occurrences - partial (Y-phase) and total (Z-phase) observation. Using the alternating renewal process in equilibrium to simulate the behaviour of the victim, the points of X are independently marked according to the mixture distribution of Theorem 2.2. In this way, the complete model W is obtained. Its realisations are sets $\{(t_1, I_1), \ldots, (t_n, I_n)\} \subset \mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$, where $I_j = (a_j, l_j)$ is the jth mark parametrisation. The pair (t_j, I_j) thus defines an interval $[t_j + a_j, t_j + a_j + l_j]$. The ensemble of all realisations is denoted by $\mathcal{N}_{\mathcal{X}\times(\mathbb{R}\times\mathbb{R}^+)}$ and equipped with the Borel σ-algebra of the weak topology. Note that W has probability density function p_X with respect to the distribution of a Poisson process on $\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$ with intensity measure $\ell \times \nu$ where ℓ is Lebesgue measure.

Due to the censoring, one does not observe the complete model W but rather the set

$$U = \bigcup_{(t,I)\in W} ((t,0) + I) \tag{2.4}$$

of interval parametrisations. When $I = (0,0)$, we have a full observation of the burglary time. For any other value of I, only the corresponding interval is observed. Our aim is to reconstruct X or W from U. In order to do so, the posterior distribution of X or W given U is needed. This will be the topic of the next chapter.

2.4 Model viability

As X and W cannot be directly observed, state estimation is used to estimate X and W from the intervals U, remembering that X is a point process, and W is a marked point process made up of the points of X together with their marks, which were obtained through the process described in Chapter 2.1. To perform state estimation, we employ a Bayesian framework. As is typical for Bayesian inference for point processes, a prior underlying form for the distribution of the exact locations of the points is assumed. A posterior distribution form is deduced after updating the prior with the interval data.

In a Bayesian framework, the posterior distribution updates prior forms in the light of data gathered [50]. Heuristically,

$$p_{X|U}(\mathbf{x} \,|\, \mathbf{u}) \propto p_{U|X}(\mathbf{u} \,|\, \mathbf{x})\; p_X(\mathbf{x}) \tag{2.5}$$

through the use of Bayes' theorem. The term $p_{U|X}(\mathbf{u} \,|\, \mathbf{x})$ describes the likelihood that the points of $\mathbf{x}$ generate the intervals in $\mathbf{u}$. In the literature, this term is referred to as a forward term, forward density or forward model [15, 85]. The term $p_X(\mathbf{x})$ captures prior beliefs about the geometry of $\mathbf{x}$. In our context, since the forward model is a mixture of discrete and absolutely continuous components, some care is required in handling (2.5).

2.4.1 Measure theoretic foundations

We now provide a suggested form for the posterior distribution and show that this form is measure-theoretically viable. The posterior form must lean on a mechanism by which points are assigned to intervals. The following proposition provides a detailed description of this mechanism.

Proposition 2.5. *Let W be a point process on the open set $\mathcal{X} \subset \mathbb{R}$ with probability density function p_X with respect to the distribution of a unit rate Poisson process on $\mathcal{X}$ marked independently with mark distribution ν defined in Theorem 2.2. Write X for the ground process of locations in $\mathcal{X}$ and consider the forward model (2.4). Let $\mathbf{u}$ be a realisation of U that consists of an atomic part $\{(a_1, 0), \ldots, (a_m, 0)\}$, $m \in \mathbb{N}_0$, and a non-atomic part $\{(a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$, $n \geq m$. Then the posterior distribution of X given $U = \mathbf{u}$ satisfies, for A in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathcal{X}}$,*

$$\mathbb{P}(X \in A \mid U = \mathbf{u}) = c(\mathbf{u}) \int_{\mathcal{X}^{n-m}} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \\ \mathbf{1}_A(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})$$

$$\times \left(\sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_i \{D_i\} = \{1,\ldots,n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \right) \prod_{i=1}^{n-m} dx_i$$

provided that $c(\mathbf{u})^{-1}$*, defined by*

$$\int_{\mathcal{X}^{n-m}} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \left(\sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_i \{D_i\} = \{1,\ldots,n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \right) \prod_{i=1}^{n-m} dx_i,$$

exists in $(0, \infty)$*.*

Proof. We must show that for each A in the Borel σ-algebra of $\mathcal{N}_{\mathcal{X}}$ with respect to the weak topology and each F in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathbb{R}\times\mathbb{R}^+}$ the following identity holds:

$$\mathbb{E}\left[\mathbf{1}_F(U)\, \mathbb{P}(X \in A \mid U)\right] = \mathbb{E}\left[\mathbf{1}_F(U)\, \mathbf{1}_A(X)\right]. \tag{2.6}$$

Use the definition of Janossy densities with respect to Lebesgue measure on $\mathcal{X}^n, n \geq 0$ [34],

$$\int_{\mathcal{N}_{\mathcal{X}}} 1_A(\mathbf{x})\, d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^n} 1_A(\mathbf{x}) j_n(x_1, \ldots, x_n) \prod_{i=1}^{n} dx_i.$$

For a valid Janossy representation, the term $j_n(x_1, \ldots, x_n)dx_1 \ldots dx_n$ refers to the probability that n distinct points x_i of X each lie in the infinitesimal region $(x_i, x_i + dx_i)$ provided absolute continuity with respect to Lebesgue measure [94]. Moreover,

$$j_n(x_1, \ldots, x_n)dx_1 \ldots dx_n = n!\rho_n \pi_n(x_1, \ldots, x_n) \tag{2.7}$$

where ρ_n is the probability of there being n points in $\mathcal{X}$, and π_n determines the finite-dimensional distribution of the individual points of X, provided there are n of them [34]. Using Equation 2.7,

$$\int_{\mathcal{N}_{\mathcal{X}}} 1_A(\mathbf{x})\, d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^n} 1_A(\mathbf{x}) j_n(x_1, \ldots, x_n) \prod_{i=1}^{n} dx_i$$

$$= \sum_{n=0}^{\infty} \int_{\mathcal{X}^n} 1_A(\mathbf{x}) \rho_n \pi_n(x_1, \ldots, x_n) \prod_{i=1}^{n} dx_i$$
$$= \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{X}|}}{n!} \int_{\mathcal{X}^n} 1_A(\mathbf{x}) p_X(x_1, \ldots, x_n) \prod_{i=1}^{n} dx_i,$$

where p_X is a prior distribution describing the location of the points in X. As densities are calculated with respect to a Poisson process with intensity 1, the term $e^{-|\mathcal{X}|}$ must also be present.

In order to consider the u-term, assume that m of the n points are perfectly observed in the alternating renewal process paradigm. Therefore m points are atomic, and $n - m$ points admit a density with respect to Lebesgue measure, meaning that a Janossy density can only be defined over these points.

Of the n total points, each individual point has a known probability of being either an atom or admitting a density. If a point $x \in X$ admits a density, it occurred in a Y-phase. Similarly, x occurring in a Z-phase leads to an atomic point. Therefore $\mathbb{P}(t \text{ in } Y\text{-phase}) = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}$, and likewise $\mathbb{P}(t \text{ in } Z\text{-phase}) = \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}$.

The conditional Janossy densities of the $n - m$ remaining points of U can be written as a product of the distribution $\rho_{n-m}(x_{m+1}, \ldots, x_n)$, which determines the particular points in the point process, and the distribution $\pi_{n-m}(\cdot \mid x_{m+1}, \ldots, x_n)$, which lays out the finite-dimensional distribution of each individual point given the existence of $n - m$ points [34]. As U is created by imposing marks on X through an alternating renewal process, the form of π_{n-m} can be exactly deduced. Using the Janossy density definition (see [34]),

$$j_{n-m}(u_{m+1}, \ldots, u_n \mid x_{m+1}, \ldots, x_n)\, du_{m+1} \ldots du_n$$
$$= (n-m)! \rho_{n-m}(x_{m+1}, \ldots, x_n)$$
$$\pi^{\text{sym}}_{n-m}(u_{m+1}, \ldots, u_n \mid x_{m+1}, \ldots, x_n).$$

As point processes are sets of points, the non-symmetrical forms of π must be used to consider the distribution of all permutations of atoms and density-admitting points. A permutation sum selects arbitrary subsets C_0 of size m over $\{1, \ldots, n\}$, including the empty set, in the non-symmetric paradigm. Letting C_0 be a subset of $\{1, \ldots, n\}$ relating to the arbitrary selection of $m = |C_0|$ atoms from U,

$$j_{n-m}(u_{m+1}, \ldots, u_n \mid x_{m+1}, \ldots, x_n)\, du_{m+1} \ldots du_n = \rho_{n-|C_0|}(x_{|C_0|+1}, \ldots, x_n)$$
$$\sum_{\text{perm } \{u_1, \ldots, u_{n-|C_0|}\}} \pi_{n-|C_0|}(\{u_1, \ldots, u_{n-|C_0|}\} \mid x_{\{1,\ldots,n\} \setminus C_0}) \prod_{j=1}^{n-|C_0|} du_j$$

over $(n - |C_0|)!$ permutations. Splitting the two classes of points,

$$\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x}) = \sum_{C_0\subseteq\{1,\ldots,n\}} [\mathbb{P}(x \text{ in } Z\text{-phase})]^{|C_0|} \frac{1}{(n-|C_0|)!}$$
$$\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} [\mathbb{P}(x \text{ in } Y\text{-phase})]^{n-|C_0|} 1_F(\{u_1,\ldots,u_{n-|C_0|}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$j_{n-|C_0|}(\{u_1,\ldots,u_{n-|C_0|}\}\,|\,x_{\{1,\ldots,n\}\setminus C_0}) \prod_{j=|C_0|+1}^{n} du_j$$

$$= \sum_{C_0\subseteq\{1,\ldots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|}$$
$$\frac{1}{(n-|C_0|)!} \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\ldots,u_{n-|C_0|}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$j_{n-|C_0|}(\{u_1,\ldots,u_{n-|C_0|}\}\,|\,x_{\{1,\ldots,n\}\setminus C_0}) \prod_{j=|C_0|+1}^{n} du_j.$$

In order to implement the permutation sum over the non-symmetric finite-dimensional distributions, sum over all sets $C_1,\ldots,C_{n-|C_0|}$ excluding the empty set such that for any j, $\cup_j C_j = \{1,\ldots,n\}\setminus C_0$. Thus

$$\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x}) = \sum_{C_0\subseteq\{1,\ldots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \frac{1}{(n-|C_0|)!}$$
$$\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\ldots,u_{n-|C_0|}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$\sum_{\substack{C_1,\ldots,C_{n-|C_0|}\\ \cup_j C_j\subseteq\{1,\ldots,n\}\setminus C_0}} \pi_{n-|C_0|}(\{u_1,\ldots,u_{n-|C_0|}\}\,|\,x_{\{1,\ldots,n\}\setminus C_0}) \prod_{j=|C_0|+1}^{n} du_j$$

after accounting for the Janossy terms.

The finite-dimensional densities $\pi_{n-|C_0|}(\cdot\,|\,\cdot)$ are admitted for all permutations of points not in the set of atoms C_0. Summing over all combinations of sets not containing points of C_0, for each point, the density function

$$q_x(a,l) = \frac{f_Y(l)}{\mathbb{E}[Y_1]}\,\mathbf{1}\{a < x < a+l\}$$

describes the density admitted by a point $u = (a,l)$ of U, given a corresponding

x (see Equation 2.2 in Chapter 2.3 for derivation). Overall,

$$\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x}) = \sum_{C_0\subseteq\{1,\dots,n\}} \frac{1}{(n-|C_0|)!} \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\dots,u_{n-|C_0|}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\})) \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \sum_{\substack{C_1,\dots,C_{n-|C_0|}\\ \cup_j C_j\subseteq\{1,\dots,n\}\setminus C_0}} \prod_{j=|C_0|+1}^{n} q_{x_{C_j}}(u_j)\, du_j \quad (2.8)$$

where x_{C_j} refers to the selection of x corresponding to C_j. After rearranging and collecting terms,

$$\begin{aligned}\mathbb{E}[1_F(U)1_A(X)] = \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{X}|}}{n!} \sum_{C_0\subseteq\{1,\dots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \\ \int_{\mathcal{X}^n} 1_A(\mathbf{x})\, p_X(x_1,\dots,x_n) \sum_{\substack{C_1,\dots,C_{n-|C_0|}\\ \cup_j C_j\subseteq\{1,\dots,n\}\setminus C_0}} \frac{1}{(n-|C_0|)!} \\ \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\dots,u_{n-|C_0|}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\})) \\ \prod_{j=|C_0|+1}^{n} q_{x_{C_j}}(u_j)\, du_j \prod_{i=1}^{n} dx_i. \end{aligned} \quad (2.9)$$

For the expression $\mathbb{E}[1_F(U)\mathbb{P}(X\in A\,|\,U)]$, expanding as before,

$$\begin{aligned}\mathbb{E}[1_F(U)\mathbb{P}(X\in A\,|\,U)] &= \int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X\in A\,|\,U=\mathbf{u})\, d\mathbb{P}(\mathbf{x},\mathbf{u}) \\ &= \int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X\in A\,|\,U=\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x}) d\mathbb{P}(\mathbf{x}).\end{aligned}$$

Using the same scheme as for Equation 2.8 and 2.9,

$$\begin{aligned}\int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X\in A\,|\,U=\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x}) d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{X}|}}{n!} \\ \sum_{C_0\subseteq\{1,\dots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \int_{\mathcal{X}^n} \frac{1}{(n-|C_0|)!} \\ \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\dots,u_{n-|C_0|}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))\end{aligned}$$

$$\sum_{\substack{\emptyset = C_1,\ldots,C_{n-|C_0|} \\ \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0}} \pi_{n-|C_0|}(\{u_1,\ldots,u_{n-|C_0|}\} \,|\, x_{\{1,\ldots,n\}\setminus C_0}) \\ \mathbb{P}(X \in A \,|\, U = \{u_1,\ldots,u_{n-|C_0|}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\}))\, d\mathbb{P}(\mathbf{x}) \prod_{j=|C_0|+1}^{n} du_j \tag{2.10}$$

for an arbitrary subset C_0. A potential form of $p_{X|U}(\mathbf{x} \,|\, \mathbf{u})$ has been detailed in Proposition 2.5. Here, $p_{X|U}(\mathbf{x} \,|\, \mathbf{u})$ is proportional to

$$p_X(\{a_1,\ldots,a_{|C_0|},x_{|C_0|+1},\ldots,x_n\}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(x_{D_i}) \\ = c(\mathbf{u})\, p_X(\{a_1,\ldots,x_n\}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(x_{D_i})$$

for a given normalisation constant $c(\mathbf{u})$ depending on the selection of a realisation $\mathbf{u}$. In order to confirm measure-theoretic viability, the provided form must satisfy Equation 2.6. Writing $\mathbb{P}(X \in A \,|\, U = \mathbf{u})$ as an integral of $p_{X|U}(\mathbf{x} \,|\, \mathbf{u})$ for putative conditional probability,

$$\mathbb{P}(X \in A \,|\, U = \mathbf{u}) = c(\mathbf{u}) \int_{\mathcal{X}^{n-m}} p_X(\{a_1,\ldots,a_m,x_{m+1},\ldots,x_n\}) \\ 1_A(\{a_1,\ldots,a_m,x_{m+1},\ldots,x_n\}) \\ \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(x_{D_i})\, dx_i$$

where n is as before, given the existence of m atoms. The indicator $1_A(\{a_1,\ldots,a_m,x_{m+1},\ldots,x_n\})$ ensures that values of a given realisation of X remain in A. Note that the term

$$\sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(x_{D_i})$$

describes the selection of certain x_i, given a permutation of the $n-m$ non-atomic points $\{u_1,\ldots,u_{n-|C_0|}\}$. The normalisation constant $c(\mathbf{u})$, after the application of Bayes' law and taking the integral of both sides of the posterior expression, is

$$\frac{1}{\int_{\mathcal{X}^{n-m}} p_X(\mathbf{x} \cup \{a_1,\ldots,a_m\}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(x_{D_i})\, dx_i}$$

over the $n-m$ non-atomic points of a realisation $\mathbf{u}$. The posed posterior distribution form provides the framework for the selection of points of X given a realisation of U. Take C_0, where $m = |C_0|$, to be the same arbitrary subset selection as before. Adding new integration variables,

$$\int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X \in A \,|\, U = \mathbf{u})\, d\mathbb{P}(\mathbf{u} \,|\, \mathbf{x}) d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{X}|}}{n!} \sum_{C_0 \subseteq \{1,\ldots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \int_{\mathcal{X}^n} p_X(\{x_1,\ldots,x_n\}) \sum_{C_1,\ldots,C_{n-|C_0|} \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0} \frac{1}{(n-|C_0|)!} \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\ldots,u_{n-|C_0|}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\})) \prod_{j=|C_0|+1}^{n} q_{x_{C_j}}(u_j)\, c(\{u_1,\ldots,u_{n-|C_0|}\} \cup (\mathbf{x}_{C_0} \times \{0\})) \int_{\mathcal{X}^{n-|C_0|}} p_X(\mathbf{y} \cup \mathbf{x}_{C_0})\, 1_A(\mathbf{y} \cup \mathbf{x}_{C_0}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(y_{D_i}) \prod_{i=|C_0|+1}^{n} dy_i \prod_{j=|C_0|+1}^{n} du_j \prod_{k=1}^{n} dx_k$$

after substitution. Substituting for the aforementioned normalisation constant $c(\{u_1,\ldots,u_{n-|C_0|}\} \cup (\mathbf{x}_{C_0}))$, we obtain

$$\int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X \in A \,|\, U = \mathbf{u})\, d\mathbb{P}(\mathbf{u} \,|\, \mathbf{x}) d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{X}|}}{n!} \sum_{C_0 \subseteq \{1,\ldots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \int_{\mathcal{X}^n} p_X(\{x_1,\ldots,x_n\}) \sum_{\substack{C_1,\ldots,C_{n-|C_0|} \\ \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0}} \frac{1}{(n-|C_0|)!} \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\ldots,u_{n-|C_0|}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\})) \prod_{j=|C_0|+1}^{n} q_{x_{C_j}}(u_j) \frac{1}{\int_{\mathcal{X}^{n-|C_0|}} p_X(\mathbf{z} \cup (\mathbf{x}_{C_0} \times \{0\})) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(z_{D_i})\, dz_i} \int_{\mathcal{X}^{n-|C_0|}} p_X(\mathbf{y} \cup \mathbf{x}_{C_0})\, 1_A(\mathbf{y} \cup \mathbf{x}_{C_0}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}}$$

$$\prod_{i=m+1}^{n} 1_{[a_i,a_i+l_i]}(y_{D_i}) \prod_{i=|C_0|+1}^{n} dy_i\, du_j \prod_{k=1}^{n} dx_k.$$

Note that in order to cancel terms, the order of integration must be changed. Since the indicator function $1_A(\mathbf{y} \cup \mathbf{x}_{C_0})$ lies within the scope of the integral over points of $\mathbf{y}$, the denominator of the normalisation constant cannot be reduced. Noting that $q_x(a,l) = \frac{f_Y(l)}{\mathbb{E}[Y_1]}\, \mathbf{1}\{x \in [a, a+l]\}$, evidently $\frac{f_Y(l)}{\mathbb{E}[Y_1]}$ does not depend on $\mathbf{x}$, and that the non-atomic points of $\mathbf{x}$ can be replaced by $\mathbf{y}$. This allows the order of integration to be changed without violating measurability constraints. By Fubini's theorem,

$$\begin{aligned}
\mathbb{E}[1_F(U)\mathbb{P}(X \in A \,|\, U)] &= \int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X \in A \,|\, U = \mathbf{u})\, d\mathbb{P}(\mathbf{u} \,|\, \mathbf{x}) d\mathbb{P}(\mathbf{x}) \\
&= \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{X}|}}{n!} \sum_{C_0 \subseteq \{1,\ldots,n\}} \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}\right)^{|C_0|} \left(\frac{\mathbb{E}[Y_1]}{\mathbb{E}[T_1]}\right)^{n-|C_0|} \\
&\int_{\mathcal{X}^n} p_X(\{x_1,\ldots,x_n\}) 1_A(\mathbf{x}) \sum_{\substack{C_1,\ldots,C_{n-|C_0|} \\ \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0}} \frac{1}{(n-|C_0|)!} \\
&\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-|C_0|}} 1_F(\{u_1,\ldots,u_{n-|C_0|}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\})) \\
&\prod_{j=|C_0|+1}^{n} q_{x_{C_j}}(u_j)\, du_j \prod_{k=1}^{n} dx_k = \mathbb{E}[1_F(U)1_A(X)] \qquad (2.11)
\end{aligned}$$

after cancelling and rearranging terms, and noting that the term in between brackets cancels out against the normalisation constant $c(\{u_1,\ldots,u_{n-|C_0|}\} \cup \{(x_k,0) : k \in C_0\})$. Hence, the posterior distribution of X given $U = \mathbf{u}$ is the union of m atoms combined with $n-m$ points that are distributed on $\mathcal{X}^{n-m}$ according to the symmetric probability density function

$$\begin{aligned}
&c(\mathbf{u})\, p_X(\{a_1,\ldots,a_m,x_1,\ldots,x_{n-m}\}) \\
&\qquad \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_i \{D_i\} = \{1,\ldots,n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \qquad (2.12)
\end{aligned}$$

with respect to Lebesgue measure. □

Corollary 2.6. *In the framework of Theorem 2.5, the conditional distribution of the mark assignments $D_1,\ldots,D_{n-m}$ for non-atomic marks is as follows. For $d_1,\ldots d_{n-m} \in \{1,\ldots,n-m\}$ such that $\{d_1,\ldots,d_{n-m}\} = \{1,\ldots,n-m\}$,*

$$\mathbb{P}(D_1 = d_1,\ldots,D_{n-m} = d_{n-m} | X = \{a_1,\ldots,a_m,x_1,\ldots,x_{n-m}\}, U = \mathbf{u})$$

$$= \frac{\prod_{i=1}^{n-m} \mathbf{1}\{x_{d_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}}{\sum_{\substack{C_1,\ldots,C_{n-m} \\ \cup_i\{C_i\}=\{1,\ldots,n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{C_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}}$$

provided that $x_i \in [a_{m+i}, a_{m+i} + l_{m+i}]$ *for* $i = 1, \ldots, n-m$ *and zero otherwise.*

As a special case, let us consider an inhomogeneous Poisson process with integrable intensity function $\lambda : \mathcal{X} \to \mathbb{R}^+$. Then, under the posterior distribution, X consists of n independent points, one in each interval of $\mathbf{u}$, with probability density function

$$\frac{\lambda(x)}{\int_{[a_i, a_i+l_i]\cap\mathcal{X}} \lambda(s)ds}$$

on $[a_i, a_i + l_i] \cap \mathcal{X}$ for intervals with $l_i > 0$. To see this, recall that for a Poisson process [86]

$$p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) = \exp\left[\int_{\mathcal{X}} (1 - \lambda(s))ds\right] \prod_{j=1}^{m} \lambda(a_j) \prod_{i=1}^{n-m} \lambda(x_i)$$

which factorises over terms associated with each interval. Hence Equation 2.12 in the Poisson case is proportional to

$$\sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_i\{D_i\}=\{1,\ldots,n-m\}}} \prod_{i=1}^{n-m} \lambda(x_{D_i})\, \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}.$$

2.5 Inference methods

To reconstruct the latent point process X from observed parametrised intervals U, statistical inference must be performed. In tandem, the censoring probability as well as the parameters η of the distribution of the non-degenerate intervals must be estimated. Parameters of the prior distribution may either be treated as fixed or subject to estimation - both approaches will be discussed in this chapter.

2.5.1 Model parameter inference scheme

Suppose that we observe the realisation

$$\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0), (a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$$

of U, where $a_i \in \mathbb{R}$, $l_i > 0$ and $n \neq 0$. Our first aim is to estimate the parameters η of the mark distribution ν. The parameter vector η comprises the parameters ζ of the probability density function f_Y as well as any other parameters χ involved in the joint distribution of the random vector $C_1 = (Y_1, Z_1)$ that defines the alternating renewal process.

The likelihood function can be obtained from the proof of Proposition 2.5 by taking A equal to $\mathcal{N}_{\mathcal{X}}$ in Equation 2.9. On a logarithmic scale,

$$L(\eta; \mathbf{u}) = m \log\left(\frac{\mathbb{E}[Z_1; \zeta, \chi]}{\mathbb{E}[T_1; \zeta, \chi]}\right) + (n-m) \log\left(\frac{\mathbb{E}[Y_1; \zeta]}{\mathbb{E}[T_1; \zeta, \chi]}\right) + \sum_{i=1}^{n-m} \log\left(\frac{f_Y(l_i; \zeta)}{\mathbb{E}[Y_1; \zeta]}\right) \tag{2.13}$$

upon ignoring terms that do not depend on η.

Equation 2.13 simplifies greatly if we assume that the mixture weight $p = \mathbb{E}[Z_1; \zeta, \chi] \,/\, \mathbb{E}[T_1; \zeta, \chi]$ does not depend on ζ. This is the case, for example, when Y_1 and Z_1 are independent and Gamma distributed with the same shape parameter k and rate parameters λ for Y_1 and $\chi\lambda$ for Z_1. Then $p = p_X(\chi) = 1/(1+\chi)$ does not depend on $\zeta = (k, \lambda)$ and

$$L(p, \zeta; \mathbf{u}) = m \log p + (n-m) \log(1-p) + \sum_{i=1}^{n-m} \log\left(\frac{f_Y(l_i; \zeta)}{\mathbb{E}[Y_1; \zeta]}\right). \tag{2.14}$$

The atom probability p may be estimated by the sample estimate m/n, the fraction of atoms in the sample $\mathbf{u}$. For ζ, we need the following result.

Proposition 2.7. *The distribution of the length L is given by a length-weighted marginal distribution $f(l) = \frac{l f_Y(l)}{\mathbb{E}[Y_1]}$ and the starting points A are, conditionally on $L = l$, uniformly distributed as $A \sim$ Unif$(-l, 0)$.*

Proof: Let $f(l)$ be the marginal distribution of the length l. Evaluating,

$$f(l) = \int f(a, l)\, da = \int \frac{f_Y(l)}{\mathbb{E}[Y_1]}\, \mathbf{1}\{a < 0 < a + l\}\, da = \int_{-l}^{0} \frac{f_Y(l)}{\mathbb{E}[Y_1]}\, da = \frac{l f_Y(l)}{\mathbb{E}[Y_1]}.$$

Let $f_{A|L=l}(a)$ be the conditional starting probability density function, where L and A are random variables generating the lengths and starting points of the intervals respectively. Using the definition of conditional density,

$$f_{A\,|L=l}(a) = \frac{f(a, l)}{f(l)} = \frac{\frac{f_Y(l)}{\mathbb{E}[Y_1]}\, \mathbf{1}\{a < 0 < a + l\}}{\frac{l f_Y(l)}{\mathbb{E}[Y_1]}} = \frac{1}{l}\, \mathbf{1}\{a \in (-l, 0)\}.$$

Thus $A \sim \text{Unif}(-l, 0)$. □

Corollary 2.8. *In the special case that $Y \sim$ Gamma(k, λ) with $k > 0$ the shape parameter and $\lambda > 0$ the rate parameter, the marginal distribution $f(l) = f_Y(l; k+1, \lambda)$.*

Proof.

$$f_Y(l;k,\lambda) = \frac{\lambda^k l^{k-1} e^{-\lambda l}}{\Gamma(k)}$$

by definition of the density of the gamma distribution. By definition, $\mathbb{E}[Y_i] = \mathbb{E}[Y_1] = \frac{k}{\lambda}$. Now

$$f(l) = \frac{l f_Y(l)}{\mathbb{E}[Y_1]} = \frac{l\lambda^k l^{k-1} e^{-\lambda l}}{\frac{k}{\lambda}\Gamma(k)} = \frac{\lambda^{k+1} l^k e^{-\lambda l}}{\Gamma(k+1)} = f_Y(l;k+1,\lambda)$$

after using the identity that $\Gamma(x+1) = x\Gamma(x)$ and by making use of the inspection paradox for renewals [126]. Hence, if $Y \sim \text{Gamma}(k,\lambda)$, the marginal distribution of the lengths l is exactly a gamma distribution with shape parameter $k+1$ and scale parameter λ, while being scaled by the length of the interval. □

One possible modelling approach would be to simulate the alternating renewal process, and then run the prior for X directly through this censoring mechanism. However, since the mark distribution for the point process is known, observations of points in Y- and Z-phases can be directly sampled from this distribution, taking the expected value for the random variables Y, Z and T along with the distribution function $f_Y(l)$. Recall that the distribution of the marks I is the mixture with an atom of size $\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}$ at $l = 0$ and $a = 0$, and density $\frac{f_Y(l)}{\mathbb{E}[Y_1]}$ for $l > 0$ with respect to Lebesgue measure on $\{(a,l) \in \mathbb{R} \times \mathbb{R}^+ : a < 0 < a + l\}$.

2.5.2 Example: Gamma distribution

Estimators will be used to estimate the parameters k and λ. First, let $\hat{k} = k+1$. Let $L_1, L_2, \ldots, L_n \overset{\text{iid}}{\sim} \text{Gamma}(\hat{k},\lambda)$ be the lengths of a given observed Y-phase, given the observation time is within a Y-phase. Clearly $\mathbb{E}[L_1] = \frac{\hat{k}}{\lambda}$ and $\mathbb{E}[L_1^2] = \frac{\hat{k}+\hat{k}^2}{\lambda^2}$. Using the method of moments,

$$(\hat{k},\lambda) = \left(\frac{L^2}{\frac{1}{n}\sum_{i=1}^n (L_i - L)^2}, \frac{L}{\frac{1}{n}\sum_{i=1}^n (L_i - L)^2}\right)$$

where L is the (sample) mean of $(L_1, \ldots, L_n)$.

For the maximum likelihood, see that

$$f(l) = f_Y(l;k+1,\lambda) = f_Y(l;\hat{k},\lambda).$$

Take a realisation of observations $\mathbf{L} = \{l_1, \ldots, l_n\}$. Now

$$L(\hat{k},\lambda \,|\, \mathbf{L}) = \frac{\lambda^{\hat{k}} l_1^{\hat{k}-1} e^{-\lambda l_1}}{\Gamma(\hat{k})} \cdots \frac{\lambda^{\hat{k}} l_n^{k} e^{-\lambda l_n}}{\Gamma(\hat{k})}$$

$$= \left(\frac{\lambda^{\hat{k}}}{\Gamma(\hat{k})}\right)^n (l_1 \dots l_n)^{\hat{k}-1} e^{-\lambda(l_1+\dots+l_n)}$$

and take

$$\log L(\hat{k}, \lambda \,|\, \mathbf{L}) = n(\hat{k} \log \lambda - \log \Gamma(\hat{k})) + (\hat{k} - 1) \sum_{i=1}^{n} \log l_i - \lambda \sum_{i=1}^{n} l_i$$

to obtain the score function. Evaluating the zeroes of the score function,

$$\begin{aligned}
\frac{\partial}{\partial \lambda}(\log L(\hat{k}, \lambda \,|\, \mathbf{L})) &= n\frac{\hat{k}}{\lambda} - \sum_{i=1}^{n} l_i = 0 \implies l = \frac{\hat{k}}{\lambda} \\
\frac{\partial}{\partial \hat{k}}(\log L(\hat{k}, \lambda \,|\, \mathbf{L})) &= n(\log \lambda - \frac{d}{d\hat{k}} \log \Gamma(\hat{k})) + \sum_{i=1}^{n} \log l_i \\
&= n(\log \hat{k} - \log l - \frac{d}{d\hat{k}} \log \Gamma(\hat{k}) + \overline{\log l}) = 0
\end{aligned}$$

after writing $\lambda = \frac{\hat{k}}{l}$. Using the digamma function, this equation can be solved for $\hat{k}$ and then subsequently λ. The variance is given by the Fisher information matrix

$$I(\hat{k}, \lambda) = n \begin{bmatrix} \frac{d^2}{d\hat{k}^2} \log \Gamma(\hat{k}) & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{\hat{k}}{\lambda^2} \end{bmatrix}.$$

2.5.3 Example: Weibull distribution

For certain datasets, one may want to use a heavy-tailed distribution such as the Weibull distribution. Like the gamma distribution, this distribution has a shape parameter k and a scale parameter λ. The marginal distribution of the lengths l is

$$f(l) = \frac{l f_Y(l)}{\mathbb{E}Y_1} = \frac{k}{\lambda \Gamma(1 + \frac{1}{k})} \left(\frac{l}{\lambda}\right)^k e^{-\left(\frac{l}{\lambda}\right)^k}.$$

The log-likelihood is

$$\begin{aligned}
\log L(k, \lambda \,|\, \mathbf{L}) = n \log k - n(k+1) \log \lambda - n \log \Gamma\left(1 + \frac{1}{k}\right) \\
+ k \sum_{i=1}^{n} \log(l_i) - \sum_{i=1}^{n} \left(\frac{l_i}{\lambda}\right)^k.
\end{aligned}$$

The score functions are

$$\frac{\partial}{\partial k} \left[\log L(k, \lambda \,|\, \mathbf{L})\right] = \frac{n}{k} - n \log(\lambda) + \frac{n\psi\left(1 + \frac{1}{k}\right)}{k^2 \, \Gamma\left(1 + \frac{1}{k}\right)}$$

$$+\sum_{i=1}^{n}\log(l_i)-\sum_{i=1}^{n}\left(\frac{l_i}{\lambda}\right)^k\log\left(\frac{l_i}{\lambda}\right) \tag{2.15}$$

where $\psi(x)=\frac{d}{dx}\log(\Gamma(x))$ is the digamma function, and

$$\frac{\partial}{\partial\lambda}\left[\log L(k,\lambda\,|\,\mathbf{L})\right]=\sum_{i=1}^{n}\frac{k{l_i}^k}{\lambda^{k+1}}-\frac{n(k+1)}{\lambda} \tag{2.16}$$

Setting Equation 2.16 to 0, we obtain

$$\lambda^*=\left(\frac{k}{n(k+1)}\sum_{i=1}^{n}{l_i}^k\right)^{\frac{1}{k}}. \tag{2.17}$$

Plugging 2.17 into Equation 2.15,

$$\begin{aligned}\frac{\partial}{\partial k}\left[\log L(k,\lambda\,|\,\mathbf{L})\right]=\frac{n}{k}-\frac{n}{k}\log\left(\frac{k}{n(k+1)}\sum_{i=1}^{n}{l_i}^k\right)+\frac{n\psi\left(1+\frac{1}{k}\right)}{k^2\,\Gamma\left(1+\frac{1}{k}\right)}\\+\sum_{i=1}^{n}\log\left(l_i\right)-\frac{n(k+1)}{k}\frac{\sum_{i=1}^{n}{l_i}^k\log\left(l_i\right)}{\sum_{i=1}^{n}{l_i}^k}\\+\frac{n(k+1)}{k^2}\frac{\sum_{i=1}^{n}{l_i}^k\log\left(\frac{k}{n(k+1)}\sum_{j=1}^{n}{l_j}^k\right)}{\sum_{i=1}^{n}{l_i}^k}.\end{aligned}$$

This equation is then solved for k numerically and subsequently λ using Expression 2.17.

2.5.4 Prior distributions

The most obvious choice for a prior is the Poisson distribution. It conveys the assumption that events occur more or less randomly. Though this is unlikely to be the case, this approach allows for benchmarking of the simulation method. In this case, the point process X is assumed to have this distribution.

If the distribution of the prior X is Poisson, then it is a counting process $X(t)$ on the real half-line with $X(0)=0$ and independent increments - i.e. for any pair of disjoint Borel sets $B_1,B_2\in\mathcal{B}(\mathcal{X})$ we have $X(B_1)\perp X(B_2)$. Introduce the integrable function $\lambda(x)$ for $x\in\mathcal{X}$. The specific property holds that $\mathbb{E}[X(a,b]]=\Lambda(b-a)=\int_a^b\lambda(x)\,dx$ for the interval $(a,b]$ and

$$\mathbb{P}(X(a,b]=n)=\frac{[\Lambda(b-a)]^n}{n!}e^{-\Lambda(b-a)}.$$

In the case where $\lambda(x)$ is constant for $x\in\mathcal{X}$, the Poisson process is homogeneous.

The area-interaction point process developed by Baddeley and Van Lieshout will be used to capture interactions between separate points in the prior, under the assumption that certain underlying processes may be affecting the behaviours of burglars. The density p with respect to the unit rate Poisson process on $\mathcal{X}$ is

$$p_X(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})}\gamma^{-|U_r(\mathbf{x})|}$$

with $U_r(\mathbf{x}) = \bigcup_{i=1}^{n} B(x_i, r)$ where $B(x_i, r) = \{a \in \mathbb{R}^d : ||a - x_i|| \leq r\}$ [9]. In one dimension, one has $U_r(\mathbf{x}) = \bigcup_{i=1}^{n}[x_i - r, x_i + r]$. Note that this is often written as

$$p_X(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})} \exp[-\log\gamma\,|U_r(\mathbf{x})|]. \tag{2.18}$$

The value α is a normalisation constant, γ is the interaction parameter, and the value of r represents the radius parameter for a given U_r in the exponent of γ. When $\gamma < 0$, the realisation is regular, and for $\gamma > 0$ the realisation is clustered. When $\gamma = 0$, then one has a Poisson process with intensity β (with the assumption that this is the density with respect to a unit rate Poisson process). The values β and r are parameters [9]. Finally, $n(\mathbf{x})$ refers to the number of points in the process, and $|\cdot|$ is the restriction of the Lebesgue measure on $\mathcal{X}$. For simulated data, known parameter values are used to generate the prior. If real data are used, these parameter values would need to be estimated. The Papangelou conditional intensity [114], interpreted as a likelihood ratio between the point process with an added point ξ as compared to the same point process without the addition of this point, is given by

$$\lambda(\xi, \mathbf{x}) = \beta \exp\left[-\log\gamma \left| \frac{B(\xi, r)}{U_r(\mathbf{x})}\right|_{\mathcal{X}}\right] \tag{2.19}$$

when using the above density formulation [9, 34].

2.5.5 Markov chain Monte Carlo simulation

Since the normalisation constant in e.g. Equation 2.19 remains unknown, Markov chain Monte Carlo methods will be used in order to obtain samples from the posterior distribution. These methods generate a Markov chain of random variables $Y_0, Y_1, \ldots$ where the conditional distribution Y_m given $Y_0, \ldots, Y_{m-1}$ is equivalent to Y_m given Y_{m-1}. The aim is to construct a chain such that the stationary distribution of the chain is exactly the posterior distribution $\mathbb{P}_{X|U}(\cdot|U)$[129].

Of these methods, a Metropolis-Hastings algorithm with a fixed number of points will be used to sample from the posterior density $p_{X|U}(\mathbf{x} \,|\, \mathbf{u})$ on $\mathcal{X}^{n-m}$, recalling that $\mathbf{u} = \{(a_i, l_i)_{i=1}^{n}\}$ be a realisation of intervals. As before, $n = n(\mathbf{u})$ and m is the number of atoms. The benefit of the Metropolis-Hastings algorithm is that one can sample from a given probability distribution $p_X(\mathbf{x})$ given that a function

$f(x)$ is known such that $f \propto p$, i.e. they may differ by a normalisation factor [102]. One would like to sample from

$$p_{X|U}(\mathbf{x} \mid \mathbf{u}) = c(\mathbf{u})\, p_X(\mathbf{x}) \sum_{\substack{\emptyset \neq D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[u_{i,1},u_{i,2}+u_{i,1}]}(x_{D_i}) \tag{2.20}$$

where $c(\mathbf{u})$ is the normalisation constant and $p_X(\mathbf{x})$ is the prior density. Note that the prior point process density must be bounded to facilitate convergence of the algorithm in time. Hence the Metropolis-Hastings algorithm allows us to sample from the unnormalised density function

$$f(\mathbf{x} \mid \mathbf{u}) = p_X(\mathbf{x}) \sum_{\substack{\emptyset \neq D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[u_{i,1},u_{i,2}+u_{i,1}]}(x_{D_i}),$$

conditional on $X(\mathcal{X}) = n - m$.

Note that sampling from $f(\mathbf{x} \mid \mathbf{u})$ becomes intractable due to the presence of the permutation sum term. We saw in Chapter 2.1 that, due to the independence of the Z_i and Y_i, for a point falling at t, a mark distribution I on $\mathcal{I}$ is induced where

$$I = \begin{cases} (0,0) & t \in [s_i + y_{i+1}, s_{i+1}) \\ (s_i - t, y_{i+1}) & t \in [s_i, s_i + y_{i+1}) \end{cases}$$

where s_i is the ith renewal. Using the parametrisation of the alternating renewal process (see Equation 2.4), one gets a marked point process of the form $W = \{(t_1, I_1), \ldots, (t_n, I_n)\}$.

Instead of looking at the attribution of points to intervals which the sum in $f(\mathbf{x} \mid \mathbf{u})$ describes, one can directly sample from the density $p_{X|U}(\mathbf{x} \mid \mathbf{u})$ by considering the marked model W given U. Write $p_W(\mathbf{w} \mid U = \mathbf{u})$ as the conditional density function for W, the complete, marked model, given U. Hence

$$p_W(\mathbf{w} \mid U = \mathbf{u}) = p_W(\{(t_1, I_1), \ldots, (t_n, I_n)\} = p_X(\{t_1, \ldots, t_n\}) = p_X(\mathbf{x})$$

with respect to the Poisson process on $\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)$ with intensity measure $\mu \times \nu$, where μ is the Lebesgue measure, over all possible configurations $\mathcal{N}_{\mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+)}$.

The Metropolis-Hastings algorithm simulates a possible realisation of this density $p_X(\{t_1, \ldots, t_n\})$, which can only be attained over the set $\overline{E}(\mathbf{u})$ of compatible configurations. The state space of the chain equipped with density function $p_X(\mathbf{x})$ is

$$\overline{E}(\mathbf{u}) = \{(x_1, \ldots, x_{n-m}) \in \mathcal{X}^{n-m} : x_i \in \mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}],$$

$$p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) > 0\}.$$

for a given realisation $\mathbf{u}$. We also assume that the state space is non-degenerate in the sense that

$$\int_{\overline{E}(\mathbf{u})} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})\, dx_1 \ldots dx_{n-m} > 0. \tag{2.21}$$

We proceed with algorithm construction in the following manner:

Denote the states of the Markov chain by $Y_0, Y_1, \ldots$. Let $\mathbf{x}_0$ be an initial realisation of X. We begin by setting $Y_0 = \mathbf{x}_0 \in \overline{E}(\mathbf{u})$. Let $Y_k = \mathbf{x}$ be the current state of the realisation of X at time k. To generate Y_{k+1}, take $J_k \sim \text{Unif}(\{1, \ldots, m\})$ where m is the number of atoms, and $R_k \sim \text{Unif}(0, 1)$. J_k generates a random index i for the state Y_k and $R_k = r$ is a uniformly sampled random number. Formally, introduce the pointwise Markov kernel

$$q_i(\mathbf{x}, x_i) = \frac{1}{l_i}\delta\{(x_i, (a_i - x_i, l_i))\} \cup \mathbf{x} \setminus \{x_i\},$$

which picks a new point ξ as an alternative to x_i. In our case, we draw the point from the ith interval, so the point is drawn from $\text{Unif}(a_i, a_i + l_i)$. Note that δ is the Dirac delta function.

Define the Hastings ratio

$$r_i(\mathbf{x}, \xi) = \frac{p_X((\mathbf{x} \setminus x_i) \cup \xi)\, q_i(\{x_1, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_n\}, x_i)}{p_X(\mathbf{x})\, q_i(\mathbf{x}, \xi)} \tag{2.22}$$

which compares the probability of occurrence of a new state Y_{k+1} with the newly generated point ξ, as compared to the old state Y_k [102]. Then

$$Y_{k+1} = \begin{cases} \{x_1, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_n\} & r \leq r_i(\mathbf{x}, \xi) \\ Y_k & \text{otherwise,} \end{cases}$$

letting ξ be the proposal point at time k. Define

$$p_X(\mathbf{x}, F) = p_X(Y_{k+1} \in F \,|\, Y_k = \mathbf{x})$$

for $F \subseteq \overline{E}(\mathbf{u})$. The acceptance rate is

$$\alpha_i(\mathbf{x}, \xi) = \min\left(1, \frac{p_X((\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\} \setminus \{x_i\}) \cup \{\xi\}, F)}{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}, F)}\right). \tag{2.23}$$

Let the target distribution be $\pi(\mathbf{x})$. Summarising, we propose the following algorithm:

Algorithm 2.9. *Suppose that $p_X > 0$ and $n > m$. Iteratively, if the current state is $\mathbf{x} \in \overline{E}(\mathbf{u})$,*

- *pick an interval $[a_{m+i}, a_{m+i} + l_{m+i}]$, $i = 1, \ldots, n-m$, uniformly at random from the non-degenerate ones;*
- *generate a uniformly randomly distributed point ξ on $\mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}]$ and propose to update x_i to ξ;*
- *accept the proposal with probability $\alpha_i(\mathbf{x}, \xi)$, otherwise stay in the current state.*

A few remarks are in order. First, note that since $\mathcal{X}$ is open, its intersection with closed intervals that contain a point in $\mathcal{X}$ is also non-degenerate when $l_i > 0$. Secondly, when p_X may take the value zero, the proposal mechanism in Algorithm 2.9 might result in a new state that does not belong to $\overline{E}(\mathbf{u})$, even when $\mathbf{x}$ does. Moreover, only changing one component at a time might lead to non-irreducible Markov chains. For example, if $\mathbf{u}$ contains the parametrisations of the intervals $[0, 1]$ and $[0.1, 1]$ and $p_X(\mathbf{x}) = 0$ for realisations $\mathbf{x}$ that contain components separated by a distance less than 0.55, then states such as $\mathbf{x} = (0.3, 0.9)$ and $\mathbf{y} = (0.9, 0.3)$ cannot be reached from one another.

Additionally, we introduce the general Markov kernel $q : \overline{E}(\mathbf{u}) \times \overline{E}(\mathbf{u})$ which takes as arguments the old state $\mathbf{x}$ and the proposed new state $\mathbf{y}$, in contrast to the pointwise version, which takes as argument only the point that will potentially be replaced. In a similar vein, we introduce a generalised acceptance probability $\alpha : \overline{E}(\mathbf{u}) \times \overline{E}(\mathbf{u})$. Formally, $\alpha(\mathbf{x}, \mathbf{y}) = 1$ if

$$p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\})\, q(\bar{\mathbf{y}}, \bar{\mathbf{x}}) \geq p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})\, q(\bar{\mathbf{x}}, \bar{\mathbf{y}}),$$

otherwise

$$\alpha(\mathbf{x}, \mathbf{y}) = \frac{p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\})\, q(\bar{\mathbf{y}}, \bar{\mathbf{x}})}{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})\, q(\bar{\mathbf{x}}, \bar{\mathbf{y}})}.$$

In the next propositions, basic properties of the algorithm are considered. The proofs are modifications to our context of classic Metropolis-Hastings proofs found in, for example, [99], [124] or [102, Chapter 7]. Note that for $\tau = 2, 3, \ldots$, $P^\tau(\mathbf{x}, F)$ denotes the τ-step transition probability.

Proposition 2.10. *Consider the setup of Theorem 2.5 with $n > m$. Then, the Metropolis-Hastings algorithm defined by Markov kernel q on $\overline{E}(\mathbf{u})$ and acceptance probabilities (2.23), is reversible with respect to π.*

Proof. Take $\mathbf{x}, \mathbf{y}$ in $\overline{E}(\mathbf{u})$ and assume that $\pi(\mathbf{y}), q(\mathbf{y}, \mathbf{x}) > \pi(\mathbf{x})\, q(\mathbf{x}, \mathbf{y}) \geq 0$. Then

$$\begin{aligned}\pi(\mathbf{x})\, q(\mathbf{x}, \mathbf{y})\, \alpha(\mathbf{x}, \mathbf{y}) &= c(\mathbf{u})\, p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})\, q(\mathbf{x}, \mathbf{y}) \\ &= c(\mathbf{u})\, p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\}) \\ &\quad q(\mathbf{y}, \mathbf{x})\, \frac{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})\, q(\mathbf{x}, \mathbf{y})}{p_X(\{a_1, \ldots, a_m, y_1, \ldots, y_{n-m}\})\, q(\mathbf{y}, \mathbf{x})} \\ &= \pi(\mathbf{y})\, q(\mathbf{y}, \mathbf{x})\, \alpha(\mathbf{y}, \mathbf{x})\end{aligned}$$

writing $c(\mathbf{u})$ for the normalisation constant. We conclude that the chain is in detailed balance and therefore reversible with respect to π. □

Recall that a Markov chain is called π-irreducible [100] if for every $\mathbf{x} \in \overline{E}(\mathbf{u})$ and every Borel set $F \subset \overline{E}(\mathbf{u})$ with $\pi(F) > 0$ there exists some natural number τ such that $P^\tau(\mathbf{x}, F) > 0$.

Proposition 2.11. *Consider the setup of Theorem 2.5 with $n > m$ and assume that condition (2.21) is met. Let Q be the one-step transition kernel of the Markov chain on $\overline{E}(\mathbf{u})$ generated by Markov kernel $q : \overline{E}(\mathbf{u}) \times \overline{E}(\mathbf{u}) \to \mathbb{R}^+$ in which every proposal is accepted. If the chain defined by Q is π-irreducible and $q(\mathbf{x}, \mathbf{y}) = 0$ if and only if $q(\mathbf{y}, \mathbf{x}) = 0$, then the Metropolis-Hastings algorithm defined by q and (2.23) is π-irreducible. In particular, the chain of Algorithm 2.9 is π-irreducible when $p_X > 0$.*

Proof. The first part follows from [124, Theorem 3.ii].

For Algorithm 2.9, $q(\mathbf{x}, \mathbf{y}) > 0$ only if $\mathbf{x}, \mathbf{y} \in \overline{E}(\mathbf{u})$ differ in at most a single component. Thus, assume that $x_j = y_j$ for all $j \neq i \in \{1, \ldots, n-m\}$ and $x_i \neq y_i$. Then $q(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}, \mathbf{x})$ so they are strictly positive or zero together. Write Q^τ for the τ-step transition kernel of the aways-accept chain. Then, for $\mathbf{x}$, $\mathbf{y} \in \overline{E}(\mathbf{u})$,

$$q^{n-m}(\mathbf{x}, \mathbf{y}) \geq \left(\frac{1}{n-m}\right)^{n-m} \prod_{i=1}^{n-m} \frac{1}{\ell(\mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}])} > 0$$

by changing each component in turn. We conclude that the Markov chain of Algorithm 2.9 is π-irreducible. □

Recall that a π-irreducible Markov chain is called aperiodic [100] if the state space cannot be partitioned into measurable sets $B_0, B_1, \ldots B_{d-1}$ such that $\pi(\overline{E}(\mathbf{u}) \setminus \cup_{j=0}^{d-1} B_j) = 0$ and $P(\mathbf{x}, B_{j+1 \bmod d}) = 1$ for all $\mathbf{x} \in B_j$ (for some $d > 1$, the period). By [102, Proposition 7.6], a π-irreducible Markov chain with invariant probability distribution π is aperiodic if and only if for some small set D with

$\pi(D) > 0$ and some $\tau \in \mathbb{N}$, the following holds: $P^i(\mathbf{x}, D) > 0$ for all $\mathbf{x} \in D$ and $i \geq \tau$.

Proposition 2.12. *Consider the setup of Theorem 2.5 with $n > m$. If $0 < p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}) \leq \delta$ for some $\delta > 0$ and all $\mathbf{x} \in \overline{E}(\mathbf{u})$, then the Markov chain of Algorithm 2.9 is aperiodic.*

Proof. Let ξ be the point on $\mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}]$ that replaces x_1. By (2.21), there exist $x_2, \ldots, x_{n-m}$ such that

$$\int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} c(\mathbf{u})\, p_X(\{a_1, \ldots, a_m, \xi, x_2, \ldots, x_{n-m})\, d\xi = \int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} \pi(\xi, x_2, \ldots, x_{n-m})\, d\xi$$

is strictly positive, where $c(\mathbf{u})$ is the normalisation constant. Define a measure μ on the Borel σ-algebra on $\overline{E}(\mathbf{u})$ by

$$\mu(F) = \int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} \mathbf{1}\{(\xi, x_2, \ldots, x_{n-m}) \in F\}\, \pi(\xi, x_2, \ldots, x_{n-m})\, d\xi$$

and note that $\mu(\overline{E}(\mathbf{u})) > 0$. Set $C = \{(\xi, x_2, \ldots, x_{n-m}) : \xi \in \mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}]\}$. We claim that C is small with respect to μ. To see this, take $\mathbf{y} = (y, x_2, \ldots, x_{n-m}) \in C$ and note that for $F \subset \overline{E}(\mathbf{u})$, the transition probability $P(\mathbf{y}, F)$ is at least

$$\frac{1}{n-m} \frac{1}{\ell(\mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}])} \int_{\mathcal{X} \cap [a_{m+1}, a_{m+1}+l_{m+1}]} \mathbf{1}_F(\xi, x_2, \ldots, x_{n-m})\, \alpha_1((y, x_2, \ldots, x_{n-m}), \xi)\, d\xi.$$

If $\pi(\xi, x_2, \ldots, x_{n-m}) \geq \pi(y, x_2, \ldots, x_{n-m})$, then

$$\alpha_1((y, x_2, \ldots, x_{n-m}), \xi) = 1 \geq \pi(\xi, x_2, \ldots, x_{n-m})/(c(\mathbf{u})\, \delta).$$

Otherwise,

$$\begin{aligned} \alpha_1((y, \ldots, x_{n-m}), \xi) &= \pi(\xi, x_2, \ldots, x_{n-m})/\pi(y, x_2, \ldots, x_{n-m}) \\ &\geq \pi(\xi, x_2, \ldots, x_{n-m})/(c(\mathbf{u})\, \delta). \end{aligned}$$

In summary,

$$P(\mathbf{y}, F) \geq \frac{1}{n-m} \frac{1}{\ell(\mathcal{X} \cap [a_{m+1}, a_{m+1} + l_{m+1}])} \frac{\mu(F)}{c(\mathbf{u})\, \delta}$$

so C is small with respect to μ. Moreover, $\pi(C) > 0$ because of the choice of $x_2, \ldots, x_{n-m}$.

Iterating the above argument one notices that for $\mathbf{y} \in C$, $P^{\tau}(\mathbf{y}, C)$ is at least as large as the τ-th power of the bound above with $F = \overline{E}(\mathbf{u})$, an observation that completes the proof. □

In conclusion, under mild conditions, from almost all initial states, Algorithm 2.9 converges in total variation to the invariant probability distribution. Conditions for general proposal kernels q can be found in [102, Chapter 7] or [124, Theorem 3].

2.6 Simulations for different priors

In this chapter, we present a few simulations to illustrate how the choice of prior affects state estimation. Calculations were carried out using the `C++` marked point process library MPPLIB, developed by Steenbeek et al. and documented in [92]. An initial realisation of the area-interaction process with parameters (β, η, r) is generated using Kendall's dominating coupling from the past (CFTP) algorithm [78] developed initially from the perfect simulation methods of [118] for coupled Markov chains. Note that the normalisation constant α does not need to be calculated, due to the process's simulation being dependent on posterior ratios. This theme runs through a considerable amount of the simulation literature for spatio-temporal point processes. The interval parametrisation dependent on the parameter vector is generated before the beginning of the CFTP algorithm, however births and deaths within this process do not depend on these intervals, but rather conditional expectation calculations relating to the suggested points $\{x_1, \ldots, x_n\}$ contained in these intervals. After this initial generation, the Metropolis-Hastings algorithm described in Chapter 2.5 moves the points within the complex interval geometry based on the conditional expectation of a uniformly chosen proposal point within a certain interval.

2.6.1 Toy example

Consider data $\mathbf{u} = \{(0.45, 0.4), (0.51, 0), (0.58, 0)\}$ that consist of two atoms and a single non-degenerate interval. By the discussion at the end of Chapter 2.4, for a Poisson prior ($\gamma = 1$), the posterior distribution of the location X_3 in $\mathcal{X} = (0, 1)$ that generated the non-degenerate interval is uniformly distributed. To see the effect of informative priors, Figure 2.3 plots the posterior distribution of X_3 when the prior is an area-interaction model with $\eta = 2r \log \gamma = 1.2$ and $r = 0.1$. Note that mass is shifted to the left side of the interval due to the presence of atoms. For $\eta = -1.2$ and $r = 0.1$, the atoms repel X_3, resulting in mass being shifted to the right side of the interval (cf. Figure 2.4). To carry out the state estimation, we ran the Metropolis-Hastings algorithm with a burn-in of 10,000 steps and calculated the histograms based on the subsequent 100,000 steps.

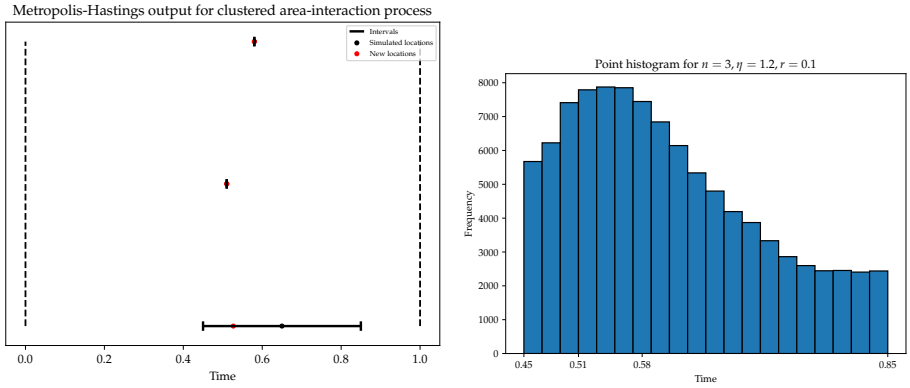


Figure 2.3 Locations of two atoms and a spanning interval together with a histogram of point locations. The interval start and end points, as well as the atom locations, are marked on the histogram x-axis.

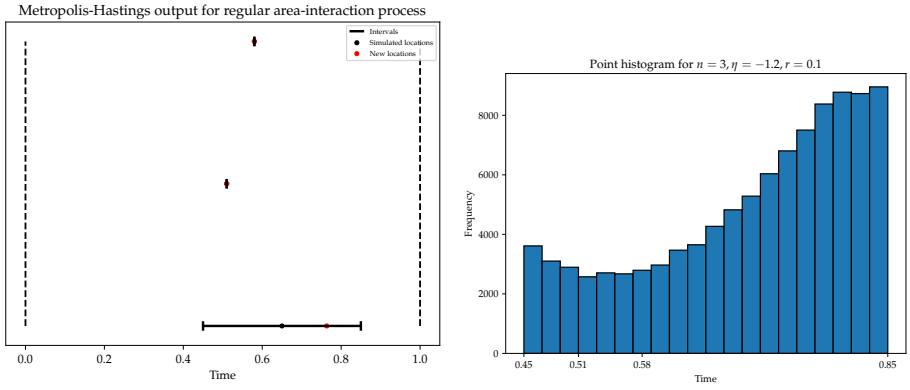


Figure 2.4 Locations of two atoms and a spanning interval together with a histogram of point locations. The interval start and end points, as well as the atom locations, are marked on the histogram x-axis.

2.6.2 The clustered case

In the clustered case, $\eta = \log \gamma > 0$. The parameter values were $(\beta, \eta, r, \lambda, k, p) = (12, 1.2, 0.05, 0.07, 2.5, 0.2)$. The gamma distribution that generated the interval lengths is therefore $L \sim \text{Gamma}(3.5, 0.07)$. A sample simulation can be found in Figure 2.5. In this simulation, 12 points were generated, of which 1 did not admit a density. The Metropolis-Hastings algorithm was run until convergence. Figure 2.5 shows the restrictive nature of the geometry that the Metropolis-Hastings algorithm has to work with for choices of individual points. Instead of having the entire window space $\mathcal{X} = [0, 1]$ to work with, proposal points are generated uniformly only within the parametrised intervals. The initial realisation of the

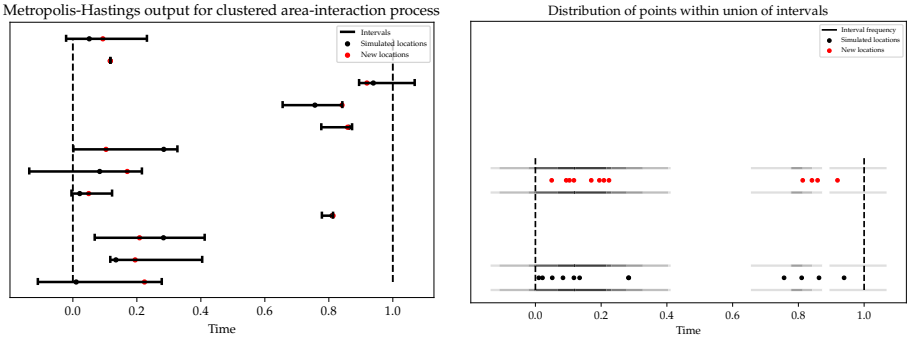


Figure 2.5 Plots of the prior and posterior locations of an area-interaction point process constrained by intervals in one-dimensional space. The left plot shows the intervals plotted above each other to show their locations within the intervals, whereas the right plot shows the raw input and output of the Metropolis-Hastings algorithm together with an interval frequency map. Black points represent the prior locations, whereas red points represent the points after the Metropolis-Hastings algorithm has converged. The dashed lines represent the edge of the window, which in this case is the interval $[0, 1]$.

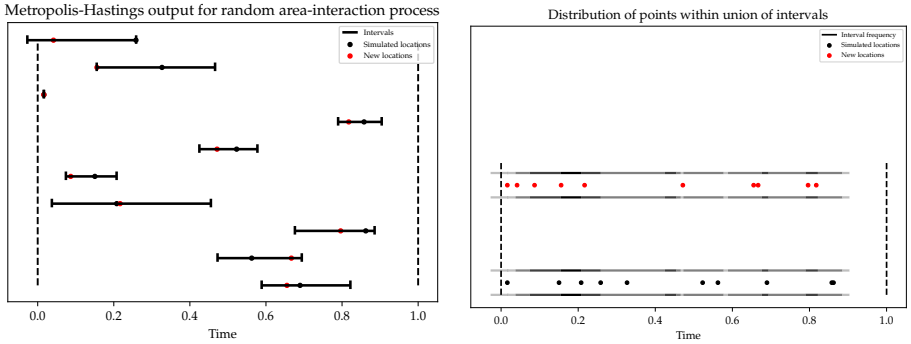


Figure 2.6 Plots of the prior and posterior locations of a random area-interaction point process.

area-interaction process does not optimise the positions of the times given their respective intervals, only in relation to the other initial times. One can see that the algorithm tends to move proposed times to areas where multiple intervals intersect, leading to clustering within these regions. The assumption of clustering in the prior crystallises with exposure to a favourable interval structure.

2.6.3 The random case

In the case where $\eta = \log \gamma = 0$, Figure 2.6 shows the behaviour of the times before and after the application of the fixed-point Metropolis-Hastings algorithm.

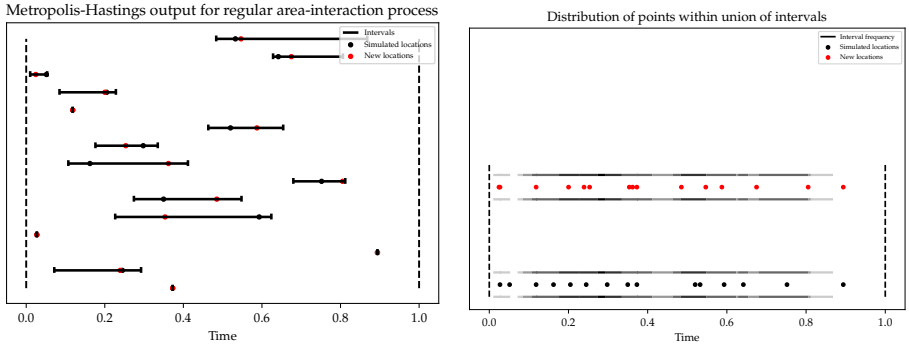


Figure 2.7 Plots of the simulated and new locations of a random area-interaction point process.

The parameter values were $(\beta, \eta, r, \lambda, k, p) = (12, 0, 0.05, 0.07, 2.5, 0.2)$, i.e. the only differing factor was the value of η. 10 points were generated, with one of those being atomic. Here, the points again tend to drift towards areas of high interval frequency, due to the geometry of the process. However, there is no real clustering or repulsive behaviour occurring, and the points settle in a random manner within the intervals.

2.6.4 The regular case

In the regular case, parameters of $(\beta, \eta, r, \lambda, k, p) = (12, -1.2, 0.05, 0.07, 2.5, 0.2)$ were chosen. This means $\eta = \log \gamma < 0$. Figure 2.7 shows that the structure of the prior point process is maintained after simulation, with points being spread out from each other. In this case, 4 of the 15 points generated have an interval length of 0, meaning that they stay in the exact same place during the entire simulation process.

A simulation study has been performed in order to test the viability of the temporal model introduced previously. The structure of the prior distribution, which encodes assumptions about the nature of the underlying process, was maintained. The data constrains the realisations to take place within certain subsets of the observation space - in this case, two times within which the realisation of the ground process must lie. The Metropolis-Hastings algorithm maintains the structure of the prior while taking this data into account.

2.7 Estimation of prior parameters

In Chapter 2.5.1, parameters of the forward model were estimated. In the simulation case, prior parameters are simply set to arbitrary values to show off the behaviour of the model under certain circumstances. However, when faced with

a dataset, prior parameters must also be estimated. To enable this, a maximum likelihood procedure is performed. For Markov point processes, typically the probability density is rewritten as $p_X(\mathbf{x};\theta) = \frac{h_X(\mathbf{x};\theta)}{Z(\theta)}$, where h_X is the unnormalised density and Z is the normalisation constant or Zustandssumme. Let θ denote the parameter vector for a given prior, and θ_0 be the reference parameter vector. Let N denote the number of samples. Assuming all derivatives exist, write the likelihood $l_N(\theta)$ for a given sample size and parameter vector

$$l_N(\theta) = \log\left(\frac{1}{N}\sum_{i=1}^{N}\frac{h_X(X_{\mathbf{u},i};\theta)}{h_X(X_{\mathbf{u},i};\theta_0)}\right) - \log\left(\frac{1}{N}\sum_{i=1}^{N}\frac{h_X(X_i;\theta)}{h_X(X_i;\theta_0)}\right)$$

where $X_{\mathbf{u},i}$ are samples from the posterior and X_i from the prior [53, 86]. An approximate maximum likelihood estimator can be obtained by taking the gradient with respect to the parameter θ and equating to zero. Hence,

$$\nabla l_N(\theta) = \frac{\sum_{i=1}^{N}\frac{\nabla h_X(X_{\mathbf{u},i};\theta)}{h_X(X_{\mathbf{u},i};\theta_0)}}{\sum_{i=1}^{N}\frac{h_X(X_{\mathbf{u},i};\theta)}{h_X(X_{\mathbf{u},i};\theta_0)}} - \frac{\sum_{i=1}^{N}\frac{\nabla h_X(X_i;\theta)}{h_X(X_i;\theta_0)}}{\sum_{i=1}^{N}\frac{h_X(X_i;\theta)}{h_X(X_i;\theta_0)}} = 0$$

must be solved to obtain a maximum likelihood estimate for θ. Note that θ_0 must be aptly chosen for this procedure to work. A common method to choose θ_0 is via the Monte Carlo EM procedure. In this process, one wishes to optimise

$$Q(\theta,\theta_k) = \mathbb{E}_{\theta_k}[\log p_X(X \cup \mathbf{a};\theta) \,|\, U = \mathbf{u}],$$

writing $\mathbf{a}$ as the atoms of $\mathbf{u}$ [53]. To do this, one must solve

$$\nabla Q(\theta,\theta_k) = \frac{1}{N}\sum_{i=1}^{N}\frac{\nabla h_X(X_{\mathbf{u},i};\theta)}{h_X(X_{\mathbf{u},i};\theta)} - \frac{\sum_{i=1}^{N}\frac{\nabla h_X(X_i;\theta)}{h_X(X_i;\theta_k)}}{\sum_{i=1}^{N}\frac{h_X(X_i;\theta)}{h_X(X_i;\theta_k)}} = 0$$

over θ.

One may solve $l_N(\theta)$ iteratively for 0, however this may lead to long processing times if candidate values are simply enumerated in ascending or descending order. To deal with this, Newton's method in optimisation can be used to provide better and better estimates for θ_k such that eventually one converges to the optimal θ. In this paradigm, set

$$\theta_{k+1} = \theta_k + \frac{\nabla l_N(\theta_k)}{\nabla^2 l_N(\theta_k)} \tag{2.24}$$

after each iteration. This sequence converges to θ_{opt} in a finite number of steps, assuming l is twice differentiable. Values for $\nabla^2 l_N(\theta)$ can be found by using the quotient rule by rewriting the score function as

$$\nabla l_N(\theta) = \frac{\sum_{i=1}^{N} f_{\mathbf{u},i}(\theta)}{\sum_{i=1}^{N} g_{\mathbf{u},i}(\theta)} - \frac{\sum_{i=1}^{N} f_i(\theta)}{\sum_{i=1}^{N} g_i(\theta)}. \tag{2.25}$$

Write $f_{\mathbf{u}}(\theta) = \sum_{i=1}^{N} f_{\mathbf{u}}(\theta)$ and subsequently, noting that the derivative operator is linear, use the quotient rule to derive the second derivative form

$$\nabla^2 l_N(\theta) = \frac{\nabla f_{\mathbf{u}}(\theta) g_{\mathbf{u}}(\theta) - \nabla g_{\mathbf{u}}(\theta) f_{\mathbf{u}}(\theta)}{(g_{\mathbf{u}}(\theta))^2} - \frac{\nabla f(\theta) g(\theta) - \nabla g(\theta) f(\theta)}{(g(\theta))^2}.$$

2.7.1 Example: Area-interaction process

In the area-interaction case, recall that $p_X(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})} \exp[-\log\gamma\,|U_r(\mathbf{x})|]$. Making the substitution $\eta = 2r\log\gamma$ and letting $A(\mathbf{x}) = |U_r(\mathbf{x})|$, we can rewrite this as $p_X(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})} e^{-\frac{\eta A(\mathbf{x})}{2r}}$. Using the definition of h,

$$h(\mathbf{x};\theta) = \beta^{n(\mathbf{x})} e^{-\frac{\eta A(\mathbf{x})}{2r}},$$

where $\theta = (\eta, \beta)$. For the likelihood equation over θ,

$$\begin{aligned} l_N(\theta) = &\log\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right) \\ &- \log\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)} e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}\right). \end{aligned}$$

Calculating the derivatives, noting that

$$\frac{\partial h}{\partial \eta} = -\frac{A(\mathbf{x})}{2r}\beta^{n(\mathbf{x})} e^{-\frac{\eta A(\mathbf{x})}{2r}}, \qquad \frac{\partial h}{\partial \beta} = n(\mathbf{x})\beta^{n(\mathbf{x})-1} e^{-\frac{\eta A(\mathbf{x})}{2r}},$$

it follows that

$$\begin{aligned} \frac{\partial l_N(\theta)}{\partial \eta} = &\frac{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})} \frac{-A(X_{\mathbf{u},i})}{2r} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}}{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}} \\ &- \frac{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)} \frac{-A(X_i)}{2r} e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}}{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)} e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial l_N(\theta)}{\partial \beta} = &\frac{\sum_{i=1}^{N} \frac{n(X_{\mathbf{u},i})\beta^{n(X_{\mathbf{u},i}))-1}}{\beta_0^{n(X_{\mathbf{u},i})}} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}}{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}} \\ &- \frac{\sum_{i=1}^{N} \frac{n(X_i))\beta^{n(X_i))-1}}{\beta_0^{n(X_i)}} e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}}{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)} e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}} \end{aligned}$$

after simplifying. These become the two elements of the row vector $\left[\frac{\partial l_N(\theta)}{\partial \eta} \;\; \frac{\partial l_N(\theta)}{\partial \beta}\right]$. Rewriting the score functions in the manner shown in Equation 2.25, we can find, using the quotient rule, the second partial derivatives

$$\frac{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}\frac{A(X_{\mathbf{u},i})}{4r^2}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}$$
$$-\frac{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}\frac{-A(X_{\mathbf{u},i})}{2r}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}-(\dots)=\frac{\partial^2 l_N(\theta)}{\partial\eta^2}$$

and

$$\frac{\partial^2 l_N(\theta)}{\partial\beta^2}=\frac{\left[\sum_{i=1}^{N}\frac{n(X_{\mathbf{u},i})(n(X_{\mathbf{u},i})-1)\beta^{n(X_{\mathbf{u},i}))-2}}{\beta_0^{n(X_{\mathbf{u},i})}}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}$$
$$\times\frac{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}$$
$$-\frac{\left[\sum_{i=1}^{N}\frac{n(X_{\mathbf{u},i})\beta^{n(X_{\mathbf{u},i}))-1}}{\beta_0^{n(X_{\mathbf{u},i})}}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}-(\dots)$$

and

$$\frac{\partial^2 l_N(\theta)}{\partial\beta\partial\eta}=\frac{\left[\sum_{i=1}^{N}\frac{n(X_{\mathbf{u},i})\beta^{n(X_{\mathbf{u},i}))-1}}{\beta_0^{n(X_{\mathbf{u},i})}}\left(\frac{-A(X_{\mathbf{u},i})}{2r}\right)e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}$$
$$\times\frac{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}$$
$$-\frac{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}\frac{-A(X_{\mathbf{u},i})}{2r}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})}e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2}$$

$$\times \frac{\left[\sum_{i=1}^{N} \frac{n(X_{\mathbf{u},i})\beta^{n(X_{\mathbf{u},i}))-1}}{\beta_0^{n(X_{\mathbf{u},i})}} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]}{\left[\sum_{i=1}^{N} \left(\frac{\beta}{\beta_0}\right)^{n(X_{\mathbf{u},i})} e^{\frac{-A(X_{\mathbf{u},i})}{2r}(\eta-\eta_0)}\right]^2} - (\dots)$$
$$= \frac{\partial^2 l_N(\theta)}{\partial\eta\partial\beta}$$

where $(\dots)$ denotes the unconditional versions of the formulas, which are the same except $X_{\mathbf{u},i}$ is replaced with X_i. By definition, the Hessian is

$$H_{l_N(\theta)} = \begin{bmatrix} \frac{\partial^2 l_N(\theta)}{\partial\eta^2} & \frac{\partial^2 l_N(\theta)}{\partial\beta\partial\eta} \\ \frac{\partial^2 l_N(\theta)}{\partial\eta\partial\beta} & \frac{\partial^2 l_N(\theta)}{\partial\beta^2} \end{bmatrix}$$

and we can find subsequent θ_k by applying Equation 2.24. In the case where $\theta = (\eta, \beta)$,

$$\begin{bmatrix}\eta_{k+1} & \beta_{k+1}\end{bmatrix} = \begin{bmatrix}\eta_k & \beta_k\end{bmatrix} + \begin{bmatrix} \frac{\partial l_N(\theta_k)}{\partial\eta_k} & \frac{\partial l_N(\theta_k)}{\partial\beta_k} \end{bmatrix} H^{-1}_{l_N(\theta_k)}$$

$$= \begin{bmatrix}\eta_k & \beta_k\end{bmatrix} + \begin{bmatrix} \frac{\partial l_N(\theta_k)}{\partial\eta_k} & \frac{\partial l_N(\theta_k)}{\partial\beta_k} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 l_N(\theta_k)}{\det(H_{l_N(\theta)})\partial\beta_k^2} & -\frac{\partial^2 l_N(\theta_k)}{\det(H_{l_N(\theta)})\partial\beta_k\partial\eta_k} \\ -\frac{\partial^2 l_N(\theta_k)}{\det(H_{l_N(\theta)})\partial\eta_k\partial\beta_k} & \frac{\partial^2 l_N(\theta_k)}{\det(H_{l_N(\theta)})\partial\eta_k^2} \end{bmatrix}$$

which means that

$$\eta_{k+1} = \eta_k + \frac{\frac{\partial l_N(\theta_k)}{\partial\eta_k}\frac{\partial^2 l_N(\theta_k)}{\partial\beta_k^2} - \frac{\partial l_N(\theta_k)}{\partial\beta_k}\frac{\partial^2 l_N(\theta_k)}{\partial\eta_k\partial\beta_k}}{\frac{\partial^2 l_N(\theta_k)}{\partial\eta_k^2}\frac{\partial^2 l_N(\theta_k)}{\partial\beta_k^2} - \frac{\partial^2 l_N(\theta_k)}{\partial\beta_k\partial\eta_k}\frac{\partial^2 l_N(\theta_k)}{\partial\eta_k\partial\beta_k}}$$

and

$$\beta_{k+1} = \beta_k + \frac{\frac{\partial l_N(\theta_k)}{\partial\beta_k}\frac{\partial^2 l_N(\theta_k)}{\partial\eta_k^2} - \frac{\partial l_N(\theta_k)}{\partial\eta_k}\frac{\partial^2 l_N(\theta_k)}{\partial\beta_k\partial\eta_k}}{\frac{\partial^2 l_N(\theta_k)}{\partial\eta_k^2}\frac{\partial^2 l_N(\theta_k)}{\partial\beta_k^2} - \frac{\partial^2 l_N(\theta_k)}{\partial\beta_k\partial\eta_k}\frac{\partial^2 l_N(\theta_k)}{\partial\eta_k\partial\beta_k}}.$$

The values η_0 and β_0 are known as reference values, and are chosen by performing the Monte Carlo EM procedure. For this, we require the partial derivatives of Q, which are given by

$$\frac{\partial Q(\theta,\theta_k)}{\partial\eta} = \frac{1}{N}\sum_{i=1}^{N}\frac{-A(X_{\mathbf{u},i})}{2r} - \frac{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)}\frac{-A(X_i)}{2r}e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}}{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)}e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}}$$

and

$$\frac{\partial Q(\theta,\theta_k)}{\partial\beta} = \frac{1}{N}\sum_{i=1}^{N}\frac{n(X_{\mathbf{u},i})}{\beta} - \frac{\sum_{i=1}^{N}\frac{n(X_i))\beta^{n(X_i))-1}}{\beta_0^{n(X_i)}}e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}}{\sum_{i=1}^{N}\left(\frac{\beta}{\beta_0}\right)^{n(X_i)}e^{\frac{-A(X_i)}{2r}(\eta-\eta_0)}}.$$

Performing a similar Newton optimisation process, we obtain the second derivatives of Q with respect to η and β, and

$$\eta_{k+1} = \eta_k + \frac{\frac{\partial Q(\theta,\theta_k)}{\partial \eta_k}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \beta_k^2} - \frac{\partial Q(\theta,\theta_k)}{\partial \beta_k}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \eta_k \partial \beta_k}}{\frac{\partial^2 Q(\theta,\theta_k)}{\partial \eta_k^2}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \beta_k^2} - \frac{\partial^2 Q(\theta,\theta_k)}{\partial \beta_k \partial \eta_k}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \eta_k \partial \beta_k}}$$

with

$$\beta_{k+1} = \beta_k + \frac{\frac{\partial Q(\theta,\theta_k)}{\partial \beta_k}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \eta_k^2} - \frac{\partial Q(\theta,\theta_k)}{\partial \eta_k}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \beta_k \partial \eta_k}}{\frac{\partial^2 Q(\theta,\theta_k)}{\partial \eta_k^2}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \beta_k^2} - \frac{\partial^2 Q(\theta,\theta_k)}{\partial \beta_k \partial \eta_k}\frac{\partial^2 Q(\theta,\theta_k)}{\partial \eta_k \partial \beta_k}}.$$

To simplify calculations and for scaling reasons, $\kappa = \log \beta$ may be used.

2.8 Discussion

In this chapter, a Bayesian inference framework for aoristic data was introduced in which an alternating renewal process is used to model the interval censoring of temporal data. A prospective point, which cannot be observed directly, was paired with an interval within which the point lies. A point process model based on this censoring mechanism was introduced. State estimation was then applied to best estimate the location of this point. Theory was developed regarding the distribution of these marks based on this renewal framework and the posterior distribution deduced. The fact that the forward model allows for a mixture of discrete and absolutely continuous components makes this process nontrivial. A state estimation procedure was outlined in the form of a Metropolis-Hastings algorithm for a fixed number of points, after which ergodicity properties were verified. Using an area-interaction prior, this procedure was applied to sample from the posterior distribution. Effects of the prior were clearly present when sampling from the complete model.

It may be limiting to assume that the length of the phase that a point is in is entirely independent of the time at which it is observed. In many applications, one might expect different times to be censored more often than others. The times at which the censoring changes from partial to full might also be time-dependent. In addition, the intensity of the prior point process may also vary depending on time. These issues are handled effectively by assuming that the temporal censoring mechanism is generated by a semi-Markov process. Mark distribution derivations, modelling approaches and parameter estimation methods for this alternative censoring mechanism are covered in Chapter 3.

CHAPTER 3

A non-homogeneous semi-Markov model for interval censoring

In Chapter 2, we developed methods for state estimation on interval-censored data. We assumed that a process exists whereby points of a marked point process fall in an interval geometry that is assumed to have been generated by an alternating renewal process censoring mechanism. Time is split up into observable and non-observable time, and the mechanism by which these phases change is known as the censoring mechanism. Either an event is fully observed, in which case the exact time of occurrence is recorded, or only a time interval is observed. A Bayesian approach was employed, with the point process being used as a prior distribution, the alternating renewal process as the forward model based on the observed interval data, and the times of occurrence within intervals as the posterior distribution. Markov chain Monte Carlo (MCMC) methods were used to estimate model parameters and times of occurrence.

One shortcoming of the alternating renewal process censoring mechanism is that it imposes time-homogeneity. In other words, the assumption is made that events are equally likely to happen at any time across the entire timespan within the interval. This has the advantage of being simpler mathematically, as one can always derive a time-independent form of the interval or mark distribution. However, this method ignores the reality that events may not occur homogeneously in time in many applications. There are times of day that may be more amenable to event occurrences due to factors such as periodicity, and not taking this into account leads to a potentially less accurate model.

This chapter introduces a non-homogeneous two-state semi-Markov process [72, 73, 81, 126], also known as a non-homogeneous alternating renewal process. Conditional intensity-based methods [59] are used to guarantee existence, and we derive the joint, marginal and conditional distributions of the recorded starting point and interval length for each occurrence time. We then propose a marked point process model [34] for the complete data using a non-homogeneous Markov point process [86] for the ground process of event occurrences and a mark kernel based on the non-homogeneous alternating renewal process. We illustrate the model by means of parametric examples that describe various types of non-homogeneous behaviour. Finally, we compare non-homogeneous and homogeneous models with simulated examples.

3.1 The non-homogeneous alternating renewal process

In the previous case, we looked at a sequence of random vectors $C_1, C_2, \ldots$ with $C_i = (Y_i, Z_i)$ representing an alternating renewal process. Then the length of phase i, which contains both a Y- and Z-phase, is represented by $T_i = Y_i + Z_i$. The time of the ith renewal is denoted by S_i. We will now redefine these concepts in a semi-Markov renewal context.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $i \in \mathbb{N}^+$. Define on this probability space the functions $S_i(\omega), \omega \in \Omega$, the ith state that the process is in, as well as the functions $X_i(\omega), \omega \in \Omega$ which have the property that $X_0(\omega) \leq X_1(\omega) \leq \dots$. We set $S(0) = 1$ and $X(0) = 0$. Since we are considering only two phases, $\Omega = \{0, 1\}$. Identify a Z-phase with state 0 and a Y-phase with state 1. For each $n \geq 0$, $\mathcal{A}$ contains the σ-algebra [25] generated by

$$\{S_m = \omega, X_m \leq x \,|\, m \in \{0, \dots, n\};\, \omega \in \Omega;\, x \in \mathbb{R}^+\}.$$

The tuple $(S_n, X_n)_{n=1}^{\infty}$ defines a non-homogeneous alternating renewal process, or two-state non-homogeneous semi-Markov process, if

$$\begin{aligned}
&\mathbb{P}(S_{n+1} = j, X_{n+1} \leq x \,|\, (S_0, X_0) = (s_0, x_0), \dots, (S_n, X_n) = (s_n, x_n)) \\
&\quad = \mathbb{P}(S_{n+1} = j, X_{n+1} \leq x \,|\, (S_n, X_n) = (s_n, x_n)) \\
&\quad = \mathbb{P}(S_{n+1} = j, X_{n+1} - X_n \leq x - x_n \,|\, (S_n, X_n) = (s_n, x_n)) \\
&\quad = \mathbb{P}(S_1 = j, X_1 - X_0 \leq x - x_n \,|\, (S_0, X_0) = (s_n, x_n)), \qquad (3.1)
\end{aligned}$$

i.e. the joint conditional probability of the sojourn time

$$T_{n+1} = X_{n+1} - X_n \qquad (3.2)$$

and the next state S_{n+1} depends only on the n-th state S_n and its jump time X_n, not on the entire history of the process [25, 73, 81, 126] nor on the index n. This process is only Markov at the jump times, hence the name semi-Markov.

To specify the distribution of the sojourn times, one need in general only define the semi-Markov kernel G, describing the transition rates from state i to state j. In our case, the process alternates - i.e. if $S_n = 0$, then $S_{n-1} = S_{n+1} = 1$. For a semi-Markov jump process, when X is a generic non-homogeneous jump process with T the sojourn times,

$$G_{ij}(x_n, \tau) = \mathbb{P}(S_{n+1} = j;\, T_{n+1} \leq \tau \,|\, S_n = i, X_n = x_n). \qquad (3.3)$$

Note that x_n is the nth jump time, and at time $x_{n+1} = x_n + T_{n+1}$, the next jump occurs. Since X is a semi-Markov process, one can safely ignore the history until X_n. As Ω only contains 2 states and the process thus alternates, we write $G_{10}(x_n, \tau) = G_Y(x_n, \tau)$ and $G_{01}(x_n, \tau) = G_Z(x_n, \tau)$, the subscript denoting the state that the process is in between jump times x_n and x_{n+1}.

Define the parametrised general form

$$G_T(x_n, \tau; \theta_{x_n}) = \begin{cases} G_Y(x_n, \tau) & \text{if } S_n = 1, \\ G_Z(x_n, \tau) & \text{otherwise.} \end{cases}$$

This form will be used when similar assumptions are made over the general form of the distribution of T, the sojourn times, for both Y- and Z-phases.

In the remainder of this chapter, we assume that, for all $x \geq 0$, $G_Y(x, \cdot)$ and $G_Z(x, \cdot)$ are absolutely continuous with respect to Lebesgue measure and write $g_Y(x, \cdot)$ and $g_Z(x, \cdot)$ respectively for their Radon–Nikodym derivatives.

3.2 Hazard rates and existence conditions

In this section, we consider the hazard rates of the sojourn times. Intuitively, the conditional (or stochastic) intensity of a temporal point process describes the infinitesimal conditional probability of occurrence given the history [77]. More precisely, for $n = 0, 1, \ldots$ and $0 = x_0 \leq x_1 \leq \cdots \leq x_n \leq x$,

$$\lambda_{n+1}(x; x_1, \ldots, x_n)\, dx = \mathbb{P}(X_{n+1} \leq x + dx | X_{n+1} \geq x, \\ X_0 = 0, X_1 = x_1, \ldots, X_n = x_n).$$

For processes with the Markov property, $\lambda_{n+1}(x; x_n) = \lambda_{n+1}(x; x_1, \ldots, x_n)$, as there is no dependence on previous values $x_1, \ldots, x_{n-1}$ of X_n. We will therefore use this notation in the sequel.

For a non-homogeneous alternating renewal process, the $\lambda_{n+1}(\cdot; \cdot)$ are closely related to the hazard rates of the sojourn times. To see this, recall that $S_0 = 1$ and assume that $n+1$ is odd. Then, using Equation 3.2, the conditional intensity of the jump process at time x given jumps at times $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n$ can be simplified as

$$\begin{aligned}
\lambda_{n+1}(x; x_n)\, dx &= \mathbb{P}(X_{n+1} \leq x + dx \,|\, X_{n+1} \geq x; X_n = x_n; S_n = 1) \\
&= \frac{\mathbb{P}(x \leq X_{n+1} \leq x + dx \,|\, X_n = x_n; S_n = 1)}{\mathbb{P}(X_{n+1} \geq x \,|\, X_n = x_n; S_n = 1)} \\
&= \frac{\mathbb{P}(x - x_n \leq T_{n+1} \leq x - x_n + dx \,|\, X_n = x_n; S_n = 1)}{\mathbb{P}(T_{n+1} \geq x - x_n \,|\, X_n = x_n; S_n = 1)} \\
&= \frac{\mathbb{P}(\tau \leq T_{n+1} \leq \tau + dx \,|\, X_n = x_n; S_n = 1)}{\mathbb{P}(T_{n+1} \geq \tau \,|\, X_n = x_n; S_n = 1)} \\
&= \frac{\mathbb{P}(\tau \leq T_{n+1} \leq \tau + dx \,|\, X_n = x_n; S_n = 1)}{1 - \mathbb{P}(T_{n+1} \leq \tau \,|\, X_n = x_n; S_n = 1)} \\
&= \frac{g_T(x_n, x - x_n) dx}{1 - G_T(x_n, x - x_n)} \qquad (3.4)
\end{aligned}$$

using the Markov property in step 1 and equation 3.3, and τ is as before. Note that this is exactly equivalent to the hazard rate function, which describes a time to failure in many statistical applications, and in this case can be interpreted as the limiting probability of a transition at time x. When $G_Y(x_n, x - x_n) = 1$, the conditional intensity is set to zero. For even $n+1$, a similar argument holds with g_Z and G_Z instead of g_Y and G_Y. Define

$$\Lambda_{n+1}(x_n + \tau; x_n) = \int_0^\tau \lambda_{n+1}(x_n + u; x_n)\, du$$

as the cumulative hazard rate. Then G_T can be rewritten as

$$G_T(x_n, \tau) = \int_0^\tau \exp[-\Lambda_{n+1}(x_n + u; x_n)]\, \lambda_{n+1}(x_n + u; x_n)\, du.$$

If the hazard rates can be estimated, so can the semi-Markov kernel G [104].

Conditions must also be established to guarantee existence. For example, one could imagine a situation in which there are infinitely many transitions in a finite time span. Haezendonck & De Vylder (1980) show that a process such as the one outlined in this chapter guarantees non-explosion [59]. We provide an outline of their reasoning below.

Proposition 3.1 (Haezendonck & De Vylder, 1980)**.** *Let X_n and X_n^* be two temporal point processes with corresponding conditional intensities λ and λ^*. If*

- *for every $n \in \mathbb{N}_0$, $\lambda_{n+1} \leq \lambda^*_{n+1}$;*
- *for every $n \in \mathbb{N}$, either $\lambda_{n+1}(x; x_n)$ or $\lambda^*_{n+1}(x; x_n)$ depends only on $x - x_n$,*

the probability of explosion at or before time x of the point process defined by λ is at most as big as that of the point process defined by λ^*. Under the same conditions, for all $n \in \mathbb{N}$ and $x \geq 0$,

$$\mathbb{P}(X_n \leq x) \leq \mathbb{P}(X_n^* \leq x).$$

Proof. See Corollaries 1,2 and 5 of [59]. □

Since non-explosion is guaranteed, we now proceed by showing existence for common families of sojourn time distributions.

Proposition 3.2. *Let $(S_n, X_n)_n$ be an alternating non-homogeneous semi-Markov process with values in $\{0, 1\} \times \mathbb{R}^+$ with $S_0 = 1$, $X_0 = 0$ and semi-Markov kernels $G_Y(x, \cdot)$, $G_Z(x, \cdot)$ that follow Gamma distributions with shape and rate parameters $\theta_Y(x) = (k_Y(x), \lambda_Y(x))$ and $\theta_Z(x) = (k_Z(x), \lambda_Z(x))$ in $[1, \infty) \times (0, \infty)$ such that, for all $x \in \mathbb{R}^+$,*

$$\lambda_Y(x) \leq c; \quad \lambda_Z(x) \leq c$$

for some $c > 0$. Write $X_\infty = \lim_{n\to\infty} X_n$ for the time of explosion. Then $\mathbb{P}(X_\infty < \infty) = 0$.

Proof. The probability density and cumulative distribution functions of the Gamma distribution with shape and rate parameters $k(x)$ and $\lambda(x)$ are, for $\tau \geq 0$,

$$g(x, \tau; k(x), \lambda(x)) = \frac{\lambda(x)^{k(x)} \tau^{k(x)-1} e^{-\lambda(x)\tau}}{\Gamma(k(x))}$$

and

$$G(x, \tau; k(x), \lambda(x)) = \frac{\gamma(k(x), \lambda(x)\tau)}{\Gamma(k(x))}$$

for Γ the gamma function and γ the lower incomplete gamma function. The conditional intensity is

$$\begin{aligned}\lambda_{n+1}(x; x_n) &= \frac{g_T(x_n, x - x_n; k_T(x_n), \lambda_T(x_n))}{1 - G_T(x_n, x - x_n; k_T(x_n), \lambda_T(x_n))} \\ &= \frac{\lambda_T(x_n)^{k_T(x_n)} (x - x_n)^{k_T(x_n)-1} e^{-\lambda_T(x_n)(x-x_n)}}{\int_{\lambda_T(x_n)(x-x_n)}^{\infty} u^{k_T(x_n)-1} e^{-u}\, du},\end{aligned}$$

where g_T is either g_Y or g_Z. We examine the limiting behaviour as $x \to \infty$. See that

$$\begin{aligned}\lim_{(x-x_n)\to\infty} g_T(x_n, x - x_n; k_T(x_n), \lambda_T(x_n)) = 0, \\ \lim_{(x-x_n)\to\infty} 1 - G_T(x_n, x - x_n; k_T(x_n)\lambda_T(x_n)) = 0,\end{aligned}$$

and both numerator and denominator are differentiable on $(0, \infty)$. By L'Hôpital's rule, we obtain

$$\begin{aligned}&\lim_{(x-x_n)\to\infty} \lambda_{n+1}(x; x_n) = \\ &\lim_{(x-x_n)\to\infty} \frac{\lambda_T(x_n)^{k_T(x_n)} (x - x_n)^{k_T(x_n)-1} e^{-\lambda_T(x_n)(x-x_n)}}{\int_{\lambda_T(x_n)(x-x_n)}^{\infty} u^{k_T(x_n)-1} e^{-u}\, du} \\ &= \lim_{(x-x_n)\to\infty} \frac{\frac{d}{d(x-x_n)} (\lambda_T(x_n)^{k_T(x_n)} (x - x_n)^{k_T(x_n)-1} e^{-\lambda_T(x_n)(x-x_n)})}{\frac{d}{d(x-x_n)} (\int_{\lambda_T(x_n)(x-x_n)}^{\infty} u^{k_T(x_n)-1} e^{-u}\, du)} \\ &= \lim_{(x-x_n)\to\infty} \frac{\lambda_T(x_n)(x - x_n) - (k_T(x_n) - 1)}{(x - x_n)} \\ &= \lambda_T(x_n)\end{aligned}$$

after simplifying. To show that the limit is monotonically reached, we have to show that $\lambda_{n+1}(x; x_n)$ is increasing. Let $t = \lambda_T(x_n)(x - x_n)$ and introduce the function $h(t)$ such that

$$h(t) = \frac{t^{k_T(x_n)-1} e^{-t}}{\int_t^{\infty} u^{k_T(x_n)-1} e^{-u}\, du}$$

and $\lambda_{n+1}(x; x_n) = \lambda_T(x_n)h(t)$. If $\log \lambda_{n+1}(x; x_n)$ is increasing then $\lambda_{n+1}(x; x_n)$ is also increasing. Therefore, it suffices to show that the function $t \to \log h(t)$ is non-decreasing in $t > 0$. Now,

$$\frac{\partial}{\partial t} \log h(t) = \frac{k_T(x_n) - 1}{t} - 1 + \frac{t^{k_T(x_n)-1} e^{-t}}{\int_t^\infty u^{k_T(x_n)-1} e^{-u}\, du}.$$

If $t < k_T(x_n) - 1$, we see directly that the derivative is positive. Otherwise, use integration by parts to simplify the last term in the right-hand side to

$$1 - \int_t^\infty \frac{k_T(x_n) - 1}{u} u^{k_T(x_n)-1} e^{-u}\, du \Big/ \int_t^\infty u^{k_T(x_n)-1} e^{-u}\, du.$$

Consequently

$$\frac{\partial}{\partial t} \log h(t) = \frac{\displaystyle\int_t^\infty \left\{ \frac{k_T(x_n) - 1}{t} - \frac{k_T(x_n) - 1}{u} \right\} u^{k_T(x_n)-1} e^{-u}\, du}{\displaystyle\int_t^\infty u^{k_T(x_n)-1} e^{-u}\, du}$$

is non-negative. We conclude that $\lambda_{n+1}(x; x_n)$ is bounded by $\lambda_T(x_n)$ for all $k_T(x_n) \geq 1$.

Under the assumption that $\sup_{x \in \mathbb{R}^+} \max(\lambda_Y(x), \lambda_Z(x)) \leq c$, we can construct a Poisson process N^* with conditional intensity $\lambda^*_{n+1}(x; x_1, \ldots, x_n) = c$. Clearly, λ^* satisfies the second condition of [59, Corollary 2]. Moreover, a Poisson process with constant intensity has probability zero to explode. We conclude that $\mathbb{P}(X_\infty < \infty) = 0$. □

Important special cases include $k_T(x) = 1$ corresponding to exponential distributions, or $k_T(x) \in \mathbb{N}$ corresponding to Erlang distributed phases.

Proposition 3.3. *Let $(S_n, X_n)_n$ be an alternating non-homogeneous semi-Markov process with values in $\{0, 1\} \times \mathbb{R}^+$ with $S_0 = 1$, $X_0 = 0$ and semi-Markov kernels $G_Y(x, \cdot)$, $G_Z(x, \cdot)$ that follow Weibull distributions with shape and rate parameters $\theta_Y(x) = (k_Y(x), \lambda_Y(x))$ and $\theta_Z(x) = (k_Z(x), \lambda_Z(x))$ in $(0, \infty) \times (0, \infty)$ such that $\lambda_Y(x) \leq c$, $\lambda_Z(x) \leq c$ for some $c > 0$, and $k_Y(x) \leq k$, $k_Z(x) \leq k$ for some $k > 0$. Write $X_\infty = \lim_{n \to \infty} X_n$ for the time of explosion. Then $\mathbb{P}(X_\infty < \infty) = 0$.*

Proof. Let $G_T(x, \cdot)$ and corresponding $(\lambda_T(x), k_T(x))$ correspond to either Y- or Z-phase cases. The probability density and cumulative distribution functions of the Weibull distribution with shape and rate parameters $k(x)$ and $\lambda(x)$ are, for $\tau \geq 0$,

$$g(x, \tau; k(x), \lambda(x)) \quad = \quad \lambda(x)k(x) \left(\lambda(x)\tau\right)^{k(x)-1} e^{-(\lambda(x)\tau)^{k(x)}};$$

$$G(x, \tau; k(x)\,\lambda(x)) \quad = \quad 1 - e^{-(\lambda(x)\tau)^{k(x)}}.$$

The conditional intensity of $(X_n)_n$ is therefore

$$\lambda_{n+1}(x; x_n) = k_T(x_n)\lambda_T(x_n)\left(\lambda_T(x_n)(x - x_n)\right)^{k_T(x_n)-1}.$$

Since the hazard rate is unbounded, we cannot use a Poisson process to bound the Weibull hazard. Instead we turn to a homogeneous renewal process X^* with sojourn times that are Weibull distributed with shape parameter k and rate parameter c. By the strong law of large numbers, since the expected sojourn times are strictly positive, X^* has explosion probability zero [126, Section 3.1]. Also,

$$\lambda_{n+1}(x; x_n) \leq \lambda^*_{n+1}(x; x_n) = kc^k(x - x_n)^{k-1}$$

and both conditional intensities are a function of $x - x_n$ only. By [59, Corollary 2], $\mathbb{P}(X_\infty < \infty) = 0$. □

The case that $k = 1$ corresponds to exponential sojourn times.

3.3 Renewal function: existence and boundedness

The counting process measuring the amount of cycles having occurred by time t can be written as

$$N(t) = \sup\left\{n \in \mathbb{N}_0 : \sum_{i=1}^{2n} T_i \leq t\right\}, \tag{3.5}$$

where a cycle is an interval of time within which each state occurs once. Because of this definition of N, we always have that $S_{2n} = 1$ and $S_{2n+1} = 0$ for any $n \in \mathbb{N}_0$ such that $N(t) = n$. The distribution of X_{2n}, the jump times of a cycle, is

$$F_{2n}(t) = \mathbb{P}\left(\sum_{i=1}^{2n} T_i \leq t\right) = \mathbb{P}(X_{2n} \leq t) = \mathbb{P}(N(t) \geq n).$$

The (Markov) renewal function is defined, analogously to the standard renewal process, as $M(t) = \mathbb{E}N(t), t \geq 0$ [73]. In our case,

$$M(t) = \mathbb{E}N(t) = \sum_{n=1}^{\infty} \mathbb{P}(N(t) \geq 2n) = \sum_{n=1}^{\infty} \mathbb{P}(X_{2n} \leq t) = \sum_{n=1}^{\infty} F_{2n}(t),$$

a $2n$-fold convolution, recalling that $N(t) \geq 2n \iff X_{2n} \leq t$.

Corollary 3.4. *The Markov renewal function $M(t)$ for a non-homogeneous jump process $(X_i)_{i=0}^{\infty}$ is bounded by the expectation of a constant rate Poisson process $N^*(t)$ if the sojourn times $(T_i)_{i=0}^{\infty}$ are exponentially or Gamma distributed.*

Proof. Take $(X_i)_{i=0}^{\infty}$, the jump process with corresponding conditional intensity $\lambda_{2n+1}(x_{n+1}; x_n)$ that is bounded by the (finite) conditional intensity of another stochastic process $\lambda^*_{2n+1}(x_{n+1}; x_n)$ (as shown in Chapter 3.2). We bound with a Poisson process, which we call $N^*(t)$, with rate c. Therefore

$$N^*(t) = \sup\left\{ n \in \mathbb{N}_0 : \sum_{i=1}^{2n} T_i^* \leq t \right\} \tag{3.6}$$

where $T_i^* \sim \text{Exp}(c)$. By Corollary 1 of [59], we have

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_{2n} \leq x_{2n}) \leq \mathbb{P}(X_1^* \leq x_1, \ldots, X_{2n}^* \leq x_{2n}),$$

where $x_1, \ldots, x_{2n}$ are in the support of the random variables $X_1, \ldots, X_{2n}$ and $X_1^*, \ldots, X_{2n}^*$. Note that due to the non-explosion property, no random variables can be defective (have mass at infinity). Take $x_1, \ldots, x_{2n-1} \to \infty$, so that we are left with the marginal cumulative distribution functions

$$\mathbb{P}(X_{2n} \leq x_{2n}) = \mathbb{P}(X_1 \leq \infty, \ldots, X_{2n-1} \leq \infty, X_{2n} \leq x_{2n}).$$

and

$$\mathbb{P}(X_{2n}^* \leq x_{2n}) = \mathbb{P}(X_1^* \leq \infty, \ldots, X_{2n-1}^* \leq \infty, X_{2n}^* \leq x_{2n}),$$

hence $\mathbb{P}(X_{2n} \leq x_{2n}) \leq \mathbb{P}(X_{2n}^* \leq x_{2n})$. Therefore, using the fact that $N^*(t)$ (see equation 3.6) is a Poisson process,

$$\begin{aligned}
\mathbb{E}N(t) &= \sum_{n=1}^{\infty} \mathbb{P}(X_{2n} \leq t) \\
&\leq \sum_{n=1}^{\infty} \mathbb{P}(X_{2n}^* \leq t) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(N^*(t) \geq 2n) = \mathbb{E}N^*(t) = ct,
\end{aligned}$$

hence $M(t)$ is bounded. □

Corollary 3.5. *The Markov renewal function $M(t)$ for a non-homogeneous jump process $(X_i)_{i=0}^{\infty}$ is bounded by the expectation of a renewal process $N^*(t)$ if the sojourn times $(T_i)_{i=0}^{\infty}$ are Weibull distributed.*

Proof. We again have that

$$N^*(t) = \sup\left\{ n \in \mathbb{N}_0 : \sum_{i=1}^{2n} T_i^* \leq t \right\},$$

however now $N^*(t)$ is no longer a Poisson process but a renewal process. Use the exact same method of proof as in proposition 3.4, since we know that $\lambda_{2n+1}(x_{n+1}; x_n) \leq \lambda^*_{2n+1}(x_{n+1}; x_n)$ under certain conditions (see proposition 3.3). Take the random variables $X_1, \ldots, X_{2n}$ and $X^*_1, \ldots, X^*_{2n}$, taking values $x_1, \ldots, x_{2n}$. Then $\mathbb{P}(X_{2n} \leq x_{2n}) \leq \mathbb{P}(X^*_{2n} \leq x_{2n})$, and since it is known that $\mathbb{E}(N^*(t)) < \infty$ (see [6, 126]), we have $\mathbb{E}(N(t)) \leq \mathbb{E}(N^*(t)) < \infty$. □

3.3.1 Renewal function derivations

Let $F(t)$ be the distribution function of the jump times. For homogeneous semi-Markov processes, we have

$$\begin{aligned} M(t) = \mathbb{E}[N(t)] &= \int_0^\infty \mathbb{E}[N(t) \,|\, X_2 = s] \, dF(s) \\ &= \int_0^t \mathbb{E}[N(t) \,|\, X_2 = s] \, dF(s) + \int_t^\infty \mathbb{E}[N(t) \,|\, X_2 = s] \, dF(s). \end{aligned}$$

If $s > t$, then $\mathbb{E}[N(t) \,|\, X_2 = s] = 0$. This is because if s occurs after t, a renewal cannot have occurred. On the other hand, if $0 \leq s \leq t$, $\mathbb{E}[N(t) \,|\, X_2 = s] = 1 + \mathbb{E}[N(t-s)]$. Because s occurs between 0 and t, 1 renewal has definitely occurred by time s, and in the remaining time $t - s$, there are $\mathbb{E}[N(t-s)]$ more renewals. Therefore

$$\begin{aligned} M(t) &= \int_0^t \mathbb{E}[N(t) \,|\, X_2 = s] \, dF(s) \\ &= \int_0^t (1 + \mathbb{E}[N(t-s)]) \, dF(s) \\ &= F(t) + \int_0^t M(t-s) \, dF(s) = F(t) + (M * F)(t), \end{aligned}$$

where * denotes a convolution.

Recall that $M(t) = \sum_{n=1}^\infty F_{2n}(t)$. In the non-homogeneous case, each function $F_{2n}(t)$ is no longer conditionally independent of the function $F_{2n-2}(t)$ that described the previous Y-phase. Therefore, one cannot simply take the Lebesgue-Stieltjes integral over a generic distribution F for all phases. So,

$$M(t) = \sum_{n=1}^\infty \mathbb{P}(X_{2n} \leq t) = \mathbb{P}(X_2 \leq t) + \mathbb{P}(X_4 \leq t) + \mathbb{P}(X_6 \leq t) \ldots$$

To discover just how difficult it is to come up with explicit expressions, take the first renewal event $\{X_2 \leq t\}$. Assume that $T_1 \sim \text{Exp}(\lambda_0(x))$ and $T_2 \sim \text{Exp}(\lambda_1(x; x_0))$. We have

$$F_2(t) = \mathbb{P}(X_2 \leq t) = \mathbb{P}(T_1 + T_2 \leq t)$$

$$
\begin{aligned}
&= \int_0^t \mathbb{P}(T_1 + T_2 \leq t \,|\, T_1 = \tau)\mathbb{P}(T_1 = \tau)\, d\tau \\
&= \int_0^t \mathbb{P}(T_2 \leq t - \tau)\lambda_{x_0} e^{-\lambda_{x_0}\tau}\, d\tau \\
&= \int_0^t [1 - e^{-\lambda_{x_1}(t-\tau)}]\lambda_{x_0} e^{-\lambda_{x_0}\tau}\, d\tau \\
&= \int_0^t \left(\lambda_{x_0} e^{-\lambda_{x_0}\tau} - \lambda_{x_0} e^{-\lambda_{x_0}\tau - \lambda_{x_1}(t-\tau)}\right) d\tau \\
&= \int_0^t \left(\lambda_{x_0} e^{-\lambda_{x_0}\tau} - \lambda_{x_0} e^{-\tau(\lambda_{x_0} - \lambda_{x_1}) - t\lambda_{x_1}}\right) d\tau \\
&= [-e^{-\lambda_{x_0}\tau}]_0^t + \frac{\lambda_{x_0}}{\lambda_{x_0} + \lambda_{x_1}}[e^{-\tau(\lambda_{x_0} - \lambda_{x_1}) - t\lambda_{x_1}}]_0^t \\
&= 1 - e^{-\lambda_{x_0} t} + \frac{\lambda_{x_0}}{\lambda_{x_0} + \lambda_{x_1}}(e^{-\lambda_{x_0} t} - e^{\lambda_{x_1} t}) \\
&= 1 - \frac{\lambda_{x_1}}{\lambda_{x_0} + \lambda_{x_1}} e^{-\lambda_{x_0} t} - \frac{\lambda_{x_0}}{\lambda_{x_0} + \lambda_{x_1}} e^{\lambda_{x_1} t}.
\end{aligned}
$$

and

$$
\begin{aligned}
f_2(t) &= \frac{d}{dt}(\mathbb{P}(X_2 \leq t)) \\
&= \frac{d}{dt}\left(1 - \frac{\lambda_{x_1}}{\lambda_{x_0} + \lambda_{x_1}} e^{-\lambda_{x_0} t} - \frac{\lambda_{x_0}}{\lambda_{x_0} + \lambda_{x_1}} e^{-\lambda_{x_1} t}\right) \\
&= \frac{\lambda_{x_0}\lambda_{x_1}}{\lambda_{x_0} + \lambda_{x_1}}(e^{-\lambda_{x_0} t} + e^{-\lambda_{x_1} t}).
\end{aligned}
$$

For higher orders, this quickly becomes infeasible. Therefore, it is intractable to write the renewal function in terms of the sum of its CDFs, and a modelling approach is required.

3.4 Age and excess

Now that the theoretical groundwork for the censoring mechanism has been laid, we proceed by determining the joint distribution of age and excess. The age $A(t)$ is the time elapsed since the last phase change, and $B(t)$, the excess, is the time remaining until the next phase change. For all t where the process is in state 0, or the Z-phase, we assume that the occurrence time can be observed perfectly. Therefore we only consider age and excess with respect to state 1, or the Y-phase. This also provides us with a concrete censoring mechanism: depending on the state within which a point falls, we note either the exact occurrence time, or the interval corresponding to the age and excess functions. Obtaining their joint distribution allows us to specify the likelihood of intervals based on their starting

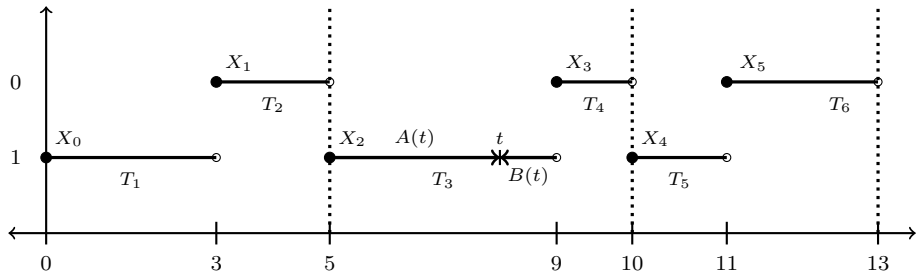


Figure 3.1 A visualisation of a semi-Markov process with initial values $S_0 = 1; X_0 = 0$. At the dotted line, one cycle has passed - i.e. the process has taken both possible state values. The jump times correspond to a change of state. For a given time t in which the process is in state 1, a non-zero age $A(t)$ and excess $B(t)$ are recorded.

point and length in terms of the semi-Markov kernel G_Y. To see a visualisation of a semi-Markov example with age $A(t)$ and excess $B(t)$, see Figure 3.1.

Proposition 3.6. *Consider a non-homogeneous alternating renewal process $(S_n, X_n)_{n=1}^{\infty}$ with values in $\{0,1\} \times \mathbb{R}^+$ with $S_0 = 1$, $X_0 = 0$, semi-Markov kernels G_Y and G_Z and associated counting measure $N(t)$, $t \geq 0$. Let the age process with respect to the Y-phase be*

$$A(t) = (t - X_{2N(t)})\, \mathbf{1}\{X_{2N(t)+1} > t\}$$

and define the excess with respect to the Y-phase as

$$B(t) = (X_{2N(t)+1} - t)\, \mathbf{1}\{X_{2N(t)+1} > t\},$$

where $X_{2N(t)}$ is the jump time immediately after $N(t)$ cycles have been completed. Then, for $t \geq 0$ and $0 \leq x \leq t$ and $z \geq 0$,

$$\mathbb{P}(A(t) \leq x, B(t) \leq z) = G_Y(0,t) - \int_{t-x}^{t} [1 - G_Y(s, t+z-s)]\, dM(s)$$
$$- \int_{0}^{t-x} [1 - G_Y(s, t-s)]\, dM(s) + \mathbf{1}\{x = t\}[G_Y(0, t+z) - G_Y(0,t)]. \quad (3.7)$$

Proof. Assume that $X_0 = 0$ and $S_0 = 1$. For the age,

$$\mathbb{P}(A(t) > x) = \mathbb{P}(t - X_{2N(t)} > x, X_{2N(t)+1} > t \,|\, S_0 = 1, X_0 = 0)$$

where $0 \leq x < t$. In this case, since $A(t) > 0$, we must be in state 1. Now

$$\mathbb{P}(A(t) > x) = \mathbb{P}(t - X_{2N(t)} > x;\, X_{2N(t)+1} > t \,|\, S_0 = 1, X_0 = 0)$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \mathbb{P}(t - X_{2n} > x;\, X_{2n+1} > t;\, N(t) = n \,|\, S_0 = 1, X_0 = 0) \\
&= \mathbb{P}(X_1 > t \,|\, S_0 = 1, X_0 = 0) \\
&+ \sum_{n=1}^{\infty} \mathbb{P}(t - X_{2n} > x;\, X_{2n+1} > t;\, N(t) = n \,|\, S_0 = 1, X_0 = 0) \\
&= \mathbb{P}(T_1 > t \,|\, S_0 = 1, X_0 = 0) \\
&+ \sum_{n=1}^{\infty} \mathbb{P}(t - X_{2n} > x;\, X_{2n+1} > t;\, X_{2n} \leq t;\, X_{2(n+1)} > t \,|\, S_0 = 1, X_0 = 0) \\
&= 1 - \mathbb{P}(T_1 \leq t \,|\, S_0 = 1, X_0 = 0) \\
&+ \sum_{n=1}^{\infty} \mathbb{P}(t - X_{2n} > x;\, X_{2n+1} > t \,|\, S_0 = 1, X_0 = 0)
\end{aligned}
$$

after simplifying and removing redundant conditions. Note that by Equation 3.3 and the fact that we know we are guaranteed to be in state 1, $\mathbb{P}(T_1 \leq t \,|\, S_0 = 1, X_0 = 0) = G_{10}(0,t)$. Since we only care about transitions from state 1 to state 0, write $G_{10}(t,\tau) = G_Y(t,\tau)$. Continuing,

$$
\begin{aligned}
\mathbb{P}(A(t) > x) &= 1 - G_Y(0,t) + \sum_{n=1}^{\infty} \mathbb{P}(t - X_{2n} > x;\, X_{2n+1} > t \,|\, S_0 = 1, X_0 = 0) \\
&= 1 - G_Y(0,t) + \sum_{n=1}^{\infty} \int_0^{t-x} \mathbb{P}(X_{2n+1} > t \,|\, S_{2n} = 1, X_{2n} = s)\, dF_{2n}(s) \\
&= 1 - G_Y(0,t) + \sum_{n=1}^{\infty} \int_0^{t-x} \mathbb{P}(T_{2n+1} > t - s \,|\, S_{2n} = 1, X_{2n} = s)\, dF_{2n}(s) \\
&= 1 - G_Y(0,t) + \sum_{n=1}^{\infty} \int_0^{t-x} [1 - \mathbb{P}(T_{2n+1} \leq t - s \,|\, S_{2n} = 1, X_{2n} = s)]\, dF_{2n}(s) \\
&= 1 - G_Y(0,t) + \sum_{n=1}^{\infty} \int_0^{t-x} [1 - G_Y(s, t-s)]\, dF_{2n}(s) \\
&= 1 - G_Y(0,t) + \int_0^{t-x} [1 - G_Y(s, t-s)]\, dM(s),
\end{aligned}
$$

using the law of total probability and Fubini's theorem. Considering the discrete components,

$$
\mathbb{P}(A(t) = 0) = 1 - P(A(t) > 0) = G_Y(0,t) - \int_0^t [1 - G_Y(s, t-s)]\, dM(s)
$$

and

$$
\mathbb{P}(A(t) = t) = 1 - G_Y(0,t)
$$

analogously. Taking the complement probability $\mathbb{P}(A(t) \leq x)$,

$$\begin{aligned}\mathbb{P}(A(t) \leq x) &= 1 - \mathbb{P}(A(t) > x) \\ &= G_Y(0,t) - \int_0^{t-x} [1 - G_Y(s, t-s)]\, dM(s).\end{aligned}$$

For the excess,

$$\mathbb{P}(B(t) > z) = \mathbb{P}(X_{2N(t)+1} - t > z, X_{2N(t)+1} > t \,|\, S_0 = 1, X_0 = 0)$$

where $z > 0$. As t and z are both positive, $\max(t, t+z) = t + z$ and we can simplify this event and write, for $z > 0$,

$$\begin{aligned}\mathbb{P}(B(t) > z) &= \mathbb{P}(X_{2N(t)+1} > t + z \,|\, S_0 = 1, X_0 = 0) \\ &= \sum_{n=0}^{\infty} \mathbb{P}(X_{2n+1} > t + z; N(t) = n \,|\, S_0 = 1, X_0 = 0) \\ &= \mathbb{P}(X_1 > t + z \,|\, S_0 = 1, X_0 = 0) \\ &+ \sum_{n=1}^{\infty} \mathbb{P}(X_{2n+1} > t + z; N(t) = n \,|\, S_0 = 1, X_0 = 0) \\ &= 1 - G_Y(0, t+z) + \sum_{n=1}^{\infty} \mathbb{P}(X_{2n+1} > t + z;\, X_{2n} \leq t; \,|\, S_0 = 1, X_0 = 0) \\ &= 1 - G_Y(0, t+z) \\ &+ \sum_{n=1}^{\infty} \int_0^t \mathbb{P}(X_{2n+1} > t + z; \,|\, S_0 = 1, X_0 = 0, X_{2n} = s)\, dF_{2n}(s) \\ &= 1 - G_Y(0, t+z) \\ &+ \sum_{n=1}^{\infty} \int_0^t [1 - \mathbb{P}(T_{2n+1} \leq t + z - s; \,|\, S_{2n} = 1, X_{2n} = s)]\, dF_{2n}(s) \\ &= 1 - G_Y(0, t+z) + \sum_{n=1}^{\infty} \int_0^t [1 - G_Y(s, t+z-s)]\, dF_{2n}(s) \\ &= 1 - G_Y(0, t+z) + \int_0^t [1 - G_Y(s, t+z-s)]\, dM(s)\end{aligned}$$

and

$$\mathbb{P}(B(t) = 0) = 1 - \mathbb{P}(B(t) > 0) = G_Y(0,t) - \int_0^t [1 - G_Y(s, t-s)]\, dM(s).$$

Taking the complement,

$$\mathbb{P}(B(t) \leq z) = 1 - \mathbb{P}(B(t) > z) = G_Y(0, t+z) - \int_0^t [1 - G_Y(s, t+z-s)]\, dM(s).$$

Going through the same steps as before, we can find an expression for $\mathbb{P}(B(t) > z; A(t) > x)$, the joint distribution of age and excess. When $z \in [0, \infty)$ and $x \in [0, t)$, as before,

$$\begin{aligned}
&\mathbb{P}(B(t) > z; A(t) > x) = \mathbb{P}(X_{2N(t)+1} > t + z, t - X_{2N(t)} > x \,|\, S_0 = 1, X_0 = 0) \\
&= 1 - G_Y(0, t + z) \\
&\qquad + \sum_{n=1}^{\infty} \mathbb{P}(X_{2n+1} > t + z; t - X_{2n} > x; N(t) = n \,|\, S_0 = 1, X_0 = 0) \\
&= 1 - G_Y(0, t + z) + \sum_{n=1}^{\infty} \int_0^{t-x} \mathbb{P}(X_{2n+1} > t + z \,|\, S_{2n} = 1, X_{2n} = s)\, dF_{2n}(s) \\
&= 1 - G_Y(0, t + z) + \sum_{n=1}^{\infty} \int_0^{t-x} \left[1 - G_Y(s, t + z - s)\right]\, dF_{2n}(s) \\
&= 1 - G_Y(0, t + z) + \int_0^{t-x} \left[1 - G_Y(s, t + z - s)\right]\, dM(s).
\end{aligned}$$

We can now handle the event $\{B(t) \leq z; A(t) \leq x\}$ for $0 \leq x \leq t$, $z \geq 0$.

$$\begin{aligned}
&\mathbb{P}(B(t) \leq z; A(t) \leq x) = \mathbb{P}(B(t) > z; A(t) > x) + \mathbb{P}(B(t) \leq z) + \mathbb{P}(A(t) \leq x) - 1 \\
&= G_Y(0, t) - \int_{t-x}^{t} \left[1 - G_Y(s, t + z - s)\right] dM(s) - \int_0^{t-x} \left[1 - G_Y(s, t - s)\right] dM(s) \\
&+ \mathbf{1}\{x = t\}[G_Y(0, t + z)]. \qquad (3.8)
\end{aligned}$$

□

From Proposition 3.6 we conclude that the probability that time $t \geq 0$ falls in a Z-phase is given by

$$w_t = \mathbb{P}(A(t) \leq 0, B(t) \leq 0) = G_Y(0, t) - \int_0^t \left[1 - G_Y(s, t - s)\right] dM(s). \qquad (3.9)$$

This case constitutes the atomic part of Equation 3.7. The singular component on the line $x = t$ has total mass $1 - G_Y(0, t)$ and represents the case that t falls before the first jump of the alternating renewal process.

The absolutely continuous component of Equation 3.7 can be written as

$$\int_0^x \int_0^z g_Y(t - u, u + v) m(t - u) du dv$$

provided that the Radon–Nikodym derivatives m of M and g_Y of G_Y exist. Recall that in our proposed censoring mechanism, when t falls in a Y-phase, the entire interval $[t - A(t), t + B(t)]$ is reported, which may be parametrised by the

left-most point $t - A(t)$ and length $A(t) + B(t)$. Suppose that $A(t) = u$, $B(t) = v$, and apply the change of variables $a = t - u$ and $l = u + v$. We find that the joint probability density function of left-most point and length is

$$q_t(a, l) = \frac{m(a) g_Y(a, l)}{\int_0^t [1 - G_Y(s, t - s)] \, dM(s)} \mathbf{1}\{0 \leq a \leq t \leq a + l; l \geq 0\} \tag{3.10}$$

upon scaling.

Proposition 3.7. *Let g_Y and m be as before, and let (A, L) be distributed according to $q_t(a, l)$ given by Equation 3.10. Then the marginal probability density function of A at $a \in [0, t]$ is*

$$f_t(a) = \frac{m(a)[1 - G_Y(a, t - a)]}{\int_0^t [1 - G_Y(s, t - s)] \, dM(s)} \tag{3.11}$$

and the conditional probability density function of L given $A = a$ is, for $l \in [t - a, \infty)$,

$$f_{t,\, L \,|A=a}(l) = \frac{g_Y(a, l)}{1 - G_Y(a, t - a)}. \tag{3.12}$$

Proof. Assume that $0 \leq a \leq t \leq a + l$ and $l \geq 0$. The marginal distribution of the starting time $f_t(a)$ is

$$\begin{aligned} f_t(a) &= \int q_t(a, l) \, dl = \frac{m(a)}{\int_0^t [1 - G_Y(s, t - s)] \, dM(s)} \int_{t-a}^{\infty} g_Y(a, l) \, dl \\ &= \frac{m(a)[1 - G_Y(a, t - a)]}{\int_0^t [1 - G_Y(s, t - s)] \, dM(s)} \end{aligned}$$

and

$$f_{t,\, L \,|A=a}(l) = \frac{q_t(a, l)}{f_a(a)} = \frac{g_Y(a, l)}{1 - G_Y(a, t - a)}.$$

□

3.4.1 Model formulation

The ensemble of potentially censored occurrence times can be mathematically formalised as a marked point process [34]. The ground process of points represents the uncensored event occurrences, which we model by a Markov point process [86] defined by a probability density with respect to a unit rate Poisson process. Temporal variations can be taken into account, as well as interactions between the points. Each point is subsequently marked, independently of other

points, either by an atom at the point when it is observed perfectly, or by the interval in which the point lies in case of censoring. The mark kernel that governs the random censoring is based on the distribution of age and excess in a non-homogeneous semi-Markov process.

Formally, let $\mathcal{X}$ be an open set on the real line. The state space $\mathcal{N}_\mathcal{X}$ of a simple point process X consists of finite sets $\{x_1, x_2, \ldots, x_n\} \subset \mathcal{X}$, $n \in \mathbb{N}_0$, which we equip with the Borel σ-algebra of the weak topology [34, Appendix A2]. Let p be a measurable, non-negative function on $\mathcal{X}$ that integrates to unity and $\sim$ a symmetric, reflexive relation on $\mathcal{X}$. A point process X on $\mathcal{X}$ having probability density p with respect to a unit rate Poisson process is Markov with respect to $\sim$ if, firstly, p is hereditary, that is, $p(\mathbf{x}) > 0$ implies that $p(\mathbf{y}) > 0$ for all subsets $\mathbf{y}$ of $\mathbf{x}$, and, secondly, the conditional intensity, defined as $p(\mathbf{x} \cup \{t\})/p(\mathbf{x})$ with $a/0 = 0$ for $a \geq 0$, depends only on the neighbourhood $\{x \in \mathbf{x} : x \sim t\}$ of t in $\mathbf{x}$ for every $t \in \mathcal{X} \setminus \mathbf{x}$ and every $\mathbf{x} = \{x_1, \ldots, x_n\} \subset \mathcal{X}$ for which $p(\mathbf{x}) > 0$ [86, 123].

An interaction function is a family $\phi_0, \phi_1, \phi_2, \ldots$ of non-negative functions ϕ_i defined on configurations of i points that take the value one whenever the configuration contains a pair $\{x_1, x_2\}$ of unrelated points, that is, $x_1 \not\sim x_2$. By the Hammersley–Clifford theorem [123], writing $|\cdot|$ for cardinality, a Markov density p can be factorised as

$$p(\mathbf{x}) = \prod_{\mathbf{y} \subset \mathbf{x}} \phi_{|\mathbf{y}|}(\mathbf{y}) \tag{3.13}$$

for some interaction function ϕ_i. The function $\phi_1(x)$ can be used to model temporal variations in the likelihood of events occurring. Higher order terms $\phi_2, \phi_3, \ldots$ govern interactions between pairs, triples or tuples of points.

The points x in a realisation $\mathbf{x}$ of X are marked independently according to a mark kernel $\nu(\cdot|x)$ on $\mathbb{R} \times \mathbb{R}^+$. A mark (a, l) represents an interval $[a, a + l]$ that starts at a and has length l. The mark kernel ν formalises the semi-Markov censoring discussed in Chapter 3.1. For demonstrative purposes, we assumed a starting time of 0, which we now set to $-\infty$. Doing so also allows us to ignore the singular component. Hence, the appropriate time-dependent mark kernel $\nu(\cdot|x)$, $x \in \mathcal{X}$, for a Borel subset $A \subset \mathbb{R} \times \mathbb{R}^+$ is

$$\begin{aligned} \nu(A|x) = &\left(1 - \int_{-\infty}^{x} \left[1 - G_Y(s, x - s)\right] \, dM(s)\right) \delta(\{(x, 0)\} \cap A) \\ &+ \int_{-\infty}^{x} \int_{x-a}^{\infty} \mathbf{1}\{(a, l) \in A\} \, G_Y(a, dl) \, dM(a). \end{aligned} \tag{3.14}$$

Write W for the marked point process defined by $p(\cdot)$ and $\nu(\cdot|\cdot)$ [34, Prop. 4.IV]. A realisation $\mathbf{w}$ is of the form

$$\mathbf{w} = \{w_1, w_2, \ldots, w_n\} = \{(x_1, (a_1, l_1)), (x_2, (a_2, l_2)), \ldots, (x_n, (a_n, l_n))\}$$

for $a_i \leq x_i \leq a_i + l_i$ for all $i = 1, 2, \ldots, n$. We denote the set of realisations by $\mathcal{N}_{\mathcal{X}\times(\mathbb{R}\times\mathbb{R}^+)}$.

The model description is complete by noting that the observable pattern of marks after censoring is

$$U = \bigcup_{(x_i,(a_i,l_i))\in W} \{(a_i, l_i)\}.$$

To obtain the probability distribution of U, write, for F in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathbb{R}\times\mathbb{R}^+}$,

$$\mathbb{P}(U \in F | X = \mathbf{x}) = \int_{(\mathbb{R}\times\mathbb{R}^+)^n} \mathbf{1}(\{(a_1, l_1), \ldots, (a_n, l_n)\} \in F) \prod_{i=1}^{n} d\nu((a_i, l_i)|x_i),$$

where $\mathbf{x} = \{x_1, \ldots, x_n\}$, and then take the expectation with respect to X.

Theorem 3.8. *Let W be a marked point process with ground process X on the open set $\mathcal{X} \subset \mathbb{R}$ defined by its probability density function p with respect to the distribution of a unit rate Poisson process having independent marks distributed according to the mark kernel $\nu(\cdot|x)$ for $x \in \mathcal{X}$ given by Equation 3.14. Let $\mathbf{u}$ be a realisation of U that consists of an atomic part $\{(a_1, 0), \ldots, (a_m, 0)\}$, $m \in \mathbb{N}_0$, and a non-atomic part $\{(a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$, $n \geq m$. Then the conditional distribution of X given $U = \mathbf{u}$ satisfies, for A in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathcal{X}}$,*

$$\begin{aligned}
&\mathbb{P}(X \in A \,|\, U = \mathbf{u}) = \\
&c(\mathbf{u}) \int_{\mathcal{X}^{n-m}} p(\{a_1, \ldots, a_m, x_{m+1}, \ldots, x_n\})\, 1_A(\{a_1, \ldots, a_m, x_{m+1}, \ldots, x_n\}) \\
&\qquad \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i, a_i+l_i]}(x_{D_i})\, dx_i
\end{aligned}$$

if the normalisation constant

$$c(\mathbf{u}) = \frac{1}{\int_{\mathcal{X}^{n-m}} p(\mathbf{x} \cup \{a_1, \ldots, a_m\}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[a_i, a_i+l_i]}(x_{D_i})\, dx_i}$$

exists in $(0, \infty)$.

Proof. We must prove, for each A in the Borel σ-algebra of $\mathcal{N}_{\mathcal{X}}$ with respect to the weak topology and each F in the Borel σ-algebra of the weak topology on $\mathcal{N}_{\mathbb{R}\times\mathbb{R}^+}$,

$$\mathbb{E}\left[\mathbf{1}_F(U)\, \mathbb{P}(X \in A \mid U)\right] = \mathbb{E}\left[\mathbf{1}_F(U)\, \mathbf{1}_A(X)\right], \tag{3.15}$$

as in Equation 4 of [90]. From the definition of expectation,

$$\begin{aligned}\mathbb{E}[1_F(U)1_A(X)] &= \int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})1_A(\mathbf{x})\, d\mathbb{P}(\mathbf{u},\mathbf{x}) \\ &= \int_{\mathcal{N}_{\mathcal{X}}} 1_A(\mathbf{x}) \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x})\, d\mathbb{P}(\mathbf{x})\end{aligned}$$

where the former integral describes the behaviour of the prior, and the latter the behaviour of the likelihood or model term. We now re-use notation introduced in Chapter 2.4.1. In the prior case, the Janossy density [34] derivation follows, and therefore

$$\int_{\mathcal{N}_{\mathcal{X}}} 1_A(\mathbf{x})\, d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \int_{\mathcal{X}^n} 1_A(\mathbf{x}) p_X(x_1,\ldots,x_n) \prod_{i=1}^{n} dx_i$$

where p is the prior distribution, and ℓ denotes Lebesgue measure. We again must split atoms from intervals, as in the renewal case. Assume that m out of n total points are perfectly observed atoms, and the remaining $n-m$ points are imperfectly observed and exist within time intervals. Let C_0 be, as in Chapter 2.4.1, an arbitrary subset of $\{1,\ldots,n\}$ of size m. Recall that

$$\begin{aligned}&j_{n-m}(u_{m+1},\ldots,u_n\,|\,x_{m+1},\ldots,x_n)\, du_{m+1}\,\ldots\, du_n = \rho_{n-m}(x_{m+1},\ldots,x_n) \\ &\sum_{\text{perm } \{u_1,\ldots,u_{n-m}\}} \pi_{n-m}(\{u_1,\ldots,u_{n-m}\}\,|\,x_{\{1,\ldots,n\}\setminus C_0}) \prod_{j=1}^{n-m} du_j,\end{aligned}$$

where j_n is Janossy density [34] and π_n the finite-dimensional distribution given n points in the process. We also introduce the atom probability for a given time $x \in \mathcal{X}$ as

$$w_x = 1 - \sum_{j=1}^{J} \delta_j \int_{A_j \cap (-\infty,x]} e^{-(x-s)\lambda(s)}\, ds. \tag{3.16}$$

Using this,

$$\begin{aligned}&\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\, d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x}) = \sum_{C_0\subseteq\{1,\ldots,n\}} \prod_{i=1}^{m} \mathbb{P}(x_i \text{ in } Z\text{-phase}) \frac{1}{(n-m)!} \\ &\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} \mathbb{P}(x_i \text{ in } Y\text{-phase}) 1_F(\{u_1,\ldots,u_{n-m}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\})) \\ &j_{n-m}(\{u_1,\ldots,u_{n-m}\}\,|\,x_{\{1,\ldots,n\}\setminus C_0}) \prod_{j=m+1}^{n} du_j \\ &= \sum_{C_0\subseteq\{1,\ldots,n\}} \prod_{i=1}^{m} w_{x_i} (1-w_x)^{n-m} \frac{1}{(n-m)!}\end{aligned}$$

$$\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1,\ldots,u_{n-m}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$\sum_{\substack{C_1,\ldots,C_{n-m}\\ \cup_j C_j\subseteq\{1,\ldots,n\}\setminus C_0}} \pi_{n-m}(\{u_1,\ldots,u_{n-m}\}\,|\,x_{\{1,\ldots,n\}\setminus C_0})\prod_{j=m+1}^{n} du_j.$$

We now replace the density function $q_x(a,l)$ for the alternating renewal process in Chapter 2.4.1 with

$$q_x(a,l)=\frac{m(a)g_Y(x,l)\,\mathbf{1}\{a\le x\le a+l\}}{\int_{-\infty}^{x}\left[1-G_Y(s,x-s)\right]\,dM(s)}$$

for measurable m and the semi-Markov process starting at $-\infty$. Whereas q_x in the renewal case only requires the time of occurrence to lie between a and $a+l$, the semi-Markov density function q is time-dependent due to the function $g_Y(x,l)$, which is the probability density of remaining in state 1 (away) for a given amount of time l since x infinitesimally. Hence

$$\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\,d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x})=\sum_{C_0\subseteq\{1,\ldots,n\}}\frac{1}{(n-m)!}$$
$$\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1,\ldots,u_{n-m}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$\prod_{i=1}^{m}\left(1-\int_{-\infty}^{x_i}1-G_Y(s,x_i-s)\,dM(s)\right)$$
$$\prod_{k=m+1}^{n}\left(\int_{-\infty}^{x_k}\left[1-G_Y(s,x_k-s)\right]dM(s)\right)$$
$$\sum_{\substack{C_1,\ldots,C_{n-m}\\ \cup_j C_j\subseteq\{1,\ldots,n\}\setminus C_0}}\prod_{j=m+1}^{n}\frac{m(a_j)g_Y(a_j,l_j)\,\mathbf{1}\{a_j\le x_{C_j}\le a_j+l_j\}}{\int_{-\infty}^{a_j}\left[1-G_Y(s,a_j-s)\right]\,dM(s)}\,da_j\,dl_j$$
$$=\sum_{C_0\subseteq\{1,\ldots,n\}}\frac{1}{(n-m)!}\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1,\ldots,u_{n-m}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$\prod_{i=1}^{m}\left(1-\int_{-\infty}^{x_i}\left[1-G_Y(s,x_i-s)\right]dM(s)\right)$$
$$\sum_{\substack{C_1,\ldots,C_{n-m}\\ \cup_j C_j\subseteq\{1,\ldots,n\}\setminus C_0}}\prod_{j=m+1}^{n} m(a_j)g_Y(a_j,l_j)\,1_{[a_j,a_j+l_j]}(x_{C_j})\,da_j\,dl_j. \tag{3.17}$$

Now

$$\mathbb{E}[1_F(U)1_A(X)]=\int_{\mathcal{N}_{\mathcal{X}}}1_A(\mathbf{x})\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}}1_F(\mathbf{u})\,d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x})\,d\mathbb{P}(\mathbf{x})=\sum_{n=0}^{\infty}\frac{e^{-\ell(\mathcal{X})}}{n!}$$

$$\int_{\mathcal{X}^n} 1_A(\mathbf{x}) p_X(x_1,\dots,x_n) \sum_{C_0\subseteq\{1,\dots,n\}} \frac{1}{(n-m)!}\prod_{i=1}^{m}$$
$$\left(1-\int_{-\infty}^{x_i}[1-G_Y(s,x_i-s)]\,dM(s)\right)$$
$$\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1,\dots,u_{n-m}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))$$
$$\sum_{\substack{C_1,\dots,C_{n-m}\\ \cup_j C_j\subseteq\{1,\dots,n\}\setminus C_0}} \prod_{j=m+1}^{n} m(a_j)g_Y(a_j,l_j)\,1_{[a_j,a_j+l_j]}(x_{C_j})\,da_j\,dl_j.\prod_{i=1}^{n}dx_i.$$

For the right-hand side,

$$\mathbb{E}[1_F(U)\mathbb{P}(X\in A\,|\,U)] = \int_{\mathcal{N}_{\mathcal{X}}}\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X\in A\,|\,U=\mathbf{u})\,d\mathbb{P}(\mathbf{x},\mathbf{u})$$
$$= \int_{\mathcal{N}_{\mathcal{X}}}\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X\in A\,|\,U=\mathbf{u})\,d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x})d\mathbb{P}(\mathbf{x}).$$

Following through,

$$\int_{\mathcal{N}_{\mathcal{X}}}\int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X\in A\,|\,U=\mathbf{u})\,d\mathbb{P}(\mathbf{u}\,|\,\mathbf{x})d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty}\frac{e^{-\ell(\mathcal{X})}}{n!}\sum_{C_0\subseteq\{1,\dots,n\}}$$
$$\int_{\mathcal{X}^n}\frac{1}{(n-m)!}\prod_{i=1}^{m}w_{x_i}\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}}\prod_{j=1}^{n-m}(1-w_{u_j})1_F(\{u_1,\dots,u_{n-m}\}$$
$$\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))\sum_{\substack{\emptyset=C_1,\dots,C_{n-m}\\ \cup_j C_j\subseteq\{1,\dots,n\}\setminus C_0}}\pi_{n-m}(\{u_1,\dots,u_{n-m}\}\,|\,x_{\{1,\dots,n\}\setminus C_0})$$
$$\mathbb{P}(X\in A\,|\,U=\{u_1,\dots,u_{n-m}\}\cup(\mathbf{x}_{C_0}\times\{\mathbf{0}\}))\prod_{j=m+1}^{n}du_j\prod_{i=1}^{m}dx_i. \tag{3.18}$$

making sure that terms are counted correctly. In analogy to Proposition 2.5 in Chapter 2, we propose the posterior form

$$\mathbb{P}(X\in A\,|\,U=\mathbf{u}) = c(\mathbf{u})\int_{\mathcal{X}^{n-m}} p(\{a_1,\dots,a_m,x_{m+1},\dots,x_n\})$$
$$1_A(\{a_1,\dots,a_m,x_{m+1},\dots,x_n\})\sum_{\substack{D_1,\dots,D_{n-m}\\ \cup_j D_j\subseteq\{1,\dots,n-m\}}}\prod_{i=m+1}^{n}1_{[a_i,a_i+l_i]}(x_{D_i})\,dx_i$$

where

$$c(\mathbf{u}) = \frac{1}{\int_{\mathcal{X}^{n-m}} p(\mathbf{x}\cup\{a_1,\dots,a_m\})\sum_{\substack{D_1,\dots,D_{n-m}\\ \cup_j D_j\subseteq\{1,\dots,n-m\}}}\prod_{i=m+1}^{n}1_{[a_i,a_i+l_i]}(x_{D_i})\,dx_i}$$

is the normalisation constant. Substituting,

$$\int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X \in A \,|\, U = \mathbf{u})\, d\mathbb{P}(\mathbf{u} \,|\, \mathbf{x}) d\mathbb{P}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \sum_{C_0 \subseteq \{1,\ldots,n\}}$$

$$\int_{\mathcal{X}^n} p(\{x_1,\ldots,x_n\}) \prod_{i=1}^{m} w_{x_i} \sum_{\substack{C_1,\ldots,C_{n-m} \\ \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0}} \frac{1}{(n-m)!} \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1,$$

$$\ldots, u_{n-m}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\})) \prod_{j=m+1}^{n} m(u_{j,1}) g_Y(x_{C_j}, u_{j,2}) 1_{[u_{j,1}, u_{j,1}+u_{j,2}]}(x_{C_j})$$

$$\frac{1}{\int_{\mathcal{X}^{n-m}} p(\mathbf{z} \cup (\mathbf{x}_{C_0} \times \{0\})) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[z_{i,1}, z_{i,1}+z_{i,2}]}(x_{D_i})\, dz_i}$$

$$\int_{\mathcal{X}^{n-m}} p(\mathbf{y} \cup \mathbf{x}_{C_0})\, 1_A(\mathbf{y} \cup \mathbf{x}_{C_0}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[y_{i,1}, y_{i,1}+y_{i,2}]}(x_{D_i})$$

$$\prod_{i=m+1}^{n} dy_i \prod_{j=m+1}^{n} du_j \prod_{k=1}^{n} dx_k.$$

We can again change the order of integration, as in the renewal case. Therefore

$$\int_{\mathcal{N}_{\mathcal{X}}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X \in A \,|\, U = \mathbf{u})\, d\mathbb{P}(\mathbf{u} \,|\, \mathbf{x}) d\mathbb{P}(\mathbf{x}) =$$

$$\sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \sum_{C_0 \subseteq \{1,\ldots,n\}} \int_{\mathcal{X}^n} p(\{x_1,\ldots,x_n\}) \prod_{i=1}^{m} w_{x_i}$$

$$\frac{1}{\int_{\mathcal{X}^{n-m}} p(\mathbf{z} \cup (\mathbf{x}_{C_0} \times \{0\})) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[z_{i,1}, z_{i,1}+z_{i,2}]}(x_{D_i})\, dz_i}$$

$$\int_{\mathcal{X}^{n-m}} p(\mathbf{y} \cup \mathbf{x}_{C_0})\, 1_A(\mathbf{y} \cup \mathbf{x}_{C_0}) \sum_{\substack{D_1,\ldots,D_{n-m} \\ \cup_j D_j \subseteq \{1,\ldots,n-m\}}} \prod_{i=m+1}^{n} 1_{[y_{i,1}, y_{i,1}+y_{i,2}]}(x_{D_i})$$

$$\sum_{\substack{C_1,\ldots,C_{n-m} \\ \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0}} \frac{1}{(n-m)!} \int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1,\ldots,u_{n-m}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\}))$$

$$\prod_{j=m+1}^{n} m(u_{j,1}) g_Y(x_{C_j}, u_{j,2}) 1_{[u_{j,1}, u_{j,1}+u_{j,2}]}(x_{C_j}) \prod_{i=m+1}^{n} du_i \prod_{j=m+1}^{n} dy_j \prod_{k=1}^{n} dx_k.$$

As the normalisation constant now cancels, after careful rearranging,

$$\int_{\mathcal{N}_\mathcal{X}} \int_{\mathcal{N}_{\mathbb{R}\times\mathbb{R}_0^+}} 1_F(\mathbf{u})\mathbb{P}(X \in A \,|\, U = \mathbf{u})\, d\mathbb{P}(\mathbf{u} \,|\, \mathbf{x}) d\mathbb{P}(\mathbf{x}) =$$
$$\sum_{n=0}^{\infty} \frac{e^{-\ell(\mathcal{X})}}{n!} \int_{\mathcal{X}^n} 1_A(\mathbf{x}) p(\{x_1, \ldots, x_n\})$$
$$\sum_{C_0 \subseteq \{1,\ldots,n\}} \frac{1}{(n-m)!} \prod_{i=1}^{m} \left(1 - \int_{-\infty}^{x_i} 1 - G_Y(s, x_i - s)\, dM(s)\right)$$
$$\int_{(\mathbb{R}\times\mathbb{R}_0^+)^{n-m}} 1_F(\{u_1, \ldots, u_{n-m}\} \cup (\mathbf{x}_{C_0} \times \{\mathbf{0}\}))$$
$$\sum_{\substack{C_1,\ldots,C_{n-m} \\ \cup_j C_j \subseteq \{1,\ldots,n\}\setminus C_0}} \prod_{j=m+1}^{n} m(u_{j,1}) g_Y(x_{C_j}, u_{j,2}) 1_{[u_{j,1}, u_{j,1}+u_{j,2}]}(x_{C_j}) du_i \prod_{k=1}^{n} dx_k$$
$$= \mathbb{E}[1_F(U) 1_A(X)].$$

Therefore,

$$\mathbb{E}\left[\mathbf{1}_F(U)\, \mathbb{P}(X \in A \mid U)\right] = \mathbb{E}\left[\mathbf{1}_F(U)\, \mathbf{1}_A(X)\right].$$

□

Strikingly, although the marking mechanism is more complicated than that in Chapter 2, the conditional distribution of X has the same form. The conditional distribution of W can be obtained in the same vein, by considering $1_A(W)$ instead of $1_A(X)$ for A a Borel set in $\mathcal{N}_{\mathcal{X}\times(\mathbb{R}\times\mathbb{R}^+)}$, the space of marked point configurations. It is given by

$$\mathbb{P}(W \in A | U = \mathbf{u}) \propto \int_{\mathcal{X}^{n-m}} p(\{a_1, ..., a_m, x_1, ..., x_{n-m}\})$$
$$1_A(\{(a_1, (a_1, 0)), \ldots, (a_m, (a_m, 0)), (x_1, (a_{m+1}, l_{m+1})), \ldots, (x_{n-m}, (a_n, l_n))\})$$
$$\prod_{i=1}^{n-m} 1_{[a_{m+i}, a_{m+i}+l_{m+i}]}(x_i)\, dx_i.$$

3.5 Semi-Markov modelling

We now turn to modelling the functions introduced in earlier sections of this chapter, specifically $p(\cdot)$ and $\nu(\cdot|x)$. As could be seen by the derivations in Chapter 3.3.1, while it is theoretically possible to model M as functions of the probability density (and thus G), it quickly becomes intractable. Recall that

$$M(t) = \sum_{n=1}^{\infty} \mathbb{P}(X_{2n} \leq t).$$

See that

$$\begin{aligned}\mathbb{P}(X_2 \leq t) &= \mathbb{P}(T_1 + T_2 \leq t) \\ &= \int_0^t \mathbb{P}(T_2 \leq t - \tau)\mathbb{P}(T_1 = \tau)\, d\tau \\ &= \int_0^t G_Y(\tau, t-\tau) g_Y(0, \tau)\, d\tau.\end{aligned}$$

In general, an expression for $\mathbb{P}(X_n \leq t)$ becomes a convolution of $T_1, \ldots, T_n$, which involves integrals that are intractable for non-trivial distributions. By extension, M can be written as sums of convolutions of intractable integrals. It is thus infeasible to directly model the Markov renewal function through the semi-Markov kernel. To proceed, parametric forms for G_Y and m must be developed.

3.5.1 Parametric modelling of the mark kernel

We begin by modelling G_Y, the semi-Markov kernel that determines the length of time until the next transition. We may take one of the time-dependent probability density functions considered in Chapter 3.2. For instance, $g_Y(a, l)$ could be the density function of an exponential distribution with rate

$$\lambda(a; \alpha) = \alpha\left(b + \sin(ca)\right), \qquad a \in \mathbb{R}, \tag{3.19}$$

where c specifies the period and $b \geq 1$ the elevation away from 0. The parameter α determines the amplitude of the harmonic.

We could proceed in a similar fashion for g_Z. However, there are two problems with such an approach. From a probabilistic point of view, tractable expressions for the renewal density m in terms of the semi-Markov kernels G_Y and G_Z do not seem to exist, and, statistically speaking, lengths of Z phases cannot be observed. Therefore, we shall model m directly. The following proposition justifies this approach.

Proposition 3.9. *Let $(S_n, X_n)_{n=1}^{\infty}$ be a semi-Markov process on $\{1\} \times \mathbb{R}^+$ with $S_0 = 1$ and $X_0 = 0$ having semi-Markov kernel G_Y defined by a density function $g_Y(t, \tau)$, $t \in \mathbb{R}^+$, $\tau \in [0, \infty)$ and write $\tilde{m}$ for the density of its renewal function*

$$\tilde{M}(t) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \leq t).$$

If $h(t) : \mathbb{R}^+ \to [0, \infty)$ is a Borel-measurable function such that $h(t) \leq \tilde{m}(t)$, then there exists an alternating semi-Markov process on $\{0, 1\} \times \mathbb{R}^+$ with $G_{01} = G_Y$ and renewal density h.

Proof. As $0 \leq h(t)/\tilde{m}(t) \leq 1$, we may use a time-dependent thinning approach with retention probability $p(t) = h(t)/\tilde{m}(t)$. Algorithmically, the sought-after process can be constructed as follows. Initialise $\hat{S}_0 = 1$, $\hat{X}_0 = 0$ and $\hat{X}_1 = X_1$. Also set $\hat{S}_{2i} = 1$, $\hat{S}_{2i-1} = 0$ for $i \in \mathbb{N}$ and $j = 1$. For each jump time X_i, $i = 1, 2, \ldots,$

- with probability $p(X_i)$, if j is even, update $\hat{X}_{j+1} = X_{i+1}$ and increment j by 1; for odd j update $\hat{X}_{j+1} = \hat{X}_j$, $\hat{X}_{j+2} = X_{i+1}$ and increment j by 2;

- else, if j is odd, update $\hat{X}_{j+1} = X_{i+1}$ and increment j by 1; for even j update $\hat{X}_j = X_{i+1}$ leaving j unchanged.

Because complete cycles correspond to intervals in between accepted points X_i, $i = 0, 1, 2, \ldots,$

$$H(t) = \sum_{n=1}^{\infty} \mathbb{P}(\hat{X}_{2n} \leq t) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \leq t; X_n \text{ retained}) = \int_0^t \frac{h(s)}{\tilde{m}(s)} d\tilde{M}(s).$$

We can therefore conclude that the intensity of the thinned process is $\frac{h(t)}{\tilde{m}(t)}\tilde{m}(t) = h(t)$ (see e.g. [34, pp. 78–79] and hence $(\hat{S}_n, \hat{X}_n)_{n=1}^{\infty}$ is a non-homogeneous alternating semi-Markov process that satisfies the proposed conditions. □

As an illustration, suppose that m is a step function

$$m(t) = \sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(t), \quad t \in \mathbb{R}^+, \tag{3.20}$$

that takes J different values $\delta_j > 0$ on Borel sets A_j forming a partition ($j = 1, \ldots, J$, with $J \in \mathbb{N}$). The following corollary lays out conditions under which $h = m$ is the renewal density of an alternating renewal process whose Y-phases are governed by Equation 3.19.

Corollary 3.10. *A sufficient condition for Equation 3.20 to be the renewal density of an alternating semi-Markov process on $\{0, 1\} \times \mathbb{R}^+$ with G_Y given by Equation 3.19 on $\mathbb{R}^+$ is that for all $j = 1, \ldots, J$ we have $\delta_j \leq \alpha(b - 1)$.*

Proof. For Equation 3.20 to induce a semi-Markov process, we require $h(t) \leq \tilde{m}(t)$, where $\tilde{m}(t)$ is the Radon-Nikodym derivative of the renewal function $\tilde{M}(t)$. In Proposition 3.9 we defined

$$\tilde{M}(t) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \leq t),$$

with $(X_n)_{n=0}^{\infty}$ being its associated jump process of only Y-phases. By construction, its conditional intensity is $\tilde{\lambda}_{n+1}(t; t_1, \ldots, t_n) = \lambda(t, \alpha)$ for all $0 \leq t_1 \leq \cdots \leq t_n \leq t$.

Observe that $\inf\{\lambda(t; \alpha) : t \in \mathbb{R}\} = \alpha(b-1)$. Construct a Poisson process $N^*(t)$ with intensity $\nu = \alpha(b-1)$. By [59, Corollary 1], since $\lambda^*_{n+1}(t; t_1, \ldots, t_n) \leq \tilde{\lambda}_{n+1}(t; t_1, \ldots, t_n)$, we may conclude that the renewal function νt of $N^*(t)$ is bounded from above by $\tilde{M}(t)$ for all t. Hence also $\nu \leq \tilde{m}(t)$.

For $h = m$ as in Equation 3.20, to have $\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(t) \leq \alpha(b-1)$, it is sufficient that $\delta_j \leq \alpha(b-1)$ for all $j = 1, \ldots, J$ to guarantee that $m(t)$ is the renewal density of a semi-Markov process. □

As noted before, in practice, the starting point 0 is moved back to $-\infty$. Realisations $\mathbf{u}$ from the specified model may be obtained as follows. First, a set of points $\mathbf{x} \subset \mathcal{X}$ in time are chosen according to the probability density function $p(\cdot)$ by, for example, coupling from the past [78] or the Metropolis–Hastings algorithm [54]. Next, for each point $x \in \mathbf{x}$, it is determined whether or not it is an atom based on w_x. If this is not the case, we appeal to Proposition 3.7 and use rejection sampling with a proposal distribution that simulates a uniformly distributed point in $A_j \cap (-\infty, x]$ chosen with probability $\delta_j \ell(A_j \cap (-\infty, x]) / \sum_{i=1}^{J} \delta_i \ell(A_i \cap (-\infty, x])$ and acceptance probability $\exp[-\lambda(a; \alpha)(x - a)]$. The result is a sample a from $f_x(a)$, cf. Equation 3.11. The length is then sampled according to an exponential distribution with parameter $\lambda(a; \alpha)$ shifted by $x - a$ (see Equation 3.12). It is interesting to observe that, in contrast to the alternating renewal case studied in Chapter 2 using the marginal distribution with respect to A and then the conditional given A is computationally simpler than sampling L first.

3.5.2 Non-homogeneous point process densities

We will now look at inhomogeneity that manifests itself via the occurrence time distribution. In view of Equation 3.13, it is natural to add inhomogeneity by means of the first-order interaction function ϕ_1, a procedure known as type I inhomogeneity [74]. The idea is to let $\phi_1(\{x\}) = \beta(x)$ vary over time according to a measurable function β that maps $x \in \mathcal{X}$ to $[0, \infty)$. In many applications, it may make sense to model β as a step function. More specifically, given a measurable partition B_k, $k = 1, \ldots, K$, of $\mathcal{X}$, set

$$\beta(x) = \sum_{k=1}^{K} \beta_k \mathbf{1}_{B_k}(x), \qquad x \in \mathcal{X} \tag{3.21}$$

where $\beta_k \geq 0$ is the value that β takes in the corresponding set B_k.

The function ϕ_1 can be combined with classic second and higher-order interaction functions. For instance, the density of the non-homogeneous area-interaction point process [9] becomes

$$p(\mathbf{x}) = \alpha_p \left(\prod_{x \in \mathbf{x}} \beta(x) \right) \exp\left[-\log \gamma \, \ell(\mathcal{X} \cap U_r(\mathbf{x})) \right] \tag{3.22}$$

with respect to a unit rate Poisson process on $\mathcal{X}$. The parameter γ quantifies the interaction strength, r the radius of interaction, and $\alpha_p = c(\beta(\cdot), \gamma)$ is a normalisation constant [9] that depends on the function β as well as on γ. Additionally, $U_r(\mathbf{x}) = \bigcup_{i=1}^{n} B(x_i, r)$ where $B(x_i, r)$ is the closed interval $[x_i - r, x_i + r]$. We observe regularity for $\gamma < 1$, clustering for $\gamma > 1$, and $\gamma = 1$ corresponds to a non-homogeneous Poisson process with intensity function β. For further examples, we refer to [86].

3.6 Statistical aspects

In practical applications, both the family of probability density functions $g_Y(t, \tau; \theta)$ for the sojourn times in phase Y and the function $m(t; \xi)$ rely on unknown parameters $\eta = (\theta, \xi)$ that must be estimated. The log-likelihood $L(\eta; \mathbf{u})$ follows directly from Equation 3.17. Upon observing $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0), (a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$,

$$\begin{aligned} L(\eta; \mathbf{u}) = &\sum_{i=1}^{m} \log \left(1 - \int_{-\infty}^{a_i} \left[1 - G_Y(s, a_i - s; \theta) \right] m(s; \xi) \, ds \right) \\ &+ \sum_{i=m+1}^{n} \log \left(m(a_i; \xi) \, g_Y(a_i, l_i; \theta) \right). \end{aligned} \tag{3.23}$$

When the sojourn time distributions G_Y and G_Z and hence the renewal density $m \equiv (\mathbb{E}Y + \mathbb{E}Z)^{-1}$ are not time-varying, we now proceed by showing that Equation 3.23 reduces to the renewal likelihood for some models, including (but not limited to) the exponential model below. Assume that $G_Y(t, \cdot)$, $t \in \mathbb{R}$, is distributed exponentially with rate parameter $\lambda(t)$ as in Equation 3.19 and m given by Equation 3.20. In the homogeneous case that $\lambda(t) \equiv \alpha > 0$ (that is, $b = 1$ and $c = 0$), $J = 1$, $A_1 = \mathbb{R}$ and $0 \leq \delta_1 \leq \alpha$ we obtain

$$\int_{-\infty}^{t} \left[1 - G_Y(s, t - s) \right] m(s) \, ds = \int_{-\infty}^{t} e^{-\alpha(t-s)} \, ds = \frac{1}{\alpha} = \mathbb{E}[Y] \tag{3.24}$$

where $\mathbb{E}[Y]$ refers to the expected length of a Y-phase in a homogeneous alternating renewal process. Using this result, we obtain $f_Y(l; \alpha) = g_Y(t, l; \lambda(t))$, where f_Y is the probability density function of the length of such a Y-phase, see

Proposition 2.7 in Chapter 2. Following through, we additionally obtain

$$L(\alpha, \delta_1; \mathbf{u}) = m \log\left(1 - \frac{\delta_1}{\alpha}\right) + (n-m)\log\delta_1 + (n-m)\log\alpha - \alpha \sum_{i=m+1}^{n} l_i,$$

which corresponds exactly to the simplified version of Equation 2.14 in Chapter 2, the log-likelihood in the homogeneous case.

The likelihood equation (see Equation 3.23) after the substitution and discarding of terms that do not depend on the parameters becomes

$$\begin{aligned} L(\delta, \alpha; \mathbf{u}) = &\sum_{i=1}^{m} \log\left(1 - \sum_{j=1}^{J} \delta_j \int_{(a_i - A_j)\cap[0,\infty)} e^{-\alpha r(b+\sin(ca_i - cr))}\, dr\right) \\ &+ \sum_{i=m+1}^{n} \log\left(\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)\right) + (n-m)\log\alpha \\ &- \alpha \sum_{i=m+1}^{n} l_i \left(b + \sin\left(ca_i\right)\right). \end{aligned}$$

The resulting equations can be solved numerically to find optimal values for $\delta_k, k = 1, \ldots, J$, and α under the inequality constraints $0 \leq \delta_j \leq \alpha(b-1)$, $j = 1, \ldots, J$.

Assume that G_Y, the semi-Markov kernel that determines the length of time before the next transition, follows one of the time-dependent distributions considered in Chapter 3.2 (exponential, gamma or Weibull). For simplicity, assume m, the derivative of the Markov renewal function, can be written as a step function taking J different values. That is, $m(x) = \sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(x)$, where A_j is the interval in which x takes the value δ_j. Writing the integral in terms of length by setting $r = a_i - s$, one can rewrite equation 3.23 as

$$\begin{aligned} L(\eta; \mathbf{u}) &= \sum_{i=1}^{m} \log\left(1 - \int_{-\infty}^{a_i} m(a_i - r; \delta)\left[1 - G_Y(a_i - r, r; \theta)\right] dr\right) \\ &\quad + \sum_{i=m+1}^{n} \log\left(m(a_i; \delta) g_Y(a_i, l_i; \theta)\right) \\ &= \sum_{i=1}^{m} \log\left(1 - \sum_{j=1}^{J} \delta_j \int_{a_i - A_j} \left[1 - G_Y(a_i - r, r; \theta)\right] dr\right) \\ &\quad + \sum_{i=m+1}^{n} \log\left(\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)\right) + \log\left(g_Y(a_i, l_i; \theta)\right). \end{aligned} \tag{3.25}$$

3.6.1 Example: Exponential case

We wish to sample from the full model W as laid out in Chapter 3.4.1; that is to say, we require n intervals and corresponding times. Let the Markov kernels $G_Y(t,\cdot)$ follow an exponential distribution with rate

$$\lambda(t;\alpha) = \alpha\left(b + \sin(ct)\right). \tag{3.26}$$

Here $c = \frac{2\pi}{d}$ denotes the period parameter for a set number of days d, and b is a vertical transformation. The harmonic function present in $\lambda(t;\alpha)$ allows for the modelling of fluctuations in the starting time. The scale parameter α determines the amplitude of the harmonic. We have

$$g_Y(a_i, l_i; \lambda_{a_i}) = \lambda_{a_i} e^{-l_i \lambda_{a_i}} \qquad G_Y(a_i, l_i; \lambda_{a_i}) = 1 - e^{-l_i \lambda_{a_i}}.$$

The bound $m(t) \leq \alpha(b-1)$ evaluates to

$$\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(t) \leq \inf_{t>a}\{\lambda(a;\alpha)\} = \alpha(b-1).$$

A more cogent way of writing this is that, for all $j = 1, \ldots, J$, $\delta_j \leq \alpha(b-1)$.

The marginal distribution for the length (see Equation 3.11) on $(-\infty, t]$ becomes:

$$f_t(a) = \frac{m(a)[1 - G_Y(a, t-a)]}{\int_{-\infty}^{t} [1 - G_Y(s, t-s)]\, dM(s)} = \frac{\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a) e^{-(t-a)\lambda_a}}{\sum_{j=1}^{J} \delta_j \int_{(-\infty,t]\cap A_j} e^{-(t-s)\lambda_s}\, ds}$$

with the conditional distribution (see Equation 3.12) on $[t-a, \infty)$ being

$$f_{t,\, L\,|A=a}(l) = \frac{g_Y(a,l)}{1 - G_Y(a, t-a)} = \lambda_a e^{-(l-t+a)\lambda_a}.$$

The likelihood equation analogous to Equation 3.25 after substitution becomes

$$\begin{aligned} L(\delta, \alpha; \mathbf{u}) = &\sum_{i=1}^{m} \log\left(1 - \sum_{j=1}^{J} \delta_j \int_{a_i - A_j} e^{-\alpha r(1+\sin(ca_i - cr))}\, dr\right) \\ &+ \sum_{i=m+1}^{n} \log\left(\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)\right) \\ &+ \sum_{i=m+1}^{n} \left[\log\alpha + \log\left(b + \sin\left(ca_i\right)\right) - \alpha l_i \left(b + \sin\left(ca_i\right)\right)\right] \end{aligned}$$

in the exponential case. The derivative of this likelihood function with respect to $\delta_k, k = 1, \ldots, J$, is

$$\frac{\partial L}{\partial \delta_k} = \sum_{i=1}^{m} \frac{-\int_{a_i - A_k} e^{-\alpha r(b+\sin(ca_i - cr))}\, dr}{1 - \sum_{j=1}^{J} \delta_j \int_{a_i - A_j} e^{-\alpha r(b+\sin(ca_i - cr))}\, dr} + \sum_{i=m+1}^{n} \frac{\mathbf{1}_{A_k}(a_i)}{\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)},$$

and with respect to α, by the Leibniz integral rule, it is

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{m} \frac{\sum_{j=1}^{J} \delta_j \int_{a_i - A_j} r(b + \sin(ca_i - cr)) e^{-\alpha r(b+\sin(ca_i - cr))}\, dr}{1 - \sum_{j=1}^{J} \delta_j \int_{a_i - A_j} e^{-\alpha r(b+\sin(ca_i - cr))}\, dr} + \frac{n-m}{\alpha} - \sum_{i=m+1}^{n} l_i(b + \sin(ca_i)).$$

These derivatives can subsequently be used to calculate the maximum likelihood.

3.6.2 Example: Weibull case

Let $Y_{i+1} \sim \text{Weibull}(k_{a_i}, \lambda_{a_i})$, with $k_{a_i} = k$, λ_{a_i} and $m(a_i)$ unchanged from Chapter 3.6.1. Then

$$\begin{aligned} g_Y(a_i, l_i; k_{a_i}, \lambda_{a_i}) &= k_{a_i} \lambda_{a_i} \left(l_i \lambda_{a_i}\right)^{k_{a_i}-1} e^{-(l_i \lambda_{a_i})^{k_{a_i}}} \\ &= k\alpha \left(b + \sin(ca_i)\right) \left(l_i \alpha \left(b + \sin(ca_i)\right)\right)^{k-1} e^{-(l_i \alpha (b+\sin(ca_i)))^k} \end{aligned}$$

and

$$G_Y(a_i, l_i; k_{a_i}, \lambda_{a_i}) = 1 - e^{-(l_i \alpha (b+\sin(ca_i)))^k}.$$

The likelihood expression becomes

$$\begin{aligned} L(\eta; \mathbf{u}) = &\sum_{i=1}^{m} \log \left(1 - \sum_{j=1}^{J} \delta_j \int_{a_i - A_j} \exp\left(- \left(\alpha r (b + \sin\left(c(a_i - r)\right))\right)^k \right) dr \right) \\ &+ \sum_{i=m+1}^{n} \log \left(\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i) \right) + \sum_{i=m+1}^{n} \left[\log k + \log \alpha + \log(b + \sin(ca_i)) \right. \\ &\left. + (k-1) \left[\log l_i + \log \alpha + \log(b + \sin(ca_i)) \right] - (l_i \alpha (b + \sin(ca_i)))^k \right]. \end{aligned}$$

The derivative for δ_d where $d = 1, \ldots, J$ is

$$\frac{\partial L}{\partial \delta_d} = \sum_{i=1}^{m} \frac{- \int_{a_i - A_d} \exp\left(- \left[\alpha r (b + \sin\left(c(a_i - r)\right))\right]^k \right) dr}{1 - \sum_{j=1}^{J} \delta_j \int_{a_i - A_j} \exp\left(- \left[\alpha r (b + \sin\left(c(a_i - r)\right))\right]^k \right) dr} + \sum_{i=m+1}^{n} \frac{\mathbf{1}_{A_k}(a_i)}{\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)}.$$

The marginal distribution for the length (see equation 3.11) on $(-\infty, t]$ becomes:

$$f_t(a) = \frac{m(a)[1 - G_Y(a, t-a)]}{\int_{-\infty}^{t} \left[1 - G_Y(s, t-s)\right] dM(s)} = \frac{\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a) e^{-((t-a)\lambda_a)^k}}{\sum_{j=1}^{J} \delta_j \int_{(-\infty, t] \cap A_j} e^{-((t-s)\lambda_s)^k}\, ds}$$

with the conditional distribution (see equation 3.12) on $[t-a, \infty)$ being

$$f_{t,\,L\,|A=a}(l) = \frac{g_Y(a,l)}{1-G_Y(a,t-a)} = k\lambda_a(l\lambda_a)^{k-1}e^{-((l-t+a)\lambda_a)^k}.$$

3.6.3 Example: Gamma case

Let $Y_{i+1} \sim \text{Gamma}(k_{a_i}, \lambda_{a_i})$. Let $\lambda_{a_i} = \alpha(b+\sin(ca_i))$ as before, take $m(a_i) = \sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)$ and let $k_{a_i} = k$. Then

$$\begin{aligned} g_T(a_i, l_i; k_{a_i}, \lambda_{a_i}) &= \frac{\lambda_{a_i}^{k_{a_i}} l_i^{k_{a_i}-1} e^{-\lambda_{a_i} l_i}}{\Gamma(k_{a_i})} \\ &= \frac{[\alpha\,(b+\sin(ca_i))]^k l_i^{k-1} e^{-\alpha(b+\sin(ca_i))l_i}}{\Gamma(k)} \end{aligned}$$

and

$$G_T(a_i, l_i; k_{a_i}, \lambda_{a_i}) = \frac{\gamma(k_{a_i}, \lambda_{a_i} l_i)}{\Gamma(k_{a_i})} = \frac{\gamma(k, \alpha l_i(b+\sin(ca_i)))}{\Gamma(k)}$$

where Γ is the gamma function and γ is the lower incomplete gamma function. The likelihood becomes

$$\begin{aligned} L(\eta; \mathbf{u}) = &\sum_{i=1}^{m} \log\left(1 - \sum_{j=1}^{J} \frac{\delta_j}{\Gamma(k)} \int_{a_i - A_j} \Gamma(k, \alpha r(b+\sin(c(a_i-r)))) \, dr\right) \\ &+ \sum_{i=m+1}^{n} \log\left(\sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(a_i)\right) + k\log\alpha + k\log(b+\sin(ca_i)) \\ &+ (k-1)\log l_i - \alpha l_i\,(b+\sin(ca_i)) - \log\Gamma(k). \end{aligned}$$

Derivatives and associated distributions can be found in analogy to the exponential and Weibull cases in Chapters 3.6.1 and 3.6.2.

3.6.4 State estimation methods

Now that parameters have been estimated, we wish to sample from the full model W as laid out in Chapter 3.4.1; that is to say, we require n intervals and corresponding times. This involves calculating the posterior distribution given in Theorem 3.8. However, due to the existence of the normalisation constant in the definition of Theorem 3.8, it is infeasible to evaluate this analytically. We will draw from the literature on spatio-temporal point process estimation [23, 102] and use a Metropolis-Hastings algorithm for a fixed number of points. Using this method, we can determine the most likely locations of points within their corresponding intervals.

There is no explicit mark distribution inherent in the points of W, but rather a time-dependent measure assigning interval lengths to each point. However, since the form of the conditional distribution of W given U according to Theorem 3.8 is identical to that for alternating renewal process-based censoring, one can still make use of Chapter 2.5.5 on state estimation to generate a sample of W using a fixed-point Metropolis-Hastings algorithm. The idea is to sample from the space $\mathcal{X}^{n-m}$ where n refers to the number of observations and m the number of atoms. Note that the state space remains the same as in the standard homogeneous renewal case. After completion, we are left with a complete sample of the posterior distribution, containing n most likely intervals with the most likely location of their corresponding points. It is known that this algorithm converges from almost all starting states (see [90], Propositions 4.3-4.5). Estimation of any prior parameters, for instance the β_k in Equation 3.21, requires a Monte Carlo EM approach such as in Chapter 2.7.

3.7 Illustrations in practice

To show how the non-homogeneous semi-Markov model behaves, we present a number of examples that compare the new model with a homogeneous one. Recall that, broadly speaking, there are three sources of inhomogeneity: the interval lengths as governed by g_Y, the renewal density m, and the ground process responsible for the uncensored event occurrences. Throughout this subchapter, we set $\mathcal{X} = (0, 1)$.

3.7.1 Model mis-specification

The first source of inhomogeneity in our model is the semi-Markov kernel $G_Y(a, l)$ for starting point $a \in \mathbb{R}$ and length $l \geq 0$, which determines the time until the next jump. In the homogeneous case, as in Equation 3.24, these may be assumed to be independent and identically Gamma or Weibull distributed. In the non-homogeneous case, the density function is dependent on a.

We have devised the following experiment to illustrate the effect of using a homogeneous model erroneously. We generate a realisation of intervals using the non-homogeneous alternating renewal process model, where the interval censoring mechanism is governed by a Weibull distribution with shape parameter $k = 1$ and rate parameter $\lambda_Y(t; \alpha) = \alpha(1.6+\sin(2\pi t))$ (see (3.19)) for $\alpha = 1$. Regarding the other model ingredients, we assume that $p(\cdot)$ is of the form given in (3.22) with $\beta = 400$ and $\gamma = 1$, that is, a homogeneous Poisson process on $\mathcal{X}$ with intensity 400. We additionally set $m(t) = 0.6\,\mathbf{1}_{[-0.2,1)}(t)$.

After generating these intervals using the non-homogeneous model, we now assume, wrongly, that these intervals were instead generated by a homogeneous interval censoring scheme. Specifically, we fit a Weibull distribution with param-

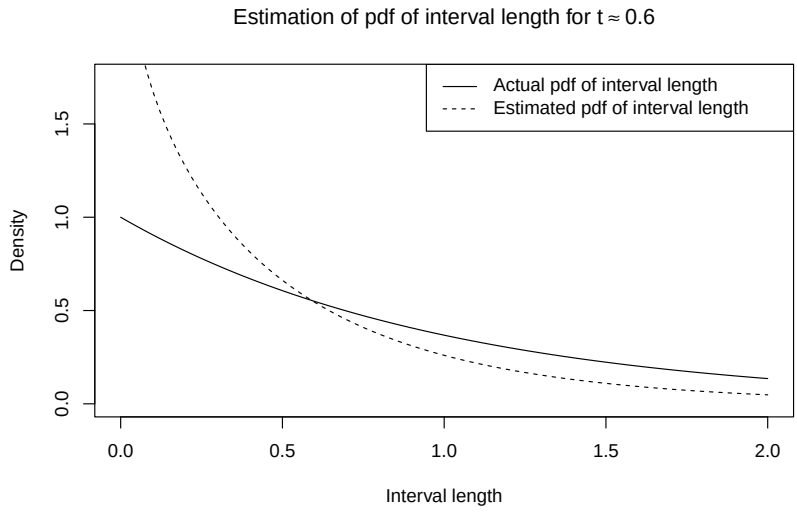


Figure 3.2 The unbroken line corresponds to the actual probability density of interval length for $k = 1$ and $\lambda(0.6; 1) = 1$. The dotted line corresponds to the estimated survival time density.

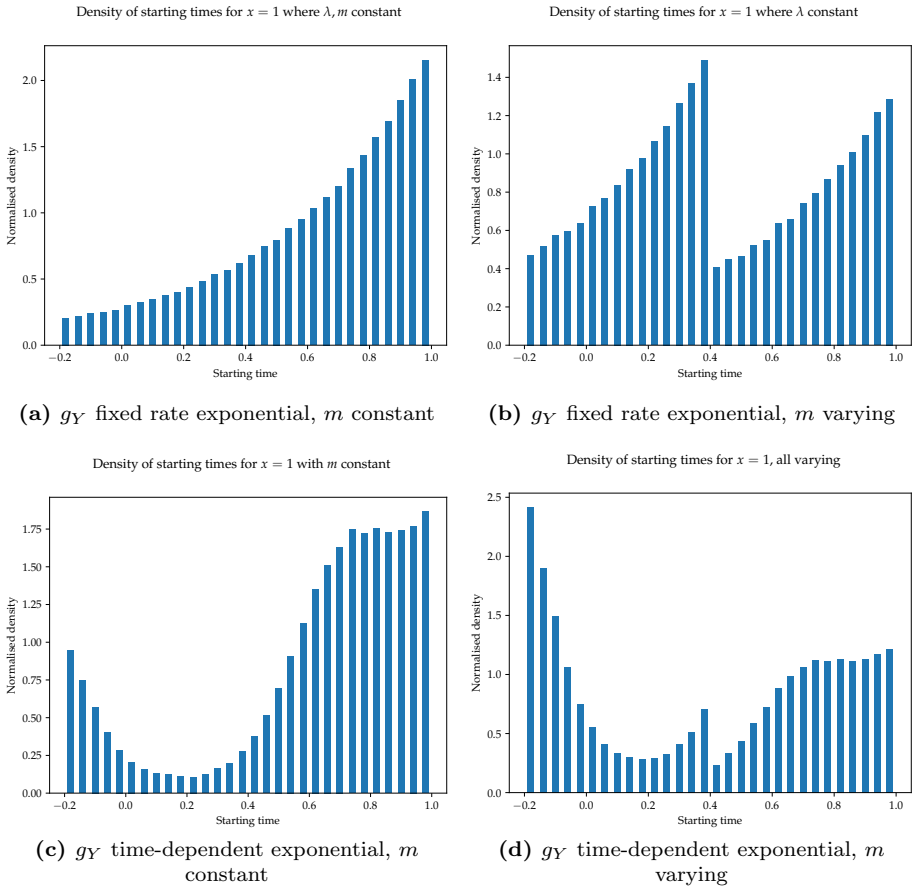
eters $k > 0$ and constant rate $\lambda_Y(t; \alpha) \equiv \alpha$ for $\alpha > 0$ using maximum likelihood estimation (see Chapter 2.5.3). We obtain parameter estimates $\hat{k} = 0.9$ and $\hat{\alpha} = 2.0$.

We plot the survival time densities for both sets of parameters. We must choose an exact point in time to obtain the Weibull parameters for the non-homogeneous model, so we assume that we are looking at time $t = 0.6$. This plot, for both models, can be seen in Figure 3.2. Compared to the actual model, the homogeneous model can roughly discern the shape of the distribution, but struggles with the scale. A homogeneous model would have generated more intervals shorter than about 0.5 and fewer of longer length.

3.7.2 Inhomogeneity in renewal density and survival time

In our second experiment, we add inhomogeneity in m to the model and study the effect on f_x, cf. (3.11). As in Chapter 3.7.1, consider an exponential semi-Markov kernel density g_Y with rate parameter either constant, $\lambda(t; \alpha) = 1.3\alpha$, or varying in time according to $\lambda(t; \alpha) = \alpha(1.3 + \sin(2\pi t))$. Furthermore, set $m(t) = 0.4$ for $t \in [-0.2, 1)$ in the constant case, and

$$m(t) = \begin{cases} 0.4 & t \in [-0.2, 0.4) \\ 0.1 & t \in [0.4, 1) \end{cases}$$



(a) g_Y fixed rate exponential, m constant

(b) g_Y fixed rate exponential, m varying

(c) g_Y time-dependent exponential, m constant

(d) g_Y time-dependent exponential, m varying

Figure 3.3 Probability density function of the starting time $f_x(\cdot)$ with $x = 1$ for various choices of g_Y and m.

in the time-varying case. We set $\alpha = 1.6$, so the largest value of δ_i which guarantees that g_Y is the Radon-Nikodym derivative of a semi-Markov kernel is $1.6 \times (1.3 - 1) = 0.48$. Figure 3.3 shows the graphs of $f_x(\cdot)$ for the four possible combinations of g_Y and m obtained from 200,000 samples from q_x for $x = 1$. In Figures 3.3a and 3.3b, we assume that λ is constant. When m is also constant as in Figure 3.3a, the marginal distribution of the starting times, by Proposition 3.7, is a shifted exponential distribution. If m is allowed to vary in time, the exponential curve is broken at $t = 0.4$, the discontinuity point of m, resulting in a zigzag pattern. In both Figures 3.3a and 3.3c, m is constant, but in Figure 3.3c the rate parameter of g_Y varies according to a harmonic. The resulting sinusoidal modulation is clearly visible. Finally, allowing m to vary too results in a break at its discontinuity point $t = 0.4$ as seen in Figure 3.3d. We

conclude that the non-homogeneous model is able to capture various types of inhomogeneity well and is far more flexible than the homogeneous one.

3.7.3 Inhomogeneity in occurrence time distribution

In the previous subsections, we have assumed that the first-order interaction function β of the point process X of occurrence times, remains constant over the entire sampling window $(0, 1)$. In our final example, we relax this assumption in that we consider a 'peak time' in which events are more likely to occur and investigate the effect on the conditional distribution of occurrences. More precisely, we take an area-interaction model (3.22) with

$$\beta(y) = \begin{cases} 3 & y \in [0, c_1) \\ 5 & y \in [c_1, c_2) \\ 3 & y \in [c_2, 1] \end{cases}$$

and critical range $[c_1, c_2) = [0.81, 0.85)$. The radius of interaction is set to $r = 0.1$ and we consider both a regular ($\eta = -1.2$) and a clustered ($\eta = 1.2$) model.

As in Chapter 2, consider the set $\mathbf{u} = \{(0.45, 0.4), (0.51, 0), (0.58, 0)\}$ that contains one non-degenerate interval. Recall that the entries are parametrised as (a, l), where a is the starting point and l is the length. Figure 3.4 plots the conditional distribution of the occurrence time on the interval $[0.45, 0.85]$ given $\mathbf{u}$ for the regular and clustered model. To create this figure, a Metropolis-Hastings algorithm (see Algorithm 4.2, [90]) has been run for 600,000 time steps, with the first 100,000 iterations being thrown out due to burn-in.

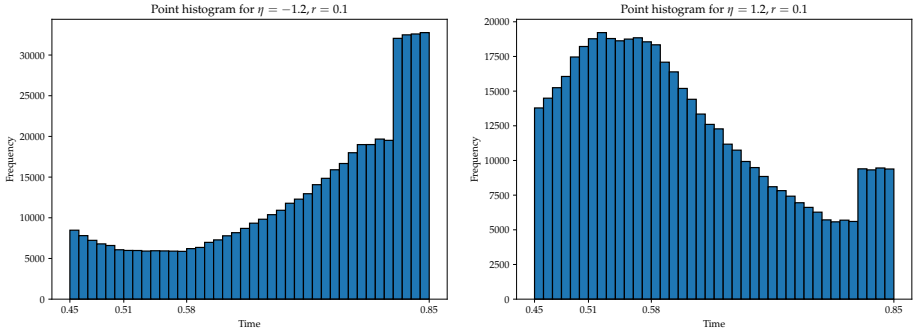


Figure 3.4 A comparison between a regular and clustered model with a 'peak time' added by changing the intensity function within a critical range.

The general shape of the graphs is similar to the corresponding plots for constant $\beta = 3$ in Figures 2.3 and 2.4 in Chapter 2.6.1. For the clustered model, the occurrence time is more likely to happen close to the atoms, for regular models the probability density is shifted away from the atoms. In the non-homogeneous

case, the higher value of β during the peak times causes a clear bump in the range $[c_1, c_2) = [0.81, 0.85)$, demonstrating the ability of the non-homogeneous model to favour certain occurrence times over others.

The previous three experiments show the effects that the added non-homogeneity may have on the model. It is able to draw from a more complicated interval censoring scheme to generate intervals, provide a more realistic starting time distribution based on the semi-Markov kernel and Markov renewal function while having the functionality of splitting the observation window into regions of differing likelihood, and deal with more complicated interaction structures in the underlying point process density. For a concrete application of both the homogeneous and non-homogeneous model to a dataset of residential burglaries, we direct the reader to Chapter 5.

3.8 Discussion

We introduced a time-dependent interval censoring mechanism that splits time into observable and partially observable phases by means of a non-homogeneous alternating renewal process on the real line. The process was shown to be well-defined for a range of Gamma and Weibull semi-Markov kernels. We extended tools from renewal theory to derive families of time-dependent joint distributions of age and excess, which in turn characterise the probability distribution of censored intervals. We then constructed a model wherein a possibly non-homogeneous point process provides a mechanism to select points on the real line, which are independently marked by the intervals resulting from the censoring mechanism. For this model, a conditional distribution was posited and verified. The influence of the model components was demonstrated through parametrised examples.

In this thesis, we have so far developed theory on stochastic temporal censoring mechanisms for interval-censored data. In Chapter 2, we assumed homogeneity in time, which was subsequently lifted in this chapter to allow for a more complicated model with less assumptions. For interval-censored data of this nature, it seems plausible that spatial effects may also considerably affect the distribution of exact occurrence times. In Chapter 4, we add a spatial component to the model, with the additional constraint that points fall on the edges of a network being inspired by potential applications in criminology. In addition, we discuss rigorous statistical methods for parameter estimation along with uncertainty measures.

CHAPTER 4

Spatio-temporal area-interaction point process models on Euclidean graphs

Chapter 4. Spatio-temporal area-interaction point process models on Euclidean graphs

In this chapter, we focus on state estimation methods for area-interaction point process models on graphs with Euclidean edges, which is a framework for embedding a graph in the Cartesian plane. One example of such a graph is a parametrised linear network, which have been studied within point process theory to a limited extent [14, 44]. We extend the theory developed in [90, 91] and applied in [96], which deal with state estimation of temporal point processes using two different censoring mechanisms, by adding spatial effects.

The complex geometry of graphs with Euclidean edges can hamper the development of point process models, the definition of suitable summary statistics and inference. Some authors [33, 65] investigated spatial prediction methods such as kriging to perform analysis on linear networks, most research to date aimed to find suitable counterparts of classic moment-based summary statistics. For example, the classic kernel-based estimator of the intensity function [37] does not immediately generalise to networks because at each node the kernel's mass must be distributed over all emanating edges. An unfortunate choice for doing so may lead to an estimator that is not mass-preserving and overestimates the denser parts of the network.

The choice of metric (i.e. how to measure distance) on the network should be carefully tailored to the application at hand. Using the shortest path distance, Moradi [105] proposed a local correction factor in the spirit of [87] that is mass preserving and fast to calculate, but that may be biased. The estimator of [111, Sec. 9.3.3] is both unbiased and mass preserving, but at a computational cost. The heat kernel estimator proposed by McSwiggan et al. [98] is a good alternative. For the second-order K-function [122], Okabe and Yamada [112] proposed a network counterpart by replacing the Euclidean distance by the shortest path distance, whereas Ang et al [4] suggested an adjustment based on the network geometry. Both statistics, however, assume an underlying stationarity that does not seem to hold widely [13]. The same remark can be made for inhomogeneity-adjusted K-functions [44], limiting the usefulness of summary statistics in practice quite severely [83].

To date, the literature on models of graphs with Euclidean edges is very sparse, perhaps due to geometry hampering technical conditions necessary for the construction of certain model classes. Van Lieshout developed theory on nearest-neighbour Markov point processes with a neighbourhood structure defined by the graph [89]. Anderes et al. used the resistance metric, an alternative to the shortest path metric, allowing for the use of isotropic covariance models [3]. Tang and Zimmerman developed a toolbox for constructing valid non-separable, parametric space-time covariance models [140], extending and complementing earlier work by Porcu et al. [116]. We refer to the recent review papers [14, 44] for a fuller discussion.

In this paper, we extend the spatio-temporal area-interaction process of [9, 70] (see also [2, 58]) to graphs with Euclidean edges. To do so, the neighbourhood relation is chosen in terms of the metric defined by the edge parametrisations. This model is then combined with aoristic censoring [90, 91] and appropriate covariates to define a full spatio-temporal model for property crimes. Since estimation of the model parameters is hampered by the presence of latent data as well as the fact that the normalisation constant is not available in closed form, we propose a novel estimating equation approach based on the Georgii–Nguyen–Zessin formula [52, 109] which relates the expected sum of a pattern-dependent weight function over the data pattern to a weighted integral of the conditional intensity. In the fully observed case, the idea dates back to Takacs and Fiksel [138, 139], of which Besag's well-known pseudo-likelihood estimator [18] is a special and popular case. Due to the poor performance of certain Takacs–Fiksel estimators in an extensive simulation study [41], it was left aside until Waagepetersen [141] and Baddeley et al [8] suggested a logistic weight function. Other choices were pursued by Cœurjolly et al. [26]. Here, we generalise the approach so that it is able to handle latent information.

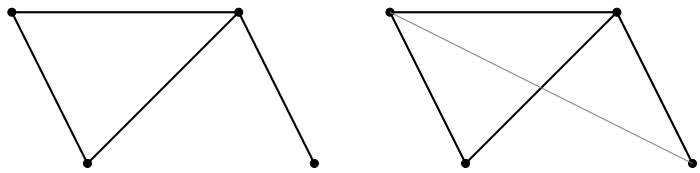
The outline of this chapter is as follows: in Chapter 4.1, we review theory on Euclidean graphs in a point process context and provide some measure-theoretic details. In Chapter 4.2, the network area-interaction point process is introduced and properties are listed. Chapter 4.3 focuses on simulation methods and Bayesian sampling approaches for the full model, which includes both the network area-interaction prior and the temporal censoring mechanism. An outline of parameter estimation methods is provided in Chapter 4.4, with uncertainty quantification methods in Chapter 4.5. A simulated example is provided in Chapter 4.6.

4.1 Parametrised Euclidean graphs

In this chapter, we consider networks in two-dimensional Cartesian space. A network consists of a set of lines. When theory on spatial analysis on linear networks was first developed, a common restriction on the structure of such a network was that lines, also referred to as links or edges, are only connected to each other at endpoints, known as nodes or vertices. In that case, we have that the union L of n lines is defined as

$$L = \bigcup_{i=1}^{n} l_i$$

where $l_i \in \mathbb{R}$, and lines are only connected at endpoints [89, 111]. Since we will be taking inspiration from street networks, this assumption may be too restrictive. There could be "crossing points", which are not technically endpoints of the lines, along which movement can occur. Examples of this may be bridges, underpasses or tunnels. We may also want to model more complex geometries such as curved



(1) (2)

Figure 4.1 Demonstration of different network types. Network (1) only has lines intersecting at endpoints, whereas Network (2) has a crossing away from an endpoint.

roads, or model a situation where flow across line segments is not constant (such as differing speed limits of roads). Additionally, if a union of straight lines is used to model curved roads, edges still do not follow the shortest possible path between two vertices.

This behaviour cannot be described effectively if we assume that a graph is defined simply as a union set in $\mathbb{R}^2$. For these reasons, we will assume that line segments can cross at other points, not just at endpoints. An example of this geometry is shown in Figure 4.1.

4.1.1 Formal graph definition

We will now formalise these notions. A *graph* henceforth refers to a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \Phi)$. The vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ refer to the set of endpoints and lines, from now on called vertices and edges, respectively. Note that vertices and edges are necessarily disjoint. Let $e_i \in \mathcal{E}$ be the ith element of the edge set, and let $v_i^1, v_i^2 \in \mathcal{V}$ be the two elements of the vertex set that are connected by e_i. We assume that edges are undirected. The set $\Phi = \{\phi_i\}$ for i in the index set of $\mathcal{E}$ is a set of homeomorphisms $\phi_i : e_i \cup \{v_i^1, v_i^2\} \to J_i$, where $J_i \subset \mathbb{R}$ is the open set corresponding to points on edge e_i. These homeomorphisms are required as edges may also be curved. Explicitly, $J_i = (j_i^{\min}, j_i^{\max})$ and $\phi_i(v_i^1) = j_i^{\min}, \phi(v_i^2) = j_i^{\max}$. In other words, the vertices v_i^1 and v_1^2, which are connected by the edge e_i are mapped to the boundary of J_i, with all remaining points strictly contained in J_i. In conjunction, we define the inverse mapping $\phi_i^{-1} : J_i \to \mathbb{R}^2$. In this way, we can refer to an edge e_i by the parametrisation $\phi^{-1}(J_i)$. The union of edges and vertices can be written as

$$L = (\{0\} \times \mathcal{V}) \cup \bigcup_{i=1}^{|\mathcal{E}|} (\{i\} \times \phi_i^{-1}(J_i)),$$

taking care to separate different edge sets by index [89]. Note that $\phi^{-1}(J_i)$ refers to the entire set of points on the edge e_i, meaning that if an arbitrary element (i, u) is selected from $\phi^{-1}(\cdot)$ where $i \neq 0$, this refers to a particular point on the

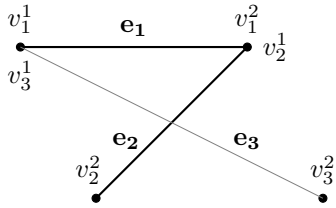


Figure 4.2 View of a simple graph with an overlapping edge, using both sets of notation introduced earlier. The index is used to distinguish edges and edge sets in the case of overlapping. Vertices are indexed multiple times if multiple edges radiate from them.

edge e_i. For an illustration of notation, we direct the reader to Figure 4.2 and Example 4.1.

Example 4.1 (The trivial graph)**.** *Let*

$$\mathcal{G} = (\{(0,0),(1,1)\}, (\{(0,0),(1,1)\}), \{\phi_1\}).$$

We define $\phi_1^{-1} : (0,1) \to \mathbb{R}^2$ *and* $\phi_1^{-1}(t) = (t,t)$*. The parametrisation* ϕ_1 *corresponds to a straight edge between vertex* $(0,0)$ *and vertex* $(1,1)$*. If instead* $\hat{\phi}_1^{-1}(t) = (t, \sqrt{t})$*, the edge becomes a parabolic arc between vertex* $(0,0)$ *and vertex* $(1,1)$*.*

4.1.2 Weighted shortest path metric

A *path* between $(i,u) \in L$ and $(j,v) \in L$ is represented by a list where the first and last element respectively refer to the start and end of the path, and all elements in between are unique vertices. If $i \neq 0$ or $j \neq 0$, the path starts at (i,u) and goes along $\phi_i^{-1}(J_i)$ first before eventually reaching $\phi_i^{-1}(J_j)$, where the path terminates at the point (j,v). This gives the path $(i,u),(0,v_i^p),\ldots,(0,v_j^q),(j,v)$ for some $p,q \in \{1,2\}$. Note that this path may not be unique, since there may exist multiple paths between two elements of L. We assume that the graph is connected, in which case for any two elements $(i,u),(j,v) \in L$, there exists a finite path $(i,u),\ldots,(j,v)$ connecting the two.

The weight of an edge e_i is defined as the Euclidean distance between the two endpoints of J_i. Recalling that these two endpoints are $\phi_i(v_i^1)$ and $\phi(v_i^2)$, the weight of e_i can be written as $|\phi_i(v_i^1) - \phi_i(v_i^2)|$. We define the shortest path distance $d_G((i,u),(j,v))$ on G to be the distance that minimises the weight of the path taken across all possible edges between (i,u) and (j,v), for $i,j \in 0,1,\ldots,n(\mathcal{E})$. Changing notation for ease of explanation, assume that a path of length n connects (i,u_1) and (j,u_n).

Definition 4.2. *Define the shortest path as* $(i,u_1),(0,u_2),\ldots,(0,u_{n-1}),(j,u_n)$*.*

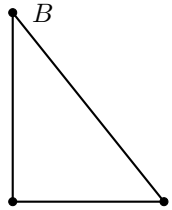


Figure 4.3 Towns A, B and C.

Notation: d is distance, v speed limit and T traversal time.
Road (A,C): $d = 3$ km, $v = 30$km/h, $T = \frac{1}{10}$h (6 min)
Road (A,B): $d = 4$ km, $v = 20$km/h, $T = \frac{1}{5}$h (12 min)
Road (B,C): $d = 5$ km, $v = 60$km/h, $T = \frac{1}{12}$h (5 min)

For any alternate path $(i, \bar{u}_1), (0, \bar{u}_2), \ldots, (0, \bar{u}_{m-1}), (j, \bar{u}_m)$ *of length* m*, where* $(i, u_1) = (i, \bar{u}_1)$ *and* $(i, u_n) = (i, \bar{u}_m)$*, the weighted shortest path metric is*

$$d_G((i, u_1), (j, u_n)) = \sum_{i=1}^{n} |\phi_i(u_{i+1}) - \phi_i(u_i)| \tag{4.1}$$

only if

$$\sum_{i=1}^{n} |\phi_i(u_{i+1}) - \phi_i(u_i)| \leq \sum_{i=1}^{m} |\phi_i(\bar{u}_{i+1}) - \phi_i(\bar{u}_i)|$$

for all possible alternate paths.

The lengths of the edges are modulated by the length induced by ϕ_i^{-1}, meaning that d_G is indeed a weighted shortest path distance and known as the geodesic or weighted shortest path metric [3, 89]. We proceed by introducing a more detailed example.

Example 4.3 (Pythagorean towns)**.** *Imagine that there are three towns* A*,* B *and* C*. Town* A *is located at* $(0, 0)$*, town* B *at* $(0, 4)$ *and town* C *is at* $(3, 0)$*. Assume that there are three straight, speed-limited roads connecting all three towns. We wish to parametrise this graph in terms of the edges' traversal times. We have the following setup: Hence* $\mathcal{V} = \{(0, 0), (3, 0), (0, 4)\}$ *and* $\mathcal{E} = (\{(0, 0), (3, 0)\}), \{(0, 0), (0, 4)\}, \{(0, 4), (3, 0)\}$*. For parametrisation, we let* t*, the time elapsed, range over the time interval that is traversed when following each line. Therefore, the parametrisations are* $\Phi = \{\phi_1, \phi_2, \phi_3\}$ *where* $\phi_1^{-1} : (0, \frac{1}{10}) \to \mathbb{R}^2$*,* $\phi_2^{-1} : (0, \frac{1}{5}) \to \mathbb{R}^2$ *and* $\phi_3^{-1} : (0, \frac{1}{12}) \to \mathbb{R}^2$*. Specifically,*

$$\phi_1^{-1}(t) = (30t, 0), \quad \phi_2^{-1}(t) = (0, 20t), \quad \phi_3^{-1}(t) = (36t, -48t + 4).$$

The weights of the roads are exactly their traversal times; hence road (A, C) *has weight* $\frac{1}{10}$*, road* (A, B) *has weight* $\frac{1}{5}$ *and road* (B, C) *has weight* $\frac{1}{12}$*.*

In Example 4.3, if we want to travel from town A to town B, there are two possible routes: travelling there directly, or going via town C. It might seem logical to

take the shortest route by distance, which is the direct route. However, due to the low speed limit on this road, it is actually faster to travel via C. This can be seen clearly when looking at the weights/traversal times: $\frac{1}{10} + \frac{1}{12} < \frac{1}{5}$. Letting the weight depend on J_i in this manner thus allows for a powerful modelling approach that is potentially decoupled from Euclidean distance between the actual vertices in the graph.

Formally, for more complex applications on road networks, weights w_i can be defined by utilising the end vertices v_i^1 and v_i^2 as well as the speed limit s_i. The ϕ-parametrisation becomes

$$\phi_i^{-1} : (0, w_i||v_i^1 - v_i^2||) \to \mathbb{R}^2, \quad \phi_i^{-1}(t) = v_i^1 + t\frac{(v_i^2 - v_i^1)}{w_i||v_i^1 - v_i^2||}.$$

One possible choice of weight might be $w_i = 1/s_i$, which represents travel time across a line segment. Another possibility would be the trivial weight choice $w_1 \equiv 1$.

4.1.3 Measure theoretic details

Since edges may not necessarily be straight lines, we need to formulate the notion of measure over L carefully. For every i in $1, \ldots, n(\mathcal{E})$, we have a corresponding σ-algebra $\mathcal{A}_i = \{\phi_i^{-1}(B) : B \subset \bar{J}_i \subset \mathbb{R}^2\}$ for a Borel set B, noting that $\bar{J}_i$ refers to the closure of J_i. We let λ_G^i be the ϕ_i-length measure on $\mathcal{A}_i$ such that for a given $A \in \mathcal{A}_i$ we have

$$\lambda_G^i(A_i) = \int_{J_i} \mathbf{1}\{\phi_i^{-1}(u) \in A_i\} \left| \frac{d}{du}\phi_i^{-1}(u) \right| du,$$

using the definition of the Lebesgue-Stieltjes integral and [3, 89]. Note that in this case, $|\cdot|$ refers to the norm, and that we therefore require differentiability for $u \in J_i$ for all i. For meaningful application, we require that $\lambda_G^i(\bar{J}_i) < \infty$ for all $i \in 1, \ldots, n(\mathcal{E})$. Construct the union σ-algebra

$$\mathcal{A} = \left\{ \bigcup_{i=0}^{n(\mathcal{E})} (i \times A_i) : A_0 = 2^{\mathcal{V}}, A_i \in \mathcal{A}_i \text{ when } i = 1, \ldots, n(\mathcal{E}) \right\}.$$

We define, for $A \in \mathcal{A}$, which can therefore be written as $A = \bigcup_{i=0}^{n(\mathcal{E})}(i \times A_i)$,

$$\lambda_G(A) = \sum_{i=1}^{n(\mathcal{E})} \lambda_G^i(A_i). \tag{4.2}$$

In analogy to [89], this is a measure on L and thus we define a measure space $(L, \mathcal{A}, \lambda_G)$ on the graph G. In this way, points can be distributed on the graph. A detailed explanation is provided in Chapter 4.2.1 and [3, 89]. In the next section, we define a spatio-temporal area-interaction process on the graph structure outlined in this section.

4.2 The spatio-temporal area-interaction process

Having outlined the network geometry, we turn now to point process models. Various point process models exist that allow for inference in the spatio-temporal paradigm, such as the spatio-temporal Neyman-Scott process [108], the spatio-temporal log-Gaussian Cox process [42, 101] the Hawkes processes [97], among many others (see [57] for a comprehensive overview). This chapter will focus on Markov point processes, as they not only provide a density with respect to a unit-rate Poisson process, but allow for interactions to be modelled in an intuitive manner via a neighbourhood relation and the Hamersley-Clifford theorem [11, 86, 123]. We will specifically be focusing on the spatio-temporal area-interaction process [9, 70]. First, we define the spatio-temporal Poisson point process on a linear network.

4.2.1 Spatio-temporal Poisson point processes on linear networks

Authors such as [3, 89] discuss the construction of a Poisson point process on a graph with Euclidean edges in their works. We wish to extend to the spatio-temporal paradigm. Recall the measure space $(L, \mathcal{A}, \lambda_G)$. Define a Borel-measurable open time window $W_T \subset \mathbb{R}^+$ and take the product measure $\mu = \lambda_G \times \ell$, where ℓ is one-dimensional Lebesgue measure acting on the measurable space $(W_T, \mathcal{B}(W_T))$. By the Hahn-Kolmogorov theorem, this product measure exists, and we obtain the measure space $(L \times W_T, \mathcal{A} \otimes \mathcal{B}(W_T), \mu)$. For every $B = ((i, A_i), D) \in \mathcal{A} \otimes W_T$ where A_i is contained in the σ-algebra generated by ϕ_i and $D \in \mathcal{B}(W_T)$, for a counting measure K as in [34],

$$\mathbb{E}(N((\{i\} \times A_i), D)) = |D|\lambda_G(\{i\} \times A_i)$$

where $|D|$ represents the length of D, and hence

$$\mathbb{P}(N((\{i\} \times A_i), D) = n_i) = \frac{(|D|\lambda_G(\{i\} \times A_i))^{n_i} e^{-|D|\lambda_G(\{i\} \times A_i)}}{n_i!}.$$

Additionally, conditioning on $N((\{i\} \times A_i), D) = n_i$, the n_i are i.i.d. with probability density $1/(|D|\lambda_G(\{i\} \times A_i))$. Finite-dimensional distributions are defined as in [34].

From a practical point of view, realisations of this Poisson point process can be obtained as follows. Use rejection sampling to pick edge i with probability $\lambda_G^i(\phi_i^{-1}(J_i))/\lambda_G(L)$. Then, sample a point j_i uniformly at random on J_i and apply the function ϕ_i^{-1} to obtain a random point on the curve $\phi_i^{-1}(J_i)$.

4.2.2 Network area-interaction process definition

We adapt the spatio-temporal area-interaction point process [9] to take values on the space $L \times W_T \subset (\mathbb{R}^+ \times \mathbb{R}^2) \times \mathbb{R}^+$. This adapted spatio-temporal area-

interaction process has density, with respect to a unit rate Poisson process on $L \times W_T$ (see Chapter 4.2.1), of

$$p(\mathbf{x}) = \alpha_p \left(\prod_{((i,u),t)\in\mathbf{x}} \beta((i,u),t) \right) \gamma^{-\mu(U_{r,h}(\mathbf{x}))},$$

where $x = ((i,u),t)$ refers to a point on a Euclidean graph at time $t \geq 0$. If $i = 0$, the point x is located on a vertex of the graph, in all other cases the point is located on an edge. For $\gamma < 1$, behaviour is regular, for $\gamma > 1$, behaviour is clustered, and $\gamma = 1$ corresponds to a non-homogeneous Poisson process with intensity function β. The function $\beta((i,u),t) \geq 0$ is a measurable and bounded function on $L \times W_T$, and is related to the intensity of the point process at location $x = ((i,u),t) \in \mathbf{x}$. If $\beta(x) = \beta$ for all x, then the term in brackets simplifies to $\beta^{n(\mathbf{x})}$, where $n(\mathbf{x})$ refers to the number of points in $\mathbf{x}$. The parameter h is the radius of interaction in time, r is the radius of interaction in space, and $\alpha_p = c(\beta(\cdot),\gamma)$ represents a normalisation constant. Covariates can be modelled by setting $\beta(z,t) = \rho(Z(z,s))$, where $Z(z,s)$ corresponds to the covariate vector [70] and $z = (i,u)$.

In applications such as [70],

$$U_{r,h}(\mathbf{x}) = \bigcup_{(z,s)\in\mathbf{x}} \mathcal{C}_r^h(z,s),$$

the union of cylinders with interaction radius r and height h, centred at (z,s). In the case of a Euclidean graph, it may not be desirable to allow interactions outside of the network structure. On a graph, analogous to the standard $\mathbb{R}^2$ case, we define

$$U_{r,h}(\mathbf{x}) = \bigcup_{((i,u),t)\in\mathbf{x}} L_{r,h}((i,u),t),$$

where

$$L_{r,h}((i,u),t) = \{((j,v),s) \in L \times W_T : d_G((i,u),(j,v)) \leq r \text{ and } t-h \leq s \leq t+h\},$$

with slight abuse of notation. In words, the set $L_{r,h}((i,u),t)$ includes all points that are at most r away from the point $((i,u),t)$ by weighted shortest-path distance, and occurred between times $t-h$ and $t+h$. The natural interpretation of r and h are as spatial and temporal radii respectively. In effect, this means that the interaction area is a shape with base set $L_{r,h}$ made up of an intersection of arcs, and height $2h$, centred at $u \in \mathbb{R}^2$ and $t \in \mathbb{R}^+$. To see an illustration of the interaction area, see Figure 4.4. Note that this is not equivalent to the intersection of the cylinder with radius r and height t with all edges that overlap with the base of the cylinder, since there is no guarantee that points on overlapping

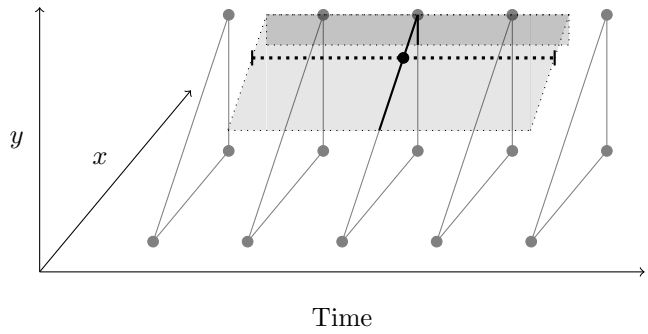


Figure 4.4 Diagram showing the interaction area, represented by the two shaded planes, of a point on a triangular Euclidean graph. Note that lines in a Euclidean graph need not be straight (constant slope), and are only straight here for demonstration purposes. The graph is plotted in the Cartesian plane with an extra time axis. The black interval represents the interval within which the point can occur in time.

edges are at most r away from (i, u) in space. We can rewrite the density of the area-interaction process on a Euclidean graph as

$$p(\mathbf{x}) = \alpha_p \left(\prod_{((i,u),t)\in\mathbf{x}} \beta((i,u),t) \right) \gamma^{-\mu(U_{r,h}(\mathbf{x}))} \tag{4.3}$$

for $\mu = \lambda_G \times \ell$.

4.2.3 Properties of the network area-interaction process

Let $\sim$ be a symmetric and reflexive relation on $L \times W_T$. In other words, for any $((i,u),t),((j,v),s) \in L \times W_T$, we have $((i,u),t) \sim ((j,v),s) \Leftrightarrow ((j,v),s) \sim ((i,u),t)$ and $((i,u),t) \sim ((i,u),t)$. If it is the case that $((i,u),t) \sim ((j,v),s)$, these two points are considered *neighbours*.

Definition 4.4. *The neighbourhood relation is defined as*

$$\begin{aligned}((i,u),t) \sim ((j,v),s) \Leftrightarrow d_G((i,u),(j,v)) &\leq 2r \\ &\Leftrightarrow L_{r,h}((i,u),t) \cap L_{r,h}((j,v),s) \neq \emptyset,\end{aligned}$$

in analogy to [70].

In other words, points are neighbours if there is any overlap between their interaction regions.

Definition 4.5. *A point process has the Markov property [86, 123] with respect to $\sim$ if, for all $\mathbf{x}$ when $p(\mathbf{x}) > 0$,*

1. $p(\mathbf{y}) > 0$ *whenever* $\mathbf{y} \subset \mathbf{x}$,

2. *the Papangelou conditional intensity* $\frac{p(\mathbf{x} \cup \{((j,v),s)\})}{p(\mathbf{x})}$ *(see [114]) when adding a point* $((j,v),s)$ *to the pattern only depends on those* $((i,u),t) \in \mathbf{x}$ *that are in the same clique as* $((j,v),s)$.

Proposition 4.6. *The spatio-temporal area-interaction process defined on the Euclidean graph as outlined in Chapter 4.1 has the Markov property.*

Proof. It suffices to show that properties *1* and *2* of Definition 4.5 are satisfied. For property *1*, assume that there is some configuration $\mathbf{y} \subset \mathbf{x}$ for which $p(\mathbf{y}) = 0$. Due to the exponential term in p, this means that necessarily $\beta((i,u),t) = 0$ for at least one $((i,u),t) \in L \times W_T$. If a point ξ is added to this configuration,

$$\prod_{((i,u),t) \in \mathbf{y} \cup \{\xi\}} (\beta((i,u),t)) = 0$$

and hence for any choice of $\mathbf{x}$, $p(\mathbf{x}) = 0$ when $p(\mathbf{y}) = 0$ and $\mathbf{y} \subset \mathbf{x}$. Therefore, by contradiction, property *1* is satisfied. For property *2*, we have

$$\begin{aligned}
&\frac{p\left(\mathbf{x} \cup \{((j,v),s)\}\right)}{p(\mathbf{x})} \\
&= \frac{\alpha_p \left(\prod_{((i,u),t) \in \mathbf{x}} \beta((i,u),t)\right) \beta((j,v),s) \gamma^{-\mu[U_{r,h}(\mathbf{x}) \cup L_{r,h}((j,v),s)]}}{\alpha_p \left(\prod_{((i,u),t) \in \mathbf{x}} \beta((i,u),t)\right) \gamma^{-\mu(U_{r,h}(\mathbf{x}))}} \\
&= \beta((j,v),s) \gamma^{-\mu(L_{r,h}((j,v),s) \setminus U_{r,h}(\mathbf{x}))},
\end{aligned}$$

with slight abuse of notation. Using the fact that $A \setminus B = A \cap B^C$, we obtain

$$\begin{aligned}
L_{r,h}((j,v),s) \setminus U_{r,h}(\mathbf{x}) &= L_{r,h}((j,v),s) \cap \left[\bigcup_{((i,u),t) \in \mathbf{x}} (L_{r,h}(i,u),t))\right]^C \\
&= L_{r,h}((j,v),s) \cap \left[\bigcup_{((i,u),t) \sim ((j,v),s) \in \mathbf{x}} (L_{r,h}(i,u),t))\right]^C.
\end{aligned}$$

As only points in the clique of $((j,v),s)$ influence the likelihood, (b) is satisfied and (4.3) is a Markov point process with respect to the relation $\sim$. □

The Papangelou conditional intensity assumes that $((j,v),s) \notin \mathbf{x}$ and $p(\mathbf{x}) > 0$. Formally,

$$\lambda(((j,v),s) \,|\, \mathbf{x}) = \frac{p\left(\mathbf{x} \cup \{((j,v),s)\}\right)}{p(\mathbf{x})}$$

$$= \beta((j,v),s)\gamma^{-\mu(L_{r,h}((j,v),s)\setminus U_{r,h}(\mathbf{x}))}.$$

Intuitively, the expression $\mu(L_{r,h}((j,v),s) \setminus U_{r,h}(\mathbf{x}))$ refers to the area of points that is not already contained in the union of $L_{r,h}(x)$ over all $x \in \mathbf{x}$ when a new point $((j,v),s) \in L \times W_T$ is added to the process.

Proposition 4.7. *The spatio-temporal area-interaction process defined on the Euclidean graph as outlined in Chapter 4.2 allows for interactions of infinite order, as described in [11].*

Proof. To do this, we use the Hamersley-Clifford theorem [123] to show that the density can be written as a product of non-negative, measurable interaction functions $\psi_0, \psi_1, \ldots$. More specifically,

$$p(\mathbf{x}) = \prod_{\mathbf{y}\subseteq\mathbf{x}} \psi_{|\mathbf{y}|}(\mathbf{y}),$$

where $\psi(\mathbf{y}) = 1$ except when $y_i \sim y_j$ for all $i,j \in 0,\ldots,|\mathbf{y}|$. We know that $U_{r,h}(\mathbf{x}) = \cup_{((i,u),t)\in\mathbf{x}} L_{r,h}((i,u),t)$. Let $\mathbf{x} = \{((i_1,u_i),t_1),\ldots,((i_n,u_n),t_n)\}$. We can write, without loss of generality,

$$\begin{aligned}
\mu(U_{r,h}(\mathbf{x})) &= \mu\left(\bigcup_{((i,u),t)\in\mathbf{x}} L_{r,h}((i,u),t)\right) = \mu(L_{r,h}((i_1,u_1),t_1)) + \ldots \\
&+ \mu(L_{r,h}((i_n,u_n),t_n)) - \mu(L_{r,h}((i_1,u_1),t_1) \cap L_{r,h}((i_2,u_2),t_2)) \\
&- \cdots - \mu(L_{r,h}((i_{n-1},u_{n-1}),t_{n-1}) \cap L_{r,h}((i_n,u_n),t_n)) \\
&+ \cdots + (-1)^{n+1}\mu(L_{r,h}((i_1,u_1),t_1) \cap \cdots \cap L_{r,h}((i_n,u_n),t_n)) \\
&= \sum_{k=1}^{n} \mu(L_{r,h}((i_k,u_k),t_k)) - \sum_{k<l} \mu(L_{r,h}((i_k,u_k),t_k) \cap L_{r,h}((i_l,u_l),t_l)) \\
&+ \cdots + (-1)^{n+1}\mu(\bigcap_{k=1}^{n} L_{r,h}((i_k,u_k),t_k))
\end{aligned}$$

via the inclusion-exclusion principle. Based on this, we can deduce that

$$\begin{aligned}
\psi_0(\emptyset) &= \alpha_p, \\
\psi_1(\{y\}) &= \beta((i,u),t)\gamma^{-\mu(L_{r,h}((i,u),t))} \\
\psi_m(\{y_1,\ldots,y_m\}) &= \gamma^{(-1)^m \mu(\bigcap_{i=1}^{m} L_{r,h}((i_m,u_m),t_m))}
\end{aligned}$$

since

$$\prod_{\mathbf{y}\subseteq\mathbf{x}} \psi_{|\mathbf{y}|}(\mathbf{y}) = \psi_0(\emptyset) \times \psi_1(\{y_1\}) \times \cdots \times \psi_1(\{Y_k\})$$

$$\times \psi_2(\{y_1, y_2\}) \times \cdots \times \psi_{|\mathbf{y}|}(\{y_1, \ldots, y_{|\mathbf{y}|}\}) = p(\mathbf{x})$$

for the provided values of ψ. Therefore, the process has interactions of infinite order. □

The Markov property is particularly useful for sampling by Monte Carlo methods that add or delete one point at a time, as the likelihood ratio for such operations can be calculated locally. For instance, dominated coupling from the past (CFTP), as in [2, 10, 79, 118], can be used to generate realisations from Equation 4.3. In the next subchapter, these simulation and sampling methods will be discussed.

4.3 Simulation methods and sampling approaches

Having determined the conditional intensity of an area-interaction model on a Euclidean graph and shown that this model is indeed a Markov point process, we now move on to estimation methods. As the spatio-temporal point process that we wish to model is temporally censored, the most likely location in time must be inferred using methods in point process state estimation [88, 90]. We now extend this to the spatial domain.

4.3.1 Modelling considerations

In practice, a number of considerations must be taken in the modelling phase. Vertices, as well as the start and end vertices of edges, are often provided when performing analysis on linear networks. However, the homeomorphism set Φ of the Euclidean graph must be modelled, since representing curved lines is non-trivial. One approach is to model an arc by a number of straight line segments. This involves picking points near to or on the arc, and connecting these using straight lines. In practice, curved line segments are often already approximated by straight line segments. Using this approach allows for existing data to be seamlessly integrated into the model, but may carry a computational cost, as points between vertices must be stored for gradients to be calculated correctly for intermediate line segments. Additionally, this approach may be useful if a functional form for the arc does not exist.

Another method by which Φ can be modelled is by using polynomial or harmonic line parametrisations. Take the trivial graph with only one edge and two vertices (see Example 4.1), with the vertices $(0, 0)$ and $(1, 1)$ connected by a single curved edge. One potential model for this curve might be parabolic: that is, $\phi_1^{-1}(t) = (t, \sqrt{t})$ for $t \in (0, 1)$. Additionally, one could use a harmonic model, which would

correspond to $\phi_1^{-1}(t) = (\cos^{-1}(t), \sin^{-1}(t))$ for $t \in (0,1)$. In this way, nonlinear homeomorphisms can be modelled directly, saving space.

4.3.2 Perfect area-interaction model simulation

Before simulation methods for the full spatio-temporal model can be described, we must first tackle the problem of simulation for the modified area-interaction process. To do this, an adapted dominated coupling from the past (CFTP) algorithm, as in [1, 78], can be used. This algorithm was originally developed by [118] in 1996, and allows for the equilibrium distribution of a Markov chain to be sampled from by using multiple coupled chains with different initial states based on the Papangelou conditional intensity [69]. The Papangelou conditional intensity of the area-interaction process extended to the linear network, when adding a point $\xi = ((j, v), s)$ is

$$\lambda(\xi \,|\, \mathbf{x}) = \frac{p\,(\mathbf{x} \cup \{\xi\})}{p(\mathbf{x})} = \beta(\xi)\gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))} \tag{4.4}$$

for $p(\mathbf{x}) \neq 0$.

Spatial birth and death processes

We would like an algorithm which generates points with rate exactly $\lambda(x \,|\, \mathbf{x})$ for $x = ((j, v), s) \in L \times W_T$. To do this, we make use of spatial birth- and death processes [117, 122]. To generate the initial point pattern to which points are either added or deleted, we use a spatio-temporal Poisson point process as outlined in Chapter 4.2.1. Let the state space Ω_n be made up of all possible configurations $\mathbf{x}$ for each n.

Definition 4.8 (Preston, 1977)**.** *Define the transition rates $D(\mathbf{x} \setminus \{x_i\}, x_i)$ (death) and $b(\mathbf{x}, u)$ (birth) as*

$$D(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} D(\mathbf{x} \setminus \{x_i\}, x_i) \qquad B(\mathbf{x}) = \int_{L \times W_T} b(\mathbf{x}, \xi)\, \mu(d\xi)$$

for $\xi \in L \times W_T$ [9, 11]. Define $\theta_n = \sup_{n(\mathbf{x})=n} B(\mathbf{x})$ and $d_n = \inf_{n(\mathbf{x})=n} D(\mathbf{x})$. If the following hold:

- *(Detailed balance) Whenever $p(\mathbf{x} \cup \xi) > 0$, for some b and D, we require*

$$b(\mathbf{x}, \xi)p(\mathbf{x}) = D(\mathbf{x} \setminus \{x_i\}, x_i)p(\mathbf{x} \cup \xi) > 0, \text{ and}$$

- *(Non-explosion, case 1) $\theta_n = 0$ for all sufficiently large n and $d_n > 0$ for all $n \geq 1$, or*

- *(Non-explosion, case 2)* $\theta_n > 0$ *for all* $n \geq 0$, $d_n > 0$ *for all* $n \geq 1$, *and*

$$\sum_{n=1}^{\infty} \frac{\theta_0 \dots \theta_{n-1}}{d_1 \dots d_n} < \infty, \qquad \sum_{n=1}^{\infty} \frac{d_1 \dots d_n}{\theta_1 \dots \theta_n} = \infty,$$

there exists a unique spatial birth and death process with the provided rates, and the process converges in distribution from any configuration of $\mathbf{x}$ *to the equilibrium measure* $p(\cdot)$.

We now need to show that these conditions are satisfied for the network area-interaction process. We consider a process with birth rate equal to the Papangelou conditional intensity of adding a point ξ to the process and a constant death rate. More precisely,

$$b(\mathbf{x}, \xi) = \begin{cases} \lambda(\xi \,|\, \mathbf{x}) = \beta(\xi)\gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))}, & p(\mathbf{x}) > 0 \\ 0, & p(\mathbf{x}) = 0 \end{cases} \tag{4.5}$$

and $D(\mathbf{x} \setminus \{x_i\}, x_i) = 1$. Detailed balance is satisfied as in the general area-interaction case (see [9, Chapter 4] and [11, Theorem 2.10] for more details).

For the non-explosion conditions, we use the same birth and death rates as in the previous case. If $\theta_n > 0$ for some n, then $p(\mathbf{x}) > 0$ for some configuration $\mathbf{x}$. For all $\mathbf{y} \subset \mathbf{x}$ we also have $p(\mathbf{y}) > 0$ by the hereditary property, and therefore $\theta_m > 0$ for $m < n$. If eventually $\theta_n = 0$ for sufficiently large n, case *1* is satisfied automatically. If not, then $\theta_n > 0$ for all n, hence $b(\mathbf{x}, \xi) = \beta(\xi)\gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))}$ always.

We now consider parameter ranges of γ, corresponding to attractive or repulsive interaction of points. For the attractive case, $\gamma \geq 1$, noting that $\gamma = 1$ implies a fully random interaction structure. Then $\beta(\xi)\gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))} \leq \beta(\xi)$ as long as $\beta(\xi) \geq 0$, which is automatically true since all probability distributions are non-negative. Let $\bar{\beta} = \sup_{\mathbf{x} \in \Omega} B(\mathbf{x})$. Then

$$B(\mathbf{x}) = \int_{L \times W_T} b(\mathbf{x}, \xi)\, \mu(d\xi) \leq \bar{\beta}$$

as long as β is integrable over A, meaning that $\theta_n \leq \bar{\beta}$ for any n. Clearly

$$\sum_{n=1}^{\infty} \frac{\theta_0 \dots \theta_{n-1}}{\delta_1 \dots \delta_n} \leq \sum_{n=1}^{\infty} \frac{[\bar{\beta}]^n}{n!} < \infty, \qquad \sum_{n=1}^{\infty} \frac{\delta_1 \dots \delta_n}{\theta_1 \dots \theta_n} \geq \sum_{n=1}^{\infty} \frac{n!}{[\bar{\beta}]^n} = \infty.$$

For the repulsive case, i.e. $0 < \gamma < 1$, the bound is $\beta(\xi)\gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))} \leq \beta(\xi)\gamma^{-\mu(U_{r,h}(\mathbf{x}))}$, and by the same argument, the summability conditions are satisfied. In summary, for both ranges of γ, non-explosivity condition *2* holds if

$\theta_n > 0$ for all n, and condition *1* holds otherwise. This is the case as long as β is non-negative and integrable over $A \subseteq L \times W_T$ with respect to μ. When these conditions are satisfied, there exists a unique spatial birth and death process with the proposed rates that converges in distribution to p, and this spatial birth and death process is the process outlined in this section.

Coupling from the past

We now move on to coupling from the past. We first require that the network area-interaction process is stochastically dominated by a reference Poisson point process of a given intensity. This is the case, as the Papangelou conditional intensity of the network area-interaction process (see equation 4.4) is bounded by the non-homogeneous Poisson point process of intensity $\beta(\xi)$ if $\gamma \geq 1$, or $\beta(\xi)\gamma^{-\mu(U_{r,h}(\mathbf{x}))}$ if $0 < \gamma < 1$, for $\xi \in L \times W_T$. Noting that the γ term is also bounded, stochastic domination is guaranteed. We therefore desire a realisation of the Poisson point process presented in Chapter 4.2.1. While choosing a random point on an edge i can be performed by using the corresponding inverse homeomorphism ϕ_i^{-1}, we must first deal with the issue of picking an edge of the graph at random. Equation 4.2 tells us that the measure λ_G is exactly the sum of i-length measures over corresponding edge sets. Therefore, we can use rejection sampling to pick edge i with probability $\lambda_G^i(A_i)/\lambda_G(A)$, where A_i and A are as in Chapter 4.1.1. This approach will lead to the generation of a non-homogeneous Poisson point process over the graph G. We proceed with the following algorithm [1, 78, 86, 118]:

Algorithm 4.9. *Let $B_{t,\xi}$ be a collection of i.i.d. $U(0,1)$ random variables, where $t \leq 0$ and $\xi \in L \times W_T$. Set $T = 1$ initially, and generate a sample $Z(0)$ of a non-homogeneous Poisson point process with intensity function $\beta(\xi) \times \gamma^{-\mu(U_{r,h}(\mathbf{x}))}$ if $0 < \gamma < 1$, and intensity function $\beta(\xi)$ if $\gamma \geq 1$. Then:*

- *The process Z is extended until time $-T$ using the spatial birth- and death process with birth rate $b(\cdot, \xi)$ from (4.5) and constant death rate,*
- *we generate two time-evolving processes Y^{min}, the lower process, and Y^{max}, the upper process, with the initial values of $Y^{min}(-T, -T) = \emptyset$, $Y^{max}(-T, -T) = Z(-T)$, where $Z(-T)$ is the extended process from the previous step,*
- *for any $t \in (-T, 0]$, if a point d was a birth in the extension of Z in reverse time, i.e. $Z(t) = Z(t-) \setminus \{d\}$ for time $t-$ just before t, the point d is a death in both $Y^{min}(-T, t-)$ and $Y^{max}(-T, t-)$,*
- *if a death occurred in the extension of the process Z in reverse time, so $Z(t) = Z(t-) \cup \{\xi\}$, we have two cases:*

1. *If* $\gamma \geq 1$, *we add the point* ξ *to* Y^{max} *only if* $B_{t,\xi} \leq \lambda(\xi \,|\, Y^{max}(-T, t-))/\beta(\xi)$, *and add* ξ *to* $Y^{min}(-T, t-)$ *only if* $B_{t,\xi} \leq \lambda(\xi \,|\, Y^{min}(-T, t-))/\beta(\xi)$.

2. *If* $0 < \gamma < 1$, *we add the point* ξ *to* Y^{max} *only if* $B_{t,\xi} \leq \lambda(\xi \,|\, Y^{min}(-T, t-))/\beta(\xi)\gamma^{-\mu(U_{r,h}(\mathbf{x}))}$, *and add* ξ *to* $Y^{min}(-T, t-)$ *only if* $B_{t,\xi} \leq \lambda(\xi \,|\, Y^{max}(-T, t-))/\beta(\xi)\gamma^{-\mu(U_{r,h}(\mathbf{x}))}$.

- *At time* 0, *if* $Y^{min}(-T, 0) = Y^{max}(-T, 0)$, *coalescence has been achieved and either* Y^{min} *or* Y^{max} *are returned. Otherwise we set* $T = 2T$ *and repeat.*

For correctness, we sketch the standard proofs in [79, 86]. First, we consider the case where $\gamma \geq 1$. Here, the addition scheme ensures that points are born at rate $\beta(\xi) \times \gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))}$. We additionally have that $\beta(\xi)\gamma^{-\mu(L_{r,h}(\xi) \setminus U_{r,h}(\mathbf{x}))} \leq \beta(\xi)$ as long as we have bounds on β as in Chapter 4.3.2, meaning that the process is monotone and bounded. Due to this stochastic domination of the non-homogeneous Poisson point process, we eventually reach the target distribution. In the repulsive case, $\lambda(\cdot)$ is no longer an upper but a lower bound on the birth rate, that is to say: $\beta(\xi)\gamma^{-\mu(U_{r,h}(\mathbf{x}))} > \beta(\xi)$. In this case, if we add points to the lower process according to the acceptance probability of the upper process, and vice versa, we end up with stochastic monotonicity as in the first case, and the target distribution can be reached in finitely many steps [78].

4.3.3 Sampling for applications with temporal censoring

The CFTP algorithm described in the previous section allows us to generate a sample of the network area-interaction process. This specifies a configuration of n points $\mathbf{x} = \{x_1, \ldots, x_n\}$, where each $x_i \in L \times W_T$ for $i \in 1, \ldots, n$. This configuration follows the probability law $p(\mathbf{x})$ with respect to a unit-rate Poisson process. A potential challenge faced when applying a model such as this one to real-life problems is that the exact time of an event is often unknown, and in that case often censored in some way. For example, police generally know where a property crime occurs, but are not able to pinpoint an occurrence time beyond an interval, which is typically the time between which the victims left the crime scene and when they returned [96, 120].

Model building

We assume that we are working with temporally interval-censored data. In that case, a Metropolis-Hastings algorithm with a fixed number of points is best suited to determine the most likely configuration of points in time [23, 90, 102]. Assume that we have a point process X such that a realisation is of the form $\mathbf{x} = \{x_1, \ldots, x_n\} = \{(z_1, t_1), \ldots, (z_n, t_n)\} \subset L \times W_T$, where we use z to rep-

resent a spatial location $(i, u) \in L$ and t to represent a time in W_T. Following [90, 91], we construct a full marked point process W. It has realisations $\mathbf{w} = \{w_1, \ldots, w_n\} = \{((z_1, t_1), (a_1, l_1)), \ldots, ((z_n, t_n), (a_n, l_n))\}$ and takes values on the exponential space $\mathcal{N}_{L \times W_T \times (\mathbb{R} \times \mathbb{R}^+)}$, the set of all possible configurations. Here, (a_i, l_i) refers to the interval parametrisation of the closed interval $[a_i, a_i + l_i]$ within which the ith point of the process is observed. In other words, each point of the process X is marked independently by an interval (a, l). We therefore observe the set

$$U = \bigcup_{((z_i,t_i),(a_i,l_i)) \in W} \{(z_i, (a_i, l_i))\}, \tag{4.6}$$

We again make the distinction between atomic and non-atomic points. If a point $w_i \in \mathbf{w}$ is of the form $w_i = ((z_i, t_i), (a_i, 0))$, i.e. $l_i = 0$ for the mark (a_i, l_i), we simply set $t_i = a_i$, and hence $x_i = (z_i, a_i)$ in X. In this way, we split the points so that we obtain

$$\mathbf{x} = \{(z_1, a_1), \ldots, (z_m, a_m), (z_{m+1}, t_{m+1}), \ldots, (z_n, t_n)\}$$

when m atoms are observed. The mechanism by which points of X are marked is through the application of the measure ν, where ν is the mark kernel defined as

$$\begin{aligned} \nu(A|t) = &\left(1 - \int_{-\infty}^{t} \left[1 - G_Y(s, t-s)\right] \, dM(s)\right) \delta(\{(t, 0)\} \cap A) \\ &+ \int_{-\infty}^{t} \int_{t-a}^{\infty} \mathbf{1}\{(a, l) \in A\} \, G_Y(a, dl) \, dM(a), \end{aligned} \tag{4.7}$$

for $t \in W_T$ and a Borel subset $A \subset \mathbb{R} \times \mathbb{R}^+$ [91]. Note that inference is performed only for the subset $\{t_{m+1}, \ldots, t_n\}$ of the conditional distribution on the product space $(W_T)^{n-m}$. The function G_Y refers to the semi-Markov kernel, and M is the renewal function extended to semi-Markov processes. The conditional distribution of the spatial observations together with their intervals, with respect to the point process $X = \mathbf{x}$, is

$$\begin{aligned} &\mathbb{P}(U \in F | X = \mathbf{x}) = \\ &\qquad \int_{(\mathbb{R} \times \mathbb{R}^+)^n} \mathbf{1}(\{(z_1, (a_1, l_1)), \ldots, (z_n, (a_n, l_n))\} \in F) \prod_{i=1}^{n} d\nu((a_i, l_i)|t_i), \end{aligned} \tag{4.8}$$

for a set F in the space $\mathcal{N}_{L \times (\mathbb{R} \times \mathbb{R}^+)}$ [91].

In a similar vein, by [91, Theorem 3.1], we obtain

$$\mathbb{P}(W \in A | U = \mathbf{u}) \propto$$

$$\int_{(L\times W_T)^{n-m}} p(\{(z_1,a_1),...,(z_m,a_m),(z_{m+1},t_{m+1}),...,(z_n,t_n)\})$$
$$1_A(\{[(z_1,a_1),(a_1,0)],\ldots,[(z_m,a_m),(a_m,0)],[(z_{m+1},t_{m+1}),(a_{m+1},l_{m+1})],$$
$$\ldots,[(z_n,t_n),(a_n,l_n)]\})\prod_{i=1}^{n-m} 1_{[a_{m+i},a_{m+i}+l_{m+i}]}(t_i)\,dt_i \tag{4.9}$$

for the conditional, or posterior, distribution of W given an explicit interval configuration $\mathbf{u}$.

Metropolis-Hastings sampling for the full model

For Metropolis-Hastings sampling, we retain the core of the state estimation method proposed in [90], with some changes in notation due to the addition of the spatial axis. The state space is

$$\overline{E}(\mathbf{u}) = \{((z_{m+1},t_{m+1}),\ldots,(z_n,t_n)) \in (L\times W_T)^{n-m} :$$
$$(z_{m+i},t_{m+i}) \in L\times(W_T\cap[a_{m+i},a_{m+i}+l_{m+i}]),$$
$$p(\{(z_1,a_1),\ldots,(z_m,a_m),(z_{m+1},t_{m+1}),\ldots,(z_n,t_n)\})>0, i\in 1,\ldots,n-m\},$$

noting that the position of the points in space does not change throughout. Ignoring impossible configurations, their joint posterior probability density is

$$q(t_1,\ldots,t_n \mid \mathbf{u}) = p(\{(z_1,a_1),\ldots,(z_m,a_m),(z_{m+1},t_{m+1}),\ldots,(z_n,t_n)\}) \tag{4.10}$$

The Metropolis-Hastings algorithm proposes to change the location in time of a given point on the graph at each step of execution. To do this, the Hastings ratio is used. The Hastings ratio expresses the likelihood that a proposed point $\xi \in L\times W_T$ replaces the ith point of a point process $\mathbf{x}$, and can be written as the quotient of the conditional intensities of the process at the two points. Note that p now refers to the density of a network area-interaction process, which has interaction in both space and time. This means that certain times may be favoured due to intersections of interaction areas in the spatial plane.

We obtain, for $i = m+1,\ldots,n$,

$$\begin{aligned} r_i(\mathbf{w},\xi) &= \frac{\lambda(\xi \mid (\mathbf{x}\setminus(z_i,t_i)))}{\lambda((z_i,t_i)\mid(\mathbf{x}\setminus(z_i,t_i)))} \\ &= \frac{\beta(\xi)\gamma^{-\mu[L_{r,h}(\xi)\setminus U_{r,h}(\mathbf{x}\setminus(z_i,t_i))]}}{\beta(z_i,t_i)\gamma^{-\mu[L_{r,h}(z_i,t_i)\setminus U_{r,h}(\mathbf{x}\setminus(z_i,t_i))]}} \\ &= \frac{\beta(\xi)}{\beta(z_i,t_i)}\gamma^{-(\mu[L_{r,h}(\xi)\setminus U_{r,h}(\mathbf{x}\setminus(z_i,t_i))]-\mu[L_{r,h}(z_i,t_i)\setminus U_{r,h}(\mathbf{x}\setminus(z_i,t_i))])}. \end{aligned} \tag{4.11}$$

Algorithm 4.2 from [90] is used to generate points, meaning that proposals within the ith density-admitting interval are selected uniformly in $[a_i, a_i+l_i]$, and subsequently the acceptance probability is

$$\alpha_i(((z_{m+1},t_{m+1}),\ldots,(z_n,t_n)),\xi) =$$

$$\min(1, r_i(\{(z_1, a_1), \ldots, (z_m, a_m), (z_{m+1}, t_{m+1}), \ldots, (z_n, t_n)\}, \xi)) \tag{4.12}$$

for proposal point $\xi \in L \times W_T$. Note that changing one point at a time can also lead to forbidden states. Assume that the underlying network is the triangle network from Example 4.3, and that we have $z_1 = (2, (0, 1.8))$ and $z_2 = (3, (0.1, 1.8))$. Then $\mathbf{u} = \{[z_1, (0.4, 0.5)], [z_2, (0.3, 0.6)]\}$. Take the states $\mathbf{x} = ((z_1, 0.45), (z_2, 0.48))$ and $\mathbf{y} = ((z_1, 0.48), (z_2, 0.45))$. Construct a density function $p(\cdot)$ such that $p(\mathbf{x}) = 0$ if $d_G(z_i, z_j) \leq 0.5$. In this case, states $\mathbf{x}$ and $\mathbf{y}$ are not reachable from each other. These kinds of "deadlocks" can also occur if $p(\mathbf{x}) = 0$ for some time range (see [90] for more details). We also assume that it is always the case that

$$\int_{\overline{E}(\mathbf{u})} p_X(\{(z_1, a_1), \ldots, (z_m, a_m), (z_{m+1}, t_{m+1}), \ldots, (z_n, t_n)\})\, dt_{m+1} \ldots dt_n > 0, \tag{4.13}$$

such that states that do not belong to $\overline{E}(\mathbf{u})$ cannot be generated from the proposal mechanism described. Convergence results from [90] hold in general, since the mechanism of the algorithm itself is unchanged. For general details regarding convergence and setup for Metropolis-Hastings algorithms, see [99, 100, 102] as well as the proofs in [90].

4.4 Parameter estimation and model fitting

Since the data are temporally censored, we make use of the following approach. First, we determine which distribution likely generated the time intervals within which events occurred. This involves parameter estimation for the temporal censoring mechanism. To do this, we use methods first developed in [90], extended to the non-homogeneous case in [91] and applied to crime data in [96]. Then, state estimation is performed to determine the most likely location of points within the provided time intervals, given spatial and temporal covariate information.

4.4.1 Forward model parameter estimation

As in [90], we will use a Bayesian approach to perform parameter estimation and sample from the conditional, or posterior distribution of the full model. Even though an extension to the spatial domain has been performed, the mark kernel itself is unchanged and only depends on time (see Chapter 4.3.3). Therefore, we are able to factorise the area-interaction density $p(\cdot)$ as was possible in the purely temporal semi-Markov case (see [91, Theorem 3.1]), and we obtain separate likelihoods for the network area-interaction density and the censoring mechanism. Within this Bayesian framework, the prior becomes the network area-interaction density, and the censoring mechanism the forward model. We follow the framework outlined in [91] to obtain the vector of mark parameters $\boldsymbol{\xi} = (\xi_S, \xi_R)$ by considering the intervals.

Parameter type	Function	Num	Notation	Model	Short description
Covariate parameters	$\beta(z,s)$	K	$\theta_0,\ldots,\theta_B$	Network area-interaction (prior)	Coefficients for covariate functions. Note that θ_0 is the intercept.
Interaction parameter	$\beta(z,s)$	1	γ, $\theta_{B+1} = \log\gamma$	Network area-interaction (prior)	Describes the interaction type of the points in the process. If $\gamma < 1$ they are regular, $\gamma = 1$ random and $\gamma > 1$ clustered. θ_9 is used in $\boldsymbol{\theta}$.
Interaction radius and distance parameters	$p(\mathbf{x})$	2	r, d	Network area-interaction (prior)	Sets the interaction area. The time radius is denoted by r and the distance radius is denoted by d.
Renewal function parameters	m	J	$\delta_1,\ldots,\delta_J$	Semi-Markov censoring mechanism	Relative weights of renewal function in subsets $A_1,\ldots,A_J$ of observation window.
Non-homogeneous rate parameters	$\lambda(g_Y)$	3	α, b, c	Semi-Markov censoring mechanism	α denotes the scale parameter of interval length and is estimated. b, c are harmonic parameters and are set to default values.
Non-homogeneous shape parameter	g_Y	1	κ	Semi-Markov censoring mechanism	Parameter denoting the shape parameter of interval length. If g_Y exponential, $\kappa = 1$.

Table 4.1 Parameter table. Provides a qualitative description of the many parameters in the full model, where parameters are categorised based on which model they belong to.

4.4.2 Equations used for prior parameter estimation

We now turn to parameter estimation for a network area-interaction process prior. Let $\theta_4 = \log\gamma$. The parameter vector is $\boldsymbol{\theta} = \{\theta_0,\ldots,\theta_B,\theta_{B+1}\}$, where B is the number of covariates in the model. For this, we proceed with a multi-step data-driven pseudolikelihood approach not dissimilar to the one developed and refined in [7, 8, 15, 75] and applied in, among others, [71, 93]. For an overview of parameters, see Table 4.1.

Referring back to equation 5.2, the Papangelou conditional intensity of the network area-interaction prior when adding the point $x = (z,s) \in L \times W_T$ can be written in the parametrised form

$$\lambda_\theta((z,s)\,|\,\mathbf{x}) = \beta(z,s)\gamma^{-A((z,s)\,|\,\mathbf{x})} \tag{4.14}$$

writing $A(z,s\,|\,\mathbf{x}) = \mu(L_{r,h}(z,s) \setminus U_{r,h}(\mathbf{x}))$. We introduce the vector of local sufficient statistics $\mathbf{S}((z,s)\,|\,\mathbf{x})$ evaluated at the point $(z,s) \in L$. In this application, the sufficient statistics are the covariate functions, which impart covariate information, along with the negative area $-A((z,s)\,|\,\mathbf{x})$ corresponding to η. Since the area A is dependent on the current configuration of the pattern, the vector of summary statistics is also dependent on $\mathbf{x}$. Rewriting, we see that we can write

λ_θ in the log-linear form

$$\lambda_\theta((z,s)\,|\,\mathbf{x}) = e^{\boldsymbol{\theta}^T \boldsymbol{S}((z,s)\,|\,\mathbf{x})}. \tag{4.15}$$

The partial derivative of equation 4.15 with respect to $\boldsymbol{\theta}$ is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \lambda_\theta((z,s)\,|\,\mathbf{x}) = \mathbf{S}((z,s)\,|\,\mathbf{x}) \lambda_\theta((z,s)\,|\,\mathbf{x}).$$

Let $n(\mathbf{x})$ be the number of points in a realisation $\mathbf{x}$ of X in the space $L \times W_T$. The unnormalised density under parameter vector $\boldsymbol{\theta}$, which refers to the point process density without the normalisation constant α_p, is

$$q_\theta(\mathbf{x}) = \left(\prod_{(z,s)\in\mathbf{x}} \beta(z,s) \right) \gamma^{-\mu(U_{r,h}(\mathbf{x}))}. \tag{4.16}$$

Let $\mathbf{Z}(\mathbf{x})$ represent the vector of (global) sufficient statistics such that $q_\theta(\mathbf{x}) = e^{\boldsymbol{\theta}^T \mathbf{Z}(\mathbf{x})}$ defines an exponential family model.

4.4.3 Prior parameter estimation with partial observation

The first step in the prior parameter estimation procedure relates to the Georgii-Nguyen-Zessin (GNZ) equation [52, 109], which states that for the point process X on $L \times W_T$, and for all non-negative measurable functions f and λ_θ,

$$\mathbb{E} \sum_{(z,s)\in X\cap(L\times W_T)} f_\theta((z,s)\,|\,X \setminus \{(z,s)\}) = \mathbb{E} \int_L \int_{W_T} f_\theta((z,s)\,|\,X) \lambda_\theta((z,s)\,|\,X)\, d\ell(s)\, d\lambda_G(z), \tag{4.17}$$

under the convention that the left-hand side is finite if and only if the right-hand side is. Note that ℓ refers to Lebesgue measure. For (z_i, s_i) not contained in $\mathbf{x}$,

$$\lambda^{(k)}((z_1,s_1),\ldots(z_k,s_k)|\mathbf{x}) = \frac{p(\mathbf{x} \cup \{((z_1,s_1),\ldots,(z_k,s_k)\})}{p(\mathbf{x})}$$

is the Papangelou conditional intensity.

In the general case, a specific function f_θ, is chosen, both sides of the GNZ equation are estimated based on an observation $\mathbf{x}$ of X, and one solves for the parameter vector $\boldsymbol{\theta}$ by way of the score equation. This is known as the Takacs-Fiksel method [138, 139]. We choose $k = 1$ and

$$f_\theta((z,s)\,|\,X) = \frac{\nabla \lambda_\theta((z,s)\,|\,X)}{\lambda_\theta((z,s)\,|\,X)},$$

the pseudolikelihood, as introduced in [18]. In this particular application, estimating both sides of the GNZ equation is non-trivial since we do not observe X perfectly. For $(z, s) \in X$, while z is always fully observed, s is only partially observed within a temporal interval and can therefore differ per observation. This presents an issue, as we cannot determine values of the Papangelou conditional intensity over time intervals. Therefore, we require sample patterns over which the values of $\lambda_\theta(\cdot \mid \mathbf{x})$ and f_θ can be calculated. We propose the use of an approach based on posterior sampling. Recall that the data pattern is

$$U = \bigcup_{((z,s),(a,l))\in W} \{(z, (a, l))\}.$$

We consider the expected value of $f_\theta(\cdot, t)$ over an interval $[a, a + l]$. The GNZ equation becomes, for each respective $s \in [a, a + l] \subseteq W_T$,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[\sum_{(z,s)\in X} f_\theta((z, s) \mid X \setminus \{(z, s)\}) \,\middle|\, U\right]\right] \\
&= \mathbb{E}\left[\int_L \int_{W_T} \mathbb{E}\left[f_\theta((z, s) \mid X)\lambda_\theta((z, s) \mid X) \mid U\right] \, d\ell(t)\, d\lambda_G(z)\right]. \qquad (4.18)
\end{aligned}$$

It follows directly due to the law of total expectation that

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{(z,s)\in X} f_\theta((z, s) \mid X \setminus \{(z, s)\}) \,\middle|\, U\right]\right] = \mathbb{E}\left[\sum_{(z,s)\in X} f_\theta((z, s) \mid X \setminus \{(z, s)\})\right]$$

and

$$\begin{aligned}
&\mathbb{E}\left[\int_L \int_{W_T} \mathbb{E}\left[f_\theta((z, s) \mid X)\lambda_\theta((z, s) \mid X) \mid U\right] \, d\ell(s)s\, d\lambda_G(z)\right] \\
&\quad = \mathbb{E}\left[f_\theta((z, s) \mid X)\lambda_\theta((z, s) \mid X)\right].
\end{aligned}$$

Since the two right-hand sides are unbiased, Equation 4.18 without the expectations is also an unbiased estimating equation.

We wish to sample from $\mathcal{L}_\theta(X|U = \mathbf{u})$, the posterior likelihood with respect to the parameter vector $\boldsymbol{\theta}$, conditioned on a particular realisation of the data pattern U. We cannot directly sample from this likelihood since $\boldsymbol{\theta}$ is unknown. To overcome this, we use a Geyer-style importance sampling approach [54, 55] where we use the Metropolis-Hastings algorithm outlined in Chapter 4.3.3 to generate some number of samples $Y_1, \ldots, Y_K$ from $\mathcal{L}_\psi(X|U = \mathbf{u})$ with known parameter vector $\boldsymbol{\psi}$. For this approach, we require that $n(Y_1) = \cdots = n(Y_K) = n(\mathbf{u})$. A realisation X_k would take the form

$$X_k = \{(z_1, a_1), \ldots, (z_m, a_m), (z_{m+1}, Y_{k,1}), \ldots, (z_n, Y_{k,n-m})\}$$

and within this context the sufficient statistics can be rewritten as

$$\begin{aligned} S_\theta(Y_k, \mathbf{u}) &= \sum_{j=1}^{m} f_\theta((z_j, a_j), \{(z_i, a_i), j \neq i = 1, \ldots, m\} \\ &\quad \cup \{(z_{m+i}, Y_{k,i}), i = 1, \ldots, n-m\}) \\ &+ \sum_{j=1}^{n-m} f_\theta((z_{m+j}, Y_{k,j}), \{(z_i, a_i, i = 1, \ldots, m\} \\ &\quad \cup \{(z_{m+i}, Y_{k,i}), j \neq i = 1, \ldots, n-m\}). \end{aligned}$$

This importance sampling approach allows us to estimate the parameter vector $\boldsymbol{\theta}$ using samples from a related distribution. For a function g of a pattern X, we obtain

$$\mathbb{E}_{K,\theta} g(X \,|\, \mathbf{u}) = \frac{1}{K} \sum_{k=1}^{K} w_\theta(Y_k) g(Y_k \,|\, \mathbf{u})$$

so long as the weights, which are given by

$$w_\theta(X) = \frac{q_\theta(X \,|\, \mathbf{u}) / q_\psi(X \,|\, \mathbf{u})}{\frac{1}{K} \sum_{k=1}^{K} (q_\theta(Y_k \,|\, \mathbf{u}) / q_\psi(Y_k \,|\, \mathbf{u}))},$$

take the form of a ratio of unnormalised densities [53, 55]. In this case $q_\psi(\cdot \,|\, \mathbf{u})$ refers to the known unnormalised density with respect to the parameter vector $\boldsymbol{\psi}$, given the set of non-degenerate intervals in U. In this application, we set $g(Y_k \,|\, \mathbf{u}) = S_\theta(Y_k, \mathbf{u})$. We now proceed by estimating the left- and right-hand sides of equation 4.18. Since we are estimating the outside expectations, estimating both sides of the generic GNZ equation (4.17) is equivalent to estimating both sides of equation 4.18. For the LHS of equation 4.17, we obtain

$$\begin{aligned} &\mathbb{E}_{K,\theta} \sum_{(z,s) \in X \cap (L \times W_T)} f_\theta((z,s) \,|\, X \setminus \{(z,s)\}) \\ &= \mathbb{E}_{K,\theta} \left[\sum_{(z,s) \in X \cap (L \times W_T)} \frac{\nabla \lambda_\theta((z,s) \,|\, X \setminus \{(z,s)\})}{\lambda_\theta((z,s) \,|\, X \setminus \{(z,s)\})} \right] \\ &= \frac{1}{K} \sum_{k=1}^{K} w_\theta(Y_k) \sum_{(z,s) \in Y_k} \frac{\nabla \lambda_\theta((z,s) \,|\, Y_k \setminus \{(z,s)\})}{\lambda_\theta((z,s) \,|\, Y_k \setminus \{(z,s)\})} \\ &= \frac{1}{K} \sum_{k=1}^{K} w_\theta(Y_k) \sum_{(z,s) \in Y_k} \frac{S_\theta(Y_k, \mathbf{u}) \lambda_\theta((z,s) \,|\, Y_k \setminus \{(z,s)\})}{\lambda_\theta((z,s) \,|\, Y_k \setminus \{(z,s)\})} \\ &= \frac{1}{K} \sum_{k=1}^{K} w_\theta(Y_k) \sum_{(z,s) \in Y_k} S_\theta(Y_k, \mathbf{u}) \end{aligned}$$

$$= \frac{\sum_{k=1}^{K}(q_\theta(Y_k)/q_\psi(Y_k))\sum_{(z,s)\in Y_k} S_\theta(Y_k,\mathbf{u})}{\sum_{k=1}^{K}(q_\theta(Y_k)/q_\psi(Y_k))}.$$

The RHS can be approximated by way of a dummy process D [8]. In this application, we assume that D is a non-homogeneous Poisson point process with density $\rho((x,t)\,|\,\mathbf{u}) = r_D\beta(x,t)$ where r_D is application-dependent and is chosen in combination with $n(\mathbf{u})$ such that there are roughly 4 times as many points in the dummy realisation as in the data pattern, see [8] for more details.

This approach allows for more points to be generated in areas where the integral is more likely to be positive, and conversely less points to be generated in areas where the integral is 0, making estimation more efficient. Denote a realisation of D by $\mathbf{d}$ and write

$$S_{D,\theta}(d,Y_k,\mathbf{u}) = f_\theta(d,\{(z_i,a_i),i=1,\ldots,m\}\cup\{(z_{m+i},Y_{k,i}),i=1,\ldots,n-m\}).$$

Following approaches in [8, 15, 70], we approximate the RHS of equation 4.17 by the Monte Carlo average

$$\begin{aligned}
&\mathbb{E}_{K,\theta}\left[\int_L\int_{W_T} f_\theta((z,s)\,|\,X)\lambda_\theta((z,s)\,|\,X)\,d\ell(s)\,d\lambda_G(z)\right]\\
&\quad= \mathbb{E}_{K,\theta}\left[\int_L\int_{W_T}\frac{\nabla\lambda_\theta((z,s)\,|\,X)}{\lambda_\theta((z,s)\,|\,X)}\lambda_\theta((z,s)\,|\,X)\,d\ell(s)\,d\lambda_G(z)\right]\\
&\quad= \mathbb{E}_{K,\theta}\left[\int_L\int_{W_T}\frac{\nabla\lambda_\theta((z,s)\,|\,X)\lambda_\theta((z,s)\,|\,X)\rho(z,s)}{\lambda_\theta((z,s)\,|\,X)\rho(z,s)}\,d\ell(s)\,d\lambda_G(z)\right]\\
&\quad= \mathbb{E}_{K,\theta}\left[\mathbb{E}\left[\sum_{(x,t)\in D}\frac{\nabla\lambda_\theta((x,t)\,|\,X)}{\rho(x,t)}\,\middle|\,X\right]\right]\\
&\quad= \frac{1}{K}\sum_{k=1}^{K} w_\theta(Y_k)\sum_{(x,t)\in D}\frac{\nabla\lambda_\theta((x,t)\,|\,Y_k)}{\rho(x,t)}\\
&\quad= \frac{\sum_{k=1}^{K}(q_\theta(Y_k)/q_\psi(Y_k))\sum_{(x,t)\in D}\left(\frac{S_{D,\theta}(d,Y_k,\mathbf{u})\lambda_\theta((x,t)\,|\,Y_k)}{\rho(x,t)}\right)}{\sum_{k=1}^{K}(q_\theta(Y_k)/q_\psi(Y_k))}.
\end{aligned}$$

Therefore, we obtain the estimating equation

$$\begin{aligned}
&s(\mathbf{y}_1,\ldots,\mathbf{y}_k,\mathbf{d};\boldsymbol{\theta}) =\\
&\quad\frac{\sum_{k=1}^{K}(q_\theta(\mathbf{y}_k)/q_\psi(\mathbf{y}_k))\left(\sum_{(z,s)\in\mathbf{y}_k} S_\theta(Y_k,\mathbf{u})\right)}{\sum_{k=1}^{K}(q_\theta(\mathbf{y}_k)/q_\psi(\mathbf{y}_k))}\\
&\quad-\frac{\sum_{k=1}^{K}(q_\theta(\mathbf{y}_k)/q_\psi(\mathbf{y}_k))\sum_{(x,t)\in\mathbf{d}}\frac{S_{D,\theta}(d,Y_k,\mathbf{u})\lambda_\theta((x,t)\,|\,\mathbf{y}_k)}{\rho(x,t)}}{\sum_{k=1}^{K}(q_\theta(\mathbf{y}_k)/q_\psi(\mathbf{y}_k))} = \mathbf{0}
\end{aligned}\qquad(4.19)$$

when written in vector form. To solve this, we use a Newton-Raphson solver from the R package `rootSolve` [134]. We then obtain the vector of parameter estimates $\hat{\boldsymbol{\theta}}$ that maximises the posterior likelihood $\mathcal{L}_\theta(X|U)$, henceforth known as the maximiser.

4.5 Uncertainty of estimates

We now wish to determine the uncertainty of the estimates obtained by the procedure in Chapter 4.4.3 when finding the root of the estimating equation or composite score $s(\mathbf{y}_1, \ldots, \mathbf{y}_K, \mathbf{d}; \boldsymbol{\theta})$ and obtaining $\hat{\boldsymbol{\theta}}$. The estimating equation is unbiased by construction since it is a direct result of the GNZ equation. To estimate standard errors, one possible approach would be to use parametric bootstrapping - i.e. generate a number of simulated samples based on $\hat{\boldsymbol{\theta}}$. However, this is not feasible in this case due to excessively long computation times. Another alternative would be to look at asymptotic behaviour. Since stationarity assumptions break down over networks (as discussed in [14]), it seems futile to define a spatial asymptotic regime. Stable estimates for $\boldsymbol{\theta}$ over time may be possible since an asymptotic regime may exist, however this is beyond the scope of this thesis. We therefore use the asymptotic approach suggested by Godambe & Heyde [56] and applied to spatial Gibbs processes in [16], where a similar pseudolikelihood form for the estimating equation has been used.

In this application, we use the Godambe information matrix [56] with the purpose of verifying the stability of the estimating equation. The Godambe information matrix, which can be seen as a generalisation of the Fisher information matrix for unbiased estimating equations, is given by

$$G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}H(\boldsymbol{\theta}).$$

where $H(\boldsymbol{\theta})$ is the sensitivity matrix and $J(\boldsymbol{\theta})$ is the composite score or covariance matrix.

For the remainder of this section, let

$$X_S = \sum_{(z,s)\in X\cap(L\times W_T)} \mathbf{S}((z,s) \,|\, X \setminus \{(z,s)\})$$

and

$$X_I = \int_L \int_{W_T} \mathbf{S}((z,s) \,|\, X)\lambda_\theta((z,s) \,|\, X)\, ds\, dz,$$

where $\mathbf{S}$ refers to the local sufficient statistics as in Equation 4.15. While a more accurate estimator may be obtained by working with the Monte Carlo estimating equation, expressions for the covariance matrix are intractable and

may not even exist in certain cases. Therefore, we will work with the non-Monte Carlo conditional estimating equation, obtained via equation 4.18. This is

$$s((X \mid U) \mid \boldsymbol{\theta}) = \mathbb{E}\left[X_S \mid U\right] - \mathbb{E}\left[X_I \mid U\right].$$

Note that due to the conditioning on the data pattern U, we have more information on X than in the standard case, where the GNZ equation takes the form outlined in equation 4.17. We therefore do not follow the standard approaches developed in [12, 16, 27] and derive a novel covariance matrix estimator for a partially observed spatial Markov point process. Expressions for the sensitivity and covariance matrices $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$, as well as plug-in estimators $\hat{H}(\hat{\boldsymbol{\theta}})$ and $\hat{J}(\hat{\boldsymbol{\theta}})$ are hence provided in the following subchapters.

4.5.1 Sensitivity matrix

Using the law of total expectation and linearity properties, the sensitivity matrix is

$$\begin{aligned} H(\boldsymbol{\theta}) &= -\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} s((X \mid U) \mid \boldsymbol{\theta})\right] = \mathbb{E}\left[-\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} X_S \,\middle|\, U\right] + \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} X_I \,\middle|\, U\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} X_I \,\middle|\, U\right]\right] = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} X_I\right] \\ &= \mathbb{E}\left[\int_L \int_{W_T} \mathbf{S}((z,s) \mid X)\mathbf{S}((z,s) \mid X)^T \lambda_\theta((z,s) \mid X)\, ds\, dz\right], \end{aligned}$$

which is identical to the unconditional case (i.e. when plugging in $s(X \mid \boldsymbol{\theta})$, see [12] for more details). A plug-in estimator for $H(\boldsymbol{\theta})$ can be obtained by generating N samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ using the coupling from the past algorithm in Chapter 4.3.2 under the estimated parameter vector $\hat{\boldsymbol{\theta}}$ and evaluating the LHS of the unconditional GNZ equation. This is

$$\hat{H}(\hat{\boldsymbol{\theta}}) = \frac{1}{J} \sum_{j=1}^{J} \sum_{(z,s) \in \mathbf{x}_j} \mathbf{S}((z,s) \mid \mathbf{x}_j)\mathbf{S}((z,s) \mid \mathbf{x}_j)^T. \tag{4.20}$$

4.5.2 Variance in the conditional case

The composite score matrix is defined as the (co)variance of the score function with respect to the parameter vector $\boldsymbol{\theta}$. The variance becomes

$$\begin{aligned} &\operatorname{var}(s((X \mid U) \mid \boldsymbol{\theta})) = \operatorname{var}(\mathbb{E}\left[X_S \mid U\right] - \mathbb{E}\left[X_I \mid U\right]) \\ &= \operatorname{var}(\mathbb{E}\left[X_S \mid U\right]) + \operatorname{var}(\mathbb{E}\left[X_I \mid U\right]) - 2\operatorname{cov}(\mathbb{E}\left[X_S \mid U\right], \mathbb{E}\left[X_I \mid U\right]) \end{aligned}$$

Let V_1 correspond to the first term, V_2 to the second term and V_3 to the third term. Using the law of total variance we obtain

$$V_1 = \operatorname{var}(\mathbb{E}\left[X_S \mid U\right]) = \operatorname{var}(X_S) - \mathbb{E}\left[\operatorname{var}(X_S \mid U)\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[X_S^2\right] - \left(\mathbb{E}\left[X_S\right]\right)^2 - \mathbb{E}\left[\mathbb{E}\left[X_S^2 \,|\, U\right]\right] + \mathbb{E}\left[\left(\mathbb{E}\left[X_S \,|\, U\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[X_S \,|\, U\right]\right)^2\right] - \left(\mathbb{E}\left[X_S\right]\right)^2,
\end{aligned}
$$

similarly

$$
\begin{aligned}
V_2 = \operatorname{var}(\mathbb{E}\left[X_I \,|\, U\right]) &= \operatorname{var}(X_I) - \mathbb{E}\left[\operatorname{var}(X_I \,|\, U)\right] \\
&= \mathbb{E}\left[X_I^2\right] - \left(\mathbb{E}\left[X_I\right]\right)^2 - \mathbb{E}\left[\mathbb{E}\left[X_I^2 \,|\, U\right]\right] + \mathbb{E}\left[\left(\mathbb{E}\left[X_I \,|\, U\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[X_I \,|\, U\right]\right)^2\right] - \left(\mathbb{E}\left[X_I\right]\right)^2.
\end{aligned}
$$

Simplifying the covariance expression,

$$
\begin{aligned}
V_3 &= -2\operatorname{cov}(\mathbb{E}\left[X_S \,|\, U\right], \mathbb{E}\left[X_I \,|\, U\right]) \\
&= 2\mathbb{E}\left[\mathbb{E}\left[X_S \,|\, U\right]\right] \mathbb{E}\left[\mathbb{E}\left[X_I \,|\, U\right]\right] - 2\mathbb{E}\left[\mathbb{E}\left[X_S \,|\, U\right] \mathbb{E}\left[X_I \,|\, U\right]\right] \\
&= 2\mathbb{E}[X_S]\mathbb{E}[X_I] - 2\mathbb{E}\left[\mathbb{E}\left[X_S \,|\, U\right] \mathbb{E}\left[X_I \,|\, U\right]\right].
\end{aligned}
$$

Note that because of equation 4.17, we have $\mathbb{E}[X_S] = \mathbb{E}[X_I]$. Using this, we obtain

$$
\begin{aligned}
\operatorname{var}(s((X \,|\, U) \,|\, \boldsymbol{\theta})) &= V_1 + V_2 + V_3 \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[X_S \,|\, U\right]\right)^2\right] - \left(\mathbb{E}\left[X_I\right]\right)^2 + \mathbb{E}\left[\left(\mathbb{E}\left[X_I \,|\, U\right]\right)^2\right] - \left(\mathbb{E}\left[X_I\right]\right)^2 \\
&\quad + 2\left(\mathbb{E}[X_I]\right)^2 - 2\mathbb{E}\left[\mathbb{E}\left[X_S \,|\, U\right] \mathbb{E}\left[X_I \,|\, U\right]\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[X_S \,|\, U\right]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}\left[X_I \,|\, U\right]\right)^2\right] - 2\mathbb{E}\left[\mathbb{E}\left[X_S \,|\, U\right] \mathbb{E}\left[X_I \,|\, U\right]\right].
\end{aligned}
\tag{4.21}
$$

The conditional variance can only be evaluated immediately in the trivial case of full information ($U = X$), or in the case that U provides no information on X. In the former case, the variance is simply 0, whereas in the latter case, we obtain the standard GNZ estimating equation variance as derived in the previous subchapter. Note that this means that the conditional variance is guaranteed to be smaller than in the standard case.

When we solve the score of the estimating equation for 0, we obtain an estimate of the parameter vector $\hat{\boldsymbol{\theta}}$ given a data observation $U = \mathbf{u}$. The conditional variance tells us how sensitive $\hat{\boldsymbol{\theta}}$ is when varying U and hence also $X \,|\, U$. To proceed, we propose a two-step algorithm based on simulation and sampling methods introduced in earlier sections.

Algorithm 4.10. *Let $\hat{\boldsymbol{\theta}}$ be the parameter vector estimated by solving equation 4.18.*

- *Generate K realisations of U, obtained by sampling from the full model W as laid out in Chapter 4.3.3 using coupling from the past (Chapter 4.9).*

Denote these realisations by $\mathbf{u}_1, \ldots, \mathbf{u}_K$. *We require one dummy process realisation* $\mathbf{d}$ *of* D *to estimate the integrals* $X_I \,|\, U = \mathbf{u}_1, \ldots, X_I \,|\, U = \mathbf{u}_K$.

- *For each* $k = 1, \ldots, K$, *estimate the inner expectations* $\mathbb{E}[X_S \,|U = \mathbf{u}_k]$ *and* $\mathbb{E}[X_I \,|U = \mathbf{u}_k]$ *by generating* N *samples* $\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kJ}$ *from the corresponding posterior likelihoods* $\mathcal{L}_{\hat{\theta}}(X \,|\, U = \mathbf{u}_k)$ *under* $\hat{\boldsymbol{\theta}}$. *To do this, we use the Metropolis-Hastings method from Chapter 4.3.3.*

- *We estimate* $\mathbb{E}[X_S \,|U = \mathbf{u}_k]$ *via the Monte Carlo plug-in estimator*

$$\hat{E}_{S,k} = \frac{1}{J} \sum_{j=1}^{J} \sum_{(z,s) \in \mathbf{x}_{kj}} S_\theta(\mathbf{x}_{kj}, \mathbf{u}_k)$$

and $\mathbb{E}[X_I \,|U = \mathbf{u}_k]$ *via*

$$\hat{E}_{I,k} = \frac{1}{J} \sum_{j=1}^{J} \sum_{(x,t) \in \mathbf{d}} \frac{S_{D,\theta}(d, \mathbf{x}_{kj}, \mathbf{u}_k) \lambda_{\hat{\theta}}((x,t) \,|\, \mathbf{x}_{kj})}{\rho(x,t)}$$

for all $k = 1, \ldots, K$.

- *An estimator* $\hat{V}$ *for the conditional variance, in analogy to equation 4.21, can be obtained via*

$$\hat{V} = \frac{1}{K} \sum_{k=1}^{K} \left((\hat{E}_{S,k})(\hat{E}_{S,k})^T + (\hat{E}_{I,k})(\hat{E}_{I,k})^T - 2(\hat{E}_{S,k})(\hat{E}_{I,k})^T \right).$$

The estimator $\hat{G}$ for the Godambe information matrix G then becomes

$$\hat{G} = \hat{H} \hat{V}^{-1} \hat{H}.$$

In the following subchapter, for completeness we outline the standard method of calculating the covariance matrix, which is used in the case of perfectly observed data.

4.5.3 Variance in the unconditional case

To obtain the composite score estimator in the unconditional case, we require some definitions. The second-order Papangelou conditional intensity is defined as

$$\lambda_\theta^{[2]}((z,s),(y,u) \,|\, X) = \lambda_\theta((z,s) \,|\, X \setminus \{(y,u)\}) \lambda_\theta((y,u) \,|\, X \cup \{(z,s)\}).$$

The variance becomes

$$\operatorname{var}[s(X \,|\, \boldsymbol{\theta})] = \operatorname{var}\Big(\sum_{(z,s) \in X \cap (L \times W_T)} \mathbf{S}((z,s) \,|\, X \setminus \{(z,s)\})$$

$$
\begin{aligned}
&- \int_L \int_{W_T} \mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X)\, ds\, dz) \\
&= \operatorname{var}\left[\sum_{(z,s) \in X \cap (L \times W_T)} \mathbf{S}((z,s) \mid X \setminus \{(z,s)\}) \right] \\
&+ \operatorname{var}\left[\int_L \int_{W_T} \mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X)\, ds\, dz \right] \\
&- 2\operatorname{cov}(\sum_{(z,s) \in X \cap (L \times W_T)} \mathbf{S}((z,s) \mid X \setminus \{(z,s)\}, \\
&\int_L \int_{W_T} \mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X)\, ds\, dz)
\end{aligned}
$$

Let T_1 correspond to the first term, T_2 to the second term and T_3 to the third term. We obtain

$$
\begin{aligned}
T_1 &= \operatorname{var}\left[\sum_{(z,s) \in X \cap (L \times W_T)} \mathbf{S}((z,s) \mid X \setminus \{(z,s)\}) \right] \\
&= \int_L \int_{W_T} \mathbb{E}\left[\mathbf{S}((z,s) \mid X) \mathbf{S}((z,s) \mid X)^T \lambda_\theta((z,s) \mid X) \right] ds\, dz \\
&+ \int_L \int_{W_T} \int_L \int_{W_T} \mathbb{E}[\mathbf{S}((z,s) \mid X \cup \{(y,u)\}) \mathbf{S}((y,u) \mid X \cup \{(z,s)\})^T \\
&\qquad \lambda_\theta^{[2]}((z,s),(y,u) \mid X)] ds\, dz\, du\, dy \\
&- \int_L \int_{W_T} \int_L \int_{W_T} \mathbb{E}\left[\mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X) \right] \\
&\qquad \mathbb{E}\left[\mathbf{S}((y,u) \mid X) \lambda_\theta((y,u) \mid X) \right] ds\, dz\, du\, dy,
\end{aligned}
$$

the second term becomes

$$
\begin{aligned}
T_2 &= \operatorname{var}\left[\int_L \int_{W_T} \mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X)\, ds\, dz \right] \\
&= \int_L \int_{W_T} \int_L \int_{W_T} \operatorname{cov}(\mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X), \\
&\qquad \mathbf{S}((y,u) \mid X) \lambda_\theta((y,u) \mid X))\, ds\, dz\, du\, dy \\
&= \int_L \int_{W_T} \int_L \int_{W_T} \mathbb{E}[\mathbf{S}((z,s) \mid X) \mathbf{S}((y,u) \mid X)^T \\
&\qquad \lambda_\theta((z,s) \mid X) \lambda_\theta((y,u) \mid X)] ds\, dz\, du\, dy \\
&- \int_L \int_{W_T} \int_L \int_{W_T} \mathbb{E}\left[\mathbf{S}((z,s) \mid X) \lambda_\theta((z,s) \mid X) \right] \\
&\qquad \mathbb{E}\left[\mathbf{S}((y,u) \mid X) \lambda_\theta((y,u) \mid X) \right] ds\, dz\, du\, dy
\end{aligned}
$$

and

$$\begin{aligned}
T_3 = -2\mathrm{cov}(& \sum_{(z,s)\in X\cap(L\times W_T)} \mathbf{S}((z,s)\,|\,X\setminus\{(z,s)\},\\
& \int_L\int_{W_T} \mathbf{S}((z,s)\,|\,X)\lambda_\theta((z,s)\,|\,X)\,ds\,dz)\\
= -2\int_L\int_{W_T}\int_L\int_{W_T} & \mathbb{E}[\mathbf{S}((z,s)\,|\,X)\mathbf{S}((y,u)\,|\,X\cup\{(z,s)\})^T\\
& \lambda_\theta((z,s)\,|\,X)\lambda_\theta((y,u)\,|\,X)]ds\,dz\,du\,dy\\
+2\int_L\int_{W_T}\int_L\int_{W_T} & \mathbb{E}\left[\mathbf{S}((z,s)\,|\,X)\lambda_\theta((z,s)\,|\,X)\right]\\
& \mathbb{E}\left[\mathbf{S}((y,u)\,|\,X)\lambda_\theta((y,u)\,|\,X)\right]ds\,dz\,du\,dy.
\end{aligned}$$

After some rearranging, noticing that the last terms of T_1, T_2 and T_3 all cancel out, we obtain

$$\begin{aligned}
&\mathrm{var}[s(X\,|\,\boldsymbol{\theta})] = T_1 + T_2 + T_3\\
&= \int_L\int_{W_T} \mathbb{E}\left[\mathbf{S}((z,s)\,|\,X)\mathbf{S}((z,s)\,|\,X)^T\lambda_\theta((z,s)\,|\,X)\right]ds\,dz\\
&+ \int_L\int_{W_T}\int_L\int_{W_T} \mathbb{E}[\mathbf{S}((z,s)\,|\,X\cup\{(y,u)\})\mathbf{S}((y,u)\,|\,X\cup\{(z,s)\})^T\\
&\qquad\qquad \lambda_\theta^{[2]}((z,s),(y,u)\,|\,X)]ds\,dz\,du\,dy\\
&+ \int_L\int_{W_T}\int_L\int_{W_T} \mathbb{E}[\mathbf{S}((z,s)\,|\,X)\mathbf{S}((y,u)\,|\,X)^T\\
&\qquad\qquad \lambda_\theta((z,s)\,|\,X)\lambda_\theta((y,u)\,|\,X)]ds\,dz\,du\,dy\\
&- 2\int_L\int_{W_T}\int_L\int_{W_T} \mathbb{E}[\mathbf{S}((z,s)\,|\,X)\mathbf{S}((y,u)\,|\,X\cup\{(z,s)\})^T\\
&\qquad\qquad \lambda_\theta((z,s)\,|\,X)\lambda_\theta((y,u)\,|\,X)]ds\,dz\,du\,dy
\end{aligned}$$

Define the set difference operator such that

$$\Delta_{y,u}\mathbf{S}((z,s)\,|\,X) = \mathbf{S}((z,s)\,|\,X\cup\{(y,u)\}) - \mathbf{S}((z,s)\,|\,X).$$

After switching expectation with integration, following similar derivations in [12, 16, 27] and copious algebra, we find that

$$\begin{aligned}
\mathrm{var}[s(X\,|\,\boldsymbol{\theta})] &= \mathrm{var}\left[\int_L\int_{W_T} \mathbf{S}((z,s)\,|\,X)\lambda_\theta((z,s)\,|\,X)\,ds\,dz\right]\\
&= \int_L\int_{W_T}\int_L\int_{W_T} \mathrm{cov}(\mathbf{S}((z,s)\,|\,X)\lambda_\theta((z,s)\,|\,X),\\
&\qquad \mathbf{S}((y,u)\,|\,X)\lambda_\theta((y,u)\,|\,X))\,ds\,dz\,du\,dy
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{K,\theta}\left[\int_L \int_{W_T} \mathbf{S}((z,s)\,|\,X)\mathbf{S}((z,s)\,|\,X)^T \lambda_\theta((z,s)\,|\,X)\,ds\,dz\right] \\
&+ \mathbb{E}_{K,\theta}\int_L \int_{W_T} \int_L \int_{W_T} \mathbf{S}((z,s)\,|\,X)\mathbf{S}((y,u)\,|\,X)^T \\
&\qquad \Big(\lambda_\theta((z,s)\,|\,X)\lambda_\theta((y,u)\,|\,X) - \lambda_\theta^{[2]}((z,s),(y,u)\,|\,X)\Big) \\
&\qquad\qquad ds\,dz\,du\,dy \\
&+ \mathbb{E}_{K,\theta}\int_L \int_{W_T} \int_L \int_{W_T} \Delta_{y,u}\mathbf{S}((z,s)\,|\,X)\Delta_{z,s}\mathbf{S}((y,u)\,|\,X)^T \\
&\qquad\qquad \lambda_\theta^{[2]}((z,s),(y,u)\,|\,X)\,ds\,dz\,du\,dy.
\end{aligned}
$$

Denote the three terms of $\mathrm{var}[s(X\,|\,\boldsymbol{\theta})]$ as A_1, A_2 and A_3 respectively. Note that $A_1 = H(\boldsymbol{\theta})$. To estimate these quantities, we first use coupling from the past to simulate a number of realisations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ under the reference parameter vector $\boldsymbol{\psi}$. We then re-weight to obtain the estimators for $\boldsymbol{\theta}$. A Monte Carlo estimator $\hat{A}_1(\boldsymbol{\theta})$ for $H(\boldsymbol{\theta})$ and A_1 is

$$
\hat{A}_1(\boldsymbol{\theta}) = \frac{\sum_{i=1}^N q_\theta(\mathbf{x}_i)/q_\psi(\mathbf{x}_i) \sum_{(x,t)\in\mathbf{x}_i} \mathbf{S}((x,t)\,|\,\mathbf{x}_i)\mathbf{S}((x,t)\,|\,\mathbf{x}_i)^T}{\sum_{i=1}^N q_\theta(\mathbf{x}_i)/q_\psi(\mathbf{x}_i)}. \tag{4.22}
$$

For A_2 and A_3, after some algebra and using the GNZ equation, we obtain

$$
\begin{aligned}
A_2 = \sum_{(x,t)\in X} \sum_{(y,u)\neq(x,t)\in X} &\mathbf{S}((x,t)\,|\,X\setminus\{(y,u)\})\,\mathbf{S}((y,u)\,|\,X\setminus\{(x,t)\})^T \\
&\left(\frac{\lambda_\theta((y,u)\,|\,X\setminus\{(x,t)\})}{\lambda_\theta((y,u)\,|\,X)} - 1\right)
\end{aligned}
$$

and

$$
\begin{aligned}
A_3 = \sum_{(x,t)\in X} \sum_{(y,u)\neq(x,t)\in X} &\Delta_{(y,u)}\mathbf{S}((x,t)\,|\,X\setminus\{(y,u)\}) \\
&\Delta_{(x,t)}\mathbf{S}((y,u)\,|\,X\setminus\{(x,t)\})^T
\end{aligned}
$$

and the estimators $\hat{A}_2(\boldsymbol{\theta})$ and $\hat{A}_3(\boldsymbol{\theta})$ become the Monte Carlo averages

$$
\hat{A}_2(\boldsymbol{\theta}) = \frac{\sum_{i=1}^N q_\theta(\mathbf{x}_i)/q_\psi(\mathbf{x}_i)A_2}{\sum_{i=1}^N q_\theta(\mathbf{x}_i)/q_\psi(\mathbf{x}_i)}
$$

and

$$
\hat{A}_3(\boldsymbol{\theta}) = \frac{\sum_{i=1}^N q_\theta(\mathbf{x}_i)/q_\psi(\mathbf{x}_i)A_3}{\sum_{i=1}^N q_\theta(\mathbf{x}_i)/q_\psi(\mathbf{x}_i)}.
$$

These estimator forms are consistent with derivations of similar estimators in [16, 27]. We finally obtain

$$
\hat{V}(\boldsymbol{\theta}) = \hat{A}_1 + \hat{A}_2 + \hat{A}_3.
$$

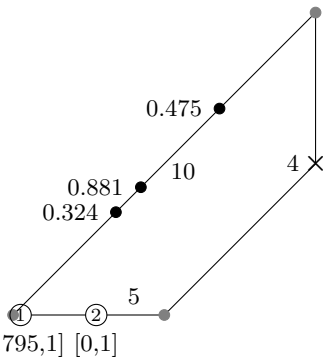


Figure 4.5 The graph on which simulations take place. The points on the graph are part of realisation $\mathbf{w}$. The vertices are at $(0,0),(1,0)$ and $(2,2)$. There are two non-atomic points, which are denoted by a circle with a number inside and are the first two points of the pattern. The time intervals become their intersection with $\mathcal{X}$. The three remaining atomic points are denoted by filled-in circles with a time of occurrence to the left.

The estimated Godambe information matrix $\hat{G}(\hat{\boldsymbol{\theta}})$ [56] is then obtained by calculating

$$\hat{G} = \hat{H}\hat{V}^{-1}\hat{H}. \tag{4.23}$$

4.6 Simulated example

To ensure that the generation and estimation process is performed correctly, we test it on a simulated example. We take $W_T = [0,1]$, and assume the graph is of the form $G = (\mathcal{V}, \mathcal{E}, \mathbf{\Phi})$ as in Chapter 4.1.1. We have

$$V = \{(0,(0,0)),(0,(1,0)),(0,(2,2))\}$$

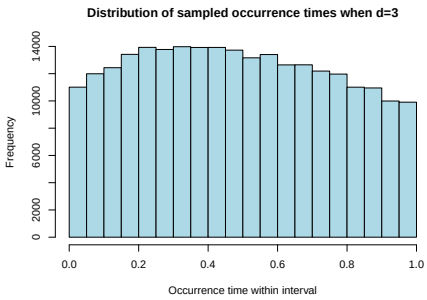
with edges

$$E = \{((0,0),(1,0)),((1,0),(2,2)),((0,0),(2,2))\}$$

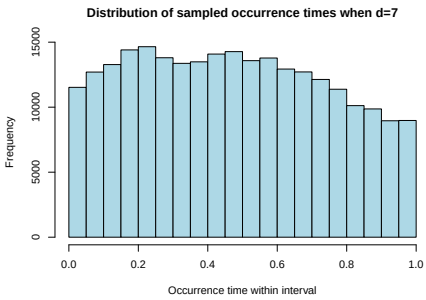
where the edge from $(1,0)$ to $(2,2)$ travels via an intermediate point $(2,1)$. The graph is shown in Figure 4.5. The numbers next to the edges denote their traversal times, which are also the edge weights.

The experimental setup is as follows: we first run Algorithm 4.9 (coupling from the past) to generate a realisation of a regular prior with parameter values $\beta = \log(2.3)$ and $\eta = -1$. The full realisation, under the assumption that a point within this realisation takes the form $w = [z,(a,a+l)]$, is

$$\mathbf{w} = \{[(0.040,0),(0.795,1.156)],[(0.552,0),(-0.176,1.250)],$$



(a) Histogram of frequencies of proposal times generated when $d = 3$.



(b) Histogram of frequencies of proposal times generated when $d = 7$.

$$[(1.366, 1.366), (0.475, 0.475)], [(0.681, 0.681), (0.324, 0.324)],$$
$$[(0.848, 0.848), (0.881, 0.881)]\}$$

In addition, we set all spatial and temporal parameters to 0 for demonstration purposes. The resulting network area-interaction point process realisation has been plotted in Figure 4.5. Non-atomic point 1 thus ranges between $[0.795, 1]$ in time, whereas non-atomic point 2 sits in the time interval $[0, 1]$. We also assume that the model parameter vector takes the form

$$\boldsymbol{\xi} = \{\delta_1, \delta_2, \alpha, k, b, c\} = \{1.7, 1.5, 6, 1, 1.3, 1/2\pi\},$$

where δ_1 ranges over $(-0.2, 0.4]$ and δ_2 ranges over $(0.4, 1]$. These parameters generate the semi-Markovian censoring mechanism, see [91, 96] for more details.

We initially set $r = 0.2$ and $d = 3$. We then run the Metropolis-Hastings algorithm from Chapter 4.3.3 for 250,000 iterations, throwing out the first 5,000 due to burn-in. From the perspective of non-atomic point 2, which can range within $[0, 1]$ in time, only non-atomic point 1 is contained in the interaction area when $d = 3$. A distribution of occurrence times for interval 2 for $d = 3$ is provided in Figure 4.6a. To see how this distribution changes when the interaction distance is changed, we set $d = 7$ and re-run the Metropolis-Hastings algorithm. This distribution is proviced in Figure 4.6b.

In the $d = 3$ case, since the point in interval 1 can take values in $[0.795, 1]$, we should expect fewer values in this range for interval 2, which can range within the entirety of $\mathcal{X}$, that is, $[0, 1]$. This is because the prior point process is regular in nature ($\eta = -1$), meaning that clustering is avoided. Figure 4.6a shows this behaviour, from which we can conclude that regular realisations are being generated correctly. Instead, occurrence times are more commonly found more than $r = 0.2$ time units away from the starting point of interval 1, which is before $t = 0.595$. If generation was incorrect, we would expect a uniform distribution,

or an increase in frequency in occurrence times in the region where both intervals overlap. Note that before $t = 0.2$, edge effects resulting from the choice of $r = 0.2$ take hold, meaning that fewer points are simulated here due to the changed proposal distribution. In the case where $d = 7$, the two atomic points occurring at time $t = 0.324$ and $t = 0.881$ are also within the interaction range. Due to the imposed regularity, the point process tends to avoid placing the point contained in the interval $[0, 1]$ very close to $t = 0.324$ and instead prefers adjacent regions. Similarly, there is increased aversion to the point $t = 0.881$. We can therefore see the effects that points in the interaction area of a non-atomic point have on the occurrence time distribution.

4.7 Discussion

In this chapter, an introduction to Euclidean graphs is provided with some simple examples. We discuss shortest path metrics, measure-theoretic details and possible methods by which graph edges may be parametrised. The area-interaction process [9] is then extended to Euclidean graphs, with Markov properties being verified under the chosen distance metric. We discuss simulation techniques for the network area-interaction process. Details on sampling from the full interval-censored model are also provided. One main novel contribution of this chapter is the development of parameter estimation methods for point processes based on estimating equations when realisations are only partially observed or obscured. These methods rely on the GNZ equation and use Geyer-style weights for the score equation. The stability of this equation is estimated via the Godambe information matrix.

In this thesis so far, we have focused entirely on probability and statistical theory, model building and simulation techniques for point processes and interval-censored data. In Chapter 5, we will provide background on potential applications of such models, outline results that have been obtained by using these models on real data and detail a full statistical application of the network area-interaction models to a car arson fire dataset.

CHAPTER 5

Applications in criminology

In this chapter, we focus on applications of the state estimation models discussed in earlier chapters of this thesis. To facilitate this, we first discuss the history of policing methods, as well as how certain crimes lead to data that may be amenable within the framework that we have developed. We will first review theories of criminal behaviour and introduce the subfield of aoristic crime data analysis. We subsequently lay out previously used methods in applied criminology and place the models developed in Chapters 2, 3 and 4 in context, showing that the models developed in this thesis add complexity that has not been considered in previous approaches. We then present results obtained by these models in an extended results subchapter.

5.1 Background on proactive and predictive policing

To increase the effectiveness of law enforcement, it is essential to have a deep understanding of criminal behaviour so that potential crimes can be averted before they are even committed. Proactive policing methods based on theories of social and criminal behaviour are crucial in this respect. Examples of this include hot-spots policing [132], broken windows policing [143], focused deterrence [80], also known as pulling levers policing, among many others. While the effectiveness of certain proactive policing methods is disputed (see, for example, [60] relating to broken windows policing, or [51] relating to police perception in general), these methods represent a broader shift from reactionary to proactive policing, whereby the police serves as a greater community force instead of simply focusing on apprehending suspects [107].

The development of statistical models for crime prediction can also be viewed through the lens of proactive policing. Such methods are often known as predictive policing methods [115], and allow for quantitative analysis of crime data. While these methods have been shown to be powerful, making significant contributions to successful proactive policing methods in certain cases [128], they are not always viable for certain types of offences [48]. Rarely reported offences are often not amenable to analysis due to a lack of data, and many otherwise viable, reported offences often contain a lot of uncertainty. An example of this is residential burglaries, as people often return home to see that their place of residence has been broken into, not knowing the actual time of occurrence. Arson presents a similar problem due to rarely being observed at the exact occurrence time. Though it may be of little consolation to the victim to know an approximate time of occurrence, it may be of great interest to the police, since policing strategies can be adapted. This information may also assist victims and the wider public in preparing for future burglaries.

While it may seem that these difficulties severely limit the effectiveness of anal-

ysis of imperfect data, criminologists have developed methods to deal with such data. Many criminologists noticed limitations in purely spatial analysis of crimes and wished to focus more on temporal data [5, 22, 64, 119, 120]. To deal with temporal uncertainty, they developed a scheme by which incomplete, censored crime data could be analysed temporally. This class of crimes was given the name "aoristic", a word derived from the Ancient Greek root aóristos meaning "without defined occurrence in time". Similarly, the aorist tense, which is present in languages such as Greek, describes a singular event occurring at an indeterminate time in the past. Aoristic crime data analysis falls under the subcategory of predictive policing methods, within the broader field of proactive policing methods, since knowing the most likely time of occurrence allows police to draft proactive policing strategies around the most likely occurrence times. These analyses could therefore be used to assist police in implementing proactive policing methods that play a role in strengthening institutions and ensuring safe and secure communities.

5.2 Review of theories on criminal behaviour

In 1979, Lawrence E. Cohen and Marcus Felson published the work "Social change and crime rate trends: a routine activity approach", describing the effect that routine activities have on crime patterns. Using criminological data, they made the argument that shifting trends in human behaviour led to an increase in the crime rate in the USA in the 1970s that initially seemed paradoxical [28]. The higher crime rate was surprising, as the country was experiencing increasing prosperity and significant social change, factors that were generally understood to lead to a decrease in crime.

This work, among others of its kind, led to a shift away from focusing on the characteristics and psychology of criminals, and towards social disorganisation theory [24]. This theory suggests that places of residence and general spatial factors play a large role in shaping the distribution of crime in a city, perhaps even more so than the psychology of an individual [49, 130]. Hawley's human ecological theory of community structure provides intuition in this respect – that time and location dependence underlie community structure itself and are crucial in organising a community [62]. Social disorganisation theory has waxed and waned in popularity through the years [82], especially since testing the theory empirically has been difficult [63]. In addition, [24] noted that a number of the perceived flaws with social disorganisation theory may come from misapplication of the theory, or that measuring social disorganisation may be difficult. Nevertheless, researchers were able to find support for the idea that "social-disorganisation theory has vitality and renewed relevance for explaining macro-level variations in crime rates" [127]. The idea therefore that time, place and community structure are relevant factors for criminal activity, instead of solely focusing on delinquent individuals, is well substantiated [82]. These ideas can be considered as building

blocks from which the statistical models applied in this chapter take inspiration.

Taking routines into account, as well as the effects that they have on crime, is often known as the routine activity approach [28]. For example, people living in neighbourhoods tend to build mental maps of their immediate surroundings, helping them avoid places that may be susceptible to crime [106]. This leads to a concentration of crimes in certain regions of cities where crime is "easy, safe and profitable" [20]. Resulting from this shift in philosophy within criminology is the realisation that spatio-temporal analysis has the potential to provide important insight into the field, as these methods aid in modelling neighbourhoods and community structures. Together with the ideas from social disorganisation theory outlined in the previous paragraph, this provides a basis upon which a model can be built.

As a result of the move towards spatio-temporal analysis, research has been done within criminology to determine the most probable locations of crimes [20, 48, 142]. However, research has not been focused on temporal interdependence to the same extent, even though the routine activity approach may also inform patterns in temporal variation [5, 28, 120]. Hawley wrote about three temporal factors of community structure: rhythm, tempo and timing. Rhythm refers to the periodicity of event occurrences, tempo to the number of events per unit of time and timing to the coordination of activities between members in a community [28, 62]. Due to several factors, such as women's increased participation in the workforce and an abundance of cars, more activities began to take place away from places of residence starting in the 1960s and 1970s, a trend that has continued to the present day [48] As a result, there was a shift in the times of the day at which residential crimes were more likely to be committed [28, 45, 46], especially for crimes that are opportunistic in nature [29, 47]. Consequently, it seems implausible to assume that crimes are equally likely to occur at any given time, an assumption often made in aoristic crime modelling methods. These methods, introduced by [120] and extended by [5, 119] among others, provide methodology for estimating the occurrence times in a temporal crime data set when only a time interval is known. This is often the case for crimes such as burglary or arson [120].

Additionally, inspired partially by Hawley's three temporal factors of community structure [62], it seems plausible that crimes in a given neighbourhood are influenced by each other. If the police decide to come down heavily on crimes at certain hours, crimes may start occurring at different times instead due to heightened police presence. Furthermore, in neighbourhoods where houses or apartments are very similar to each other in layout, a successful burglary may lead to an increase in crime in that neighbourhood due to the near-repeat effect [17, 133]. Therefore, it may be erroneous to assume that burglary times and frequencies remain constant over time.

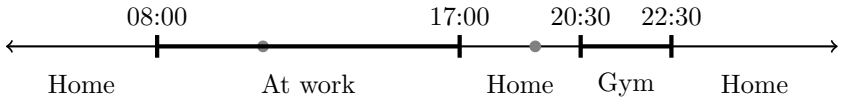


Figure 5.1 Example of the schedule of a potential victim. Note: The bold intervals denote times at which the victim is away from home.

5.3 Proactive and reactive policing methods

There is a key difference in philosophy between proactive policing methods, which is where the work presented in this chapter falls, and reactive policing methods. In the past, the police were often seen primarily as a reactive force, apprehending alleged offenders and responding to requests from citizens [121]. Throughout the 20th century, police departments pushed to modernise and professionalise, with an increased understanding that the police's role includes serving and protecting the community. Along with this came the idea of proactive policing, snuffing out crime before it even occurs [107].

The models developed in this thesis are most applicable to data gathered by hot spots policing, which is a method developed in 1995 by Sherman and Weisburd in the Minneapolis Hot Spots Patrol Experiment. The idea is to focus on certain geographic regions where crime incidents seem to be, or are empirically shown to have been, more likely than in other regions [132, 43]. This strategy has been shown to be effective at reducing crime in the identified hot spots, without it spreading to adjacent areas [131]. While there are certain drawbacks to this strategy, chief among them being its limited lasting effectiveness and narrow use [125], the National Research Council summarised that there is strong evidence for the effectiveness of hot spots policing [107].

The statistical models developed in this thesis work with the idea that hot spots do not only occur in place but also in time, an idea for which there is significant evidence. It is informed by theories such as the near-repeat effect [17, 133], the routine activities approach [28] and a number of other insights mentioned previously in this chapter (see, for example, [29, 106]). While this is a predictive policing method, relying on the use of statistical models and data to perform estimation within time windows, it assumes an inherent difference in likelihood between different times of the day, taking inspiration from hot spot-related methods.

5.4 Overview of methods

Numerous methods have been proposed to perform aoristic crime data analysis in the years since theory on this topic was developed. When performing this

analysis, one would like to find the most likely time at which the crime occurred, for every time interval in a given data set. Alternatively, one might like to know the entire distribution of occurrence times. In the spatial case, the data is the set

$$\mathbf{u} = \bigcup_{((z_i,t_i),(a_i,l_i))\in W} \{(z_i,(a_i,l_i))\},$$

with z_i being the location in space. We assume that there are n crimes. Let t_i denote the estimated time of occurrence for the ith crime in the data set so that $t_i \in [a_i, a_i + l_i]$.

5.4.1 Methods based on summary statistics

A naive approach might be to simply take the midpoint of each time interval as an estimate. One would then record this time as the time of occurrence [64]. Possible advantages of this method include that it is easy to calculate for large data sets, and that for small enough intervals, the exact estimate of the time becomes less important. However, the major disadvantage of this approach is that it is entirely arbitrary. Based on what is known about property crimes, there is no good reason why this particular value should be chosen over others.

Similar methods, such as arbitrarily picking the start or end points of the interval, have been described [5, 120]. [5] also suggested picking, with uniform probability, a completely random point within the interval. These methods all suffer from the same pitfall as the midpoint method - arbitrary time selection. Possible interactions between occurrence times are ignored, and the assumption that all occurrence times are equally likely within a given interval is still implicitly present, if occurrence times are selected uniformly.

5.4.2 Statistical approaches

Some more advanced methods have been developed to perform data analysis on aoristic crime data sets. Consider a weight function over t [119, 120]. Here, for a given value of t over $\mathbf{u}$

$$W(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}\{a_i \le t \le a_i + l_i\}}{l_i}.$$

Note that $\mathbf{1}$ is the indicator function, meaning that 1 is returned if t is contained in the interval $[a_i, a_i + l_i]$, and 0 otherwise. In this way, a length-weighted sum is returned for a given point in time. This denotes the likelihood of a crime occurring at time t. To make this more clear, consider the following example where

$$D = \{[0.1, 0.5], [0.2, 0.4], [0.3, 0.6]\}.$$

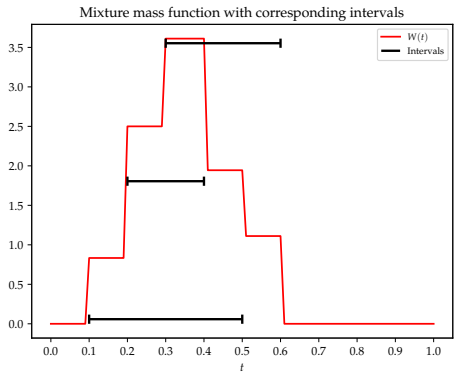


Figure 5.2 Example $D = \{[0.1, 0.5], [0.2, 0.4], [0.3, 0.6]\}$ plotted together with the values of $W(t)$ for t between 0 and 1.

The union of this set is $U = [0.1, 0.6]$. Evaluating W at $t = 0.2$ results in

$$\begin{aligned} W(0.2) &= \frac{1}{3}\left[\frac{\mathbf{1}\{0.1 \leq 0.2 \leq 0.5\}}{0.4} + \frac{\mathbf{1}\{0.2 \leq 0.2 \leq 0.4\}}{0.2} + \frac{\mathbf{1}\{0.3 \leq 0.2 \leq 0.6\}}{0.3}\right] \\ &= \frac{1}{3}\left[\frac{1}{0.4} + \frac{1}{0.2} + \frac{0}{0.3}\right] \\ &= 2.5. \end{aligned}$$

The weight function $W(t)$ can be evaluated for discretely many values of t in the observation window. In practice, one would generate many potential values of t, evenly spaced out across the observation window, and apply this function. As this is a probability mass function, one can sample from this distribution, picking random points based on how likely it is that points follow this pattern. While this method does add some complexity and considers multiple time intervals at once, occurrence times within a certain interval are biased towards time ranges within which multiple intervals intersect. It may be erroneous to assume that criminal activity always clusters in this way. Figure 5.2 shows the value of this function with the intervals overlain. One can see that the value of $W(t)$ is highest in the intersection of the three intervals.

5.4.3 Model-based approaches

While the methods outlined earlier can have relative success in certain cases, they fail to model interaction in the data which inherently affects the locations of intervals. For example, people are less likely to be home during work hours, resulting that a criminal would encounter less resistance when trying to break in at this time. Such behaviour patterns lead to a dynamic environment in which criminals act. It would be of interest to implement an interaction model, which would result in more accurate estimation of break-in times. We would also like to

Method	NAS	IM	TD	SP
Midpoint	No	No	No	No
Start/End	No	No	No	No
Random selection	No	No	No	No
Weight function	Yes	Partially	No	No
Homogeneous temporal model	Yes	Yes	No	No
Non-homogeneous temporal model	Yes	Yes	Yes	No
Spatial network model	Yes	Yes	Yes	Yes

Table 5.1 Qualitative analysis of aoristic crime analysis methods discussed in the text. *NAS* stands for non-arbitrary selection, *IM* for interaction modelled, *TD* for time-dependent and *SP* for spatial effects.

model potential occurrence times not as separate data points, but following some underlying distribution. Additionally, we want to partially separate occurrence times from the intervals within which they range. The models introduced in this thesis satisfy these requirements.

The models consider two processes: the burglar process, which contains the occurrence times, and the victim process, which contains the time intervals within which crimes occurred. The burglar process is the ground point process and therefore the prior, whereas the victim process is the forward model. The three models developed during this thesis are the fully homogeneous temporal model from Chapter 2, the non-homogeneous temporal model from Chapter 3 and the non-homogeneous spatial model from Chapter 4.

Benefits and drawbacks of all models introduced in this subchapter are outlined in Table 5.1. In the following subchapter, these models are run on real data sets, culminating in a full statistical application of the network model from Chapter 4, which is also the most complete model.

5.5 Results for temporal models

In this subchapter, we outline results for the two temporal models outlined in this thesis.

5.5.1 Description of data

For the temporal models, one standard crime data set has been used for model fitting. We have chosen the Washington D.C. burglaries data set, since this data set is one of very few to include precise start time and end times for burglaries within a small enough spatial area, as well as having enough data points to facilitate comparison (though this is still not optimal, see discussion for details). The data set is called *Crime Incidents in 2016*, and is a record of all reported

crimes within the city limits of Washington D.C. in the year 2016 [113]. It contains 37,189 records, of which 2,121 are burglaries. The model is run on a subset of this data set, specifically burglaries where the end times fall within the month of February. Only burglaries with a defined start and end date have been selected and other crimes have been filtered out. The model has been restricted to a subset of the data for computational reasons. While frequency of occurrence of crimes almost certainly depends on time of day and season, this can be accounted for by the model for both the victim and the burglar process by choosing different time ranges (see Table 4.1).

5.5.2 Exploratory results for homogeneous temporal model

In the case of the Washington D.C. data, a Gamma model proved to be unsatisfactory for the length distribution of intervals. Specifically, the salient feature of the collection of observed non-degenerate intervals is that it combines a large number of small intervals with a few very long ones, which cannot be captured by a Gamma distribution with shape parameter larger than one. Therefore, we instead used the heavy-tailed Weibull distribution. Maximum likelihood methods for the Weibull distribution can be found in Chapter 2.5.3. For the Washington D.C. data, $\hat{\alpha} = 0.27$ and $\hat{k} = 0.000046$.

Using the methods outlined in Chapter 2.7, the model was run on the Washington D.C. burglary data set for February 2016. With model parameters tuned, prior parameters were estimated and realisations were generated with this set of parameter values after the Monte Carlo EM procedure was run until convergence. One such realisation has been plotted in Figure 5.3. Note that for demonstration purposes, only the last week of February is shown.

The prior parameter estimation led to estimates of $\hat{\beta} = 115.469$ and $\hat{\eta} = -0.256$, meaning that the model assumed a regular underlying structure. While one may assume that criminology data sets would lend themselves better to clustered priors, this may not always be the case. For example, it may be the case that burglars wish not to interfere with each other, leading to burglary times being more spread out.

The homogeneous model provides solid parameter estimates for β and η, but more information and insights can be obtained by using the non-homogeneous model from Chapter 3.

5.5.3 Results for updated Bayesian model

The updated Bayesian model is run on the aforementioned data. By visual inspection of the data, there seem to be fewer victims reporting crimes in the last few days of the month. Following this intuition, the time range A_1 is set to

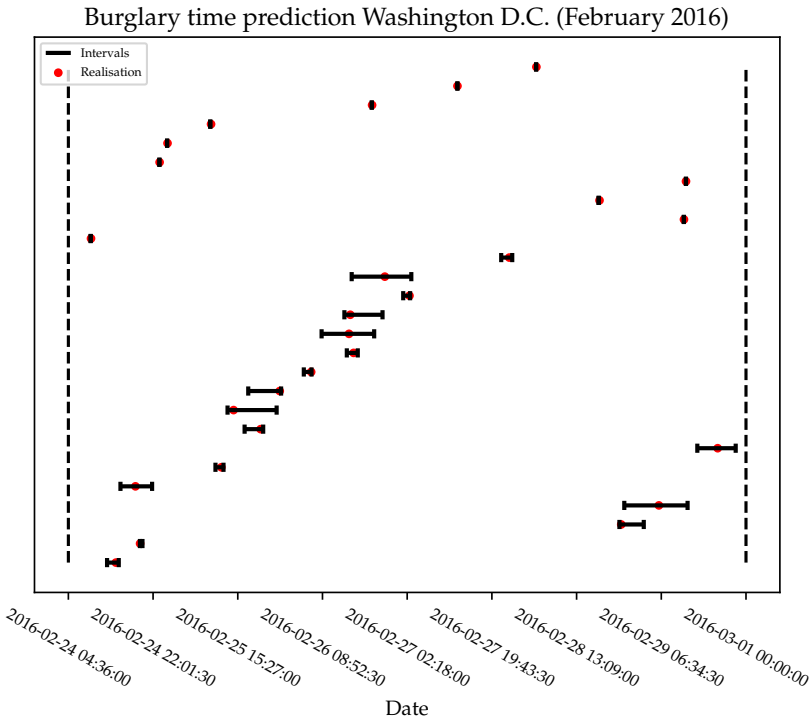


Figure 5.3 Model output for the final week of February 2016 with one realisation in black. Prior parameters were estimated to be $\beta = 115.469$, $\eta = -0.256$ with model parameters $r = 0.008$, $k = 0.27$ and $\lambda = 0.000046$.

$[-0.2, 0.9]$, and $A_2 = [0.9, 1]$ (see Table 4.1 for interpretations). For these sets, the value 0 refers to midnight on the 1st of February, and 1 likewise represents midnight on the 1st of March. The choice of a negative value in A_1 ensures that crimes that may have occurred before the 1st of February, but whose end times still fall in the month, are also included. Hence, the values of δ_1 and δ_2, corresponding to the time ranges in question, must be estimated. We set $K = 1$, meaning we must only calculate one β.

Figure 5.4 shows the time intervals and corresponding estimated occurrence time for the last week of February. As there were 120 reported burglaries in February 2016, plotting the entire month would lead to a much less visually appealing graphic. See Table 5.2 for the exact occurrence times estimated by the model.

Estimation resulted in the following parameter estimates: $\hat{\delta}_1 = 0.8, \hat{\delta}_2 = 0.5, \hat{\beta} = 115.06$ and $\hat{\gamma} \approx 0$. Recalling that $\eta = \log \gamma$, we see that $\hat{\beta}$ is similar to the homogeneous model output, and that interactions are still regular. That $\hat{\delta}_1$ does

not equal $\hat{\delta}_2$ shows that there does seem to be an inherent difference in likelihood of crime occurrence in the last few days of February. The parameter β can be seen to be approximately equal to the number of crimes that one would expect in this configuration. As γ is much smaller than 1, this implies that crimes do not seem to cluster, and instead occur further apart from each other in time.

This means that there is strong evidence for regularity. A possible counterpoint could be that there were "only" 120 crimes in 28 days. For non-overlapping crimes, no assumption can be made regarding behaviour with respect to other crimes, leading to less information being fed to the model. While the estimation procedure for γ can still be carried out, as less information is available, the exact value may be somewhat imprecise.

One might expect greater clustering, as it makes sense that crimes do not occur at an equal rate across time. However, it may be the case that the spaced-out nature of the estimated occurrence times is the model taking into account the inherent periodicity of the data. That is to say, as more likely times are followed by less likely times in roughly equal measure, points may be clustered in more likely regions, but are spread out with respect to the entire data set due to this phenomenon.

The parameters α and k take the values 2.2 and 0.3 respectively, meaning intervals are relatively long compared to the length scale, but that there are comparatively more "short" intervals than "long" intervals. In other words, people tended to be away for a relatively long time, but there were very few instances where people were away from their properties for a period that was significantly longer compared to fellow victims. Though this can be determined in other ways, such as visual inspection or summary statistic estimation methods, it shows that the model accounts for this behaviour when performing estimation.

While potential hot spots can be detected by using approaches such as the mixture mass function [119, 120] (see also Figure 5.2), this ignores other factors almost undoubtedly present for residential burglaries, and biases potential occurrence times towards intersections of time intervals. By looking at the most likely occurrence times estimated by the non-homogeneous estimating procedure, potential hidden hot spots in time can be detected, leading to the potential implementation of proactive policing methods based on this model. In addition, the non-homogeneous parametrisation method for time intervals used in the estimation process allows for the generation of more accurate test data sets. This may aid future researchers who wish to implement different occurrence time estimation methods to validate their models.

This model rectifies issues present in the estimation of occurrence times for aoristic crime data, but some shortcomings should also be mentioned. Firstly, no

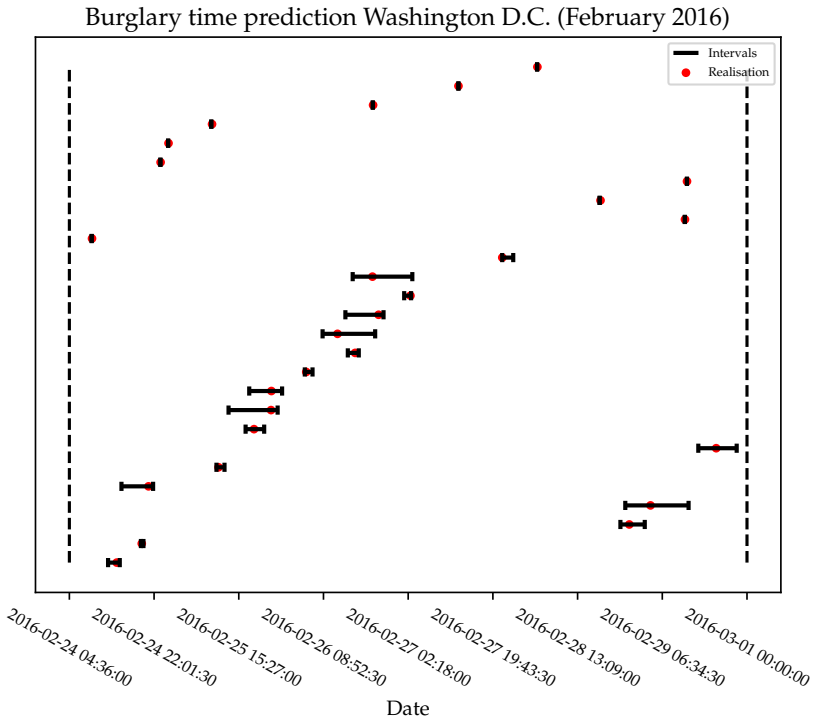


Figure 5.4 Simulation of the occurrence time of crimes from the Washington D.C. aoristic crime data set. Estimated burglary times are marked by a red point, the intervals in black.

spatial information at all is taken into account, which undoubtedly limits the effectiveness of the model. Secondly, this particular model assumes that the burglar and victim processes are more or less independent. Relaxing this assumption leads to greater difficulty in applying a model of this type, but may be more accurate in describing the actual behaviour of criminals and victims.

A more powerful model might make use of spatial covariates to inform crime rates at specific locations since it is known that location affects crime rate [20, 132]. Therefore, in the following subchapter, we apply the spatio-temporal model from Chapter 4 to a data set of car arson fires.

5.6 Spatio-temporal model application

In this application, we will use car arson data from the Safety Region Twente. The data was collated by the Twente fire department in collaboration with the police, and describes locations of car arson attempts within the regional subdivi-

Start time	End time	Estimated time
2016-02-24 12:33:42	2016-02-24 14:57:56	2016-02-24 14:20:31
2016-02-24 19:19:06	2016-02-24 19:50:11	2016-02-24 19:29:24
2016-02-28 21:57:38	2016-02-29 02:58:04	2016-02-28 23:47:12
2016-02-28 22:57:47	2016-02-29 11:58:53	2016-02-29 04:09:10
2016-02-24 15:18:58	2016-02-24 21:49:32	2016-02-24 20:52:49
2016-02-25 10:51:39	2016-02-25 12:30:48	2016-02-25 11:16:54
2016-02-29 13:59:05	2016-02-29 21:52:47	2016-02-29 17:39:41
2016-02-25 16:51:11	2016-02-25 20:40:31	2016-02-25 18:35:16
2016-02-25 13:20:44	2016-02-25 23:26:38	2016-02-25 22:05:17
2016-02-25 17:36:05	2016-02-26 00:21:40	2016-02-25 22:09:38
2016-02-26 05:02:09	2016-02-26 06:37:17	2016-02-26 05:19:53
2016-02-26 13:52:58	2016-02-26 16:08:10	2016-02-26 15:23:10
2016-02-26 08:41:32	2016-02-26 19:30:27	2016-02-26 11:45:55
2016-02-26 13:23:02	2016-02-26 21:13:43	2016-02-26 20:12:42
2016-02-27 01:28:44	2016-02-27 02:53:50	2016-02-27 02:45:38
2016-02-26 14:53:04	2016-02-27 03:09:08	2016-02-26 18:57:18
2016-02-27 21:38:32	2016-02-27 23:54:42	2016-02-27 21:39:30

Table 5.2 Start times, end times and estimated times of the burglaries in the Washington D.C. data set which have non-zero interval lengths.

sion of Twente, located in the province of Overijssel in the eastern Netherlands in a five-year period between 2017 and 2021. We would like to thank the Twente fire department for providing us with this data. For privacy reasons, the locations provided in the data set do not match up exactly with the real physical location, but have been shifted through the use of random noise. We have decided to perform analysis on car arson fires occurring solely within the city limits of Enschede, the largest city in the region.

The car arson data consist of 142 points, and have both a censored temporal component and a spatial location. The censored temporal component consists of a starting time and an end time, which produces a time interval within which the unknown occurrence time is located. The spatial location takes the form of x-y coordinates in the RD coordinate system, a mechanism by which distances within the European Netherlands can be accurately estimated. See Figure 5.5 for the spatial component of the data, and Figure 5.6 for the temporal component, with the median of each interval being taken. We aim to provide an estimate of the actual occurrence time given the spatial location and time interval structure.

In the data, the points are sometimes not exactly positioned on roads for a few reasons. Firstly, parking spaces for cars are often not on the street itself, but adjacent to it. Secondly, due to rounding of RD coordinates to three (3) decimal places, it is possible that a car that was parked on the street appears slightly to

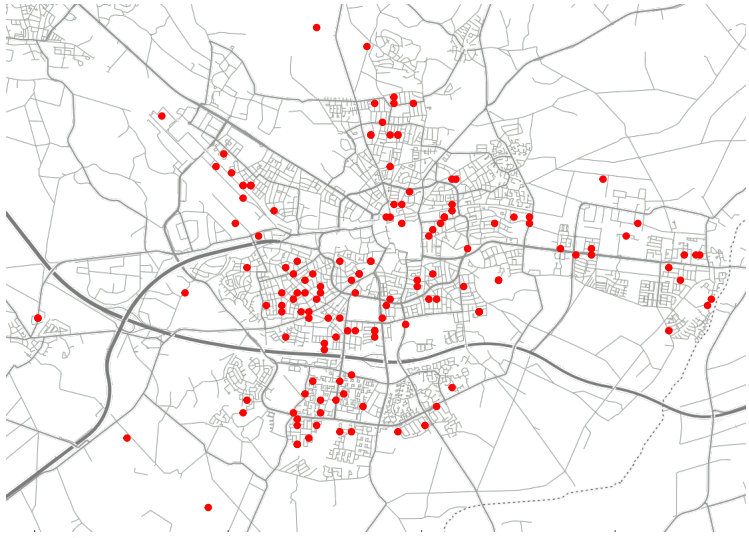


Figure 5.5 A map of the locations of car arson fires in Enschede, with occurrences marked in red.

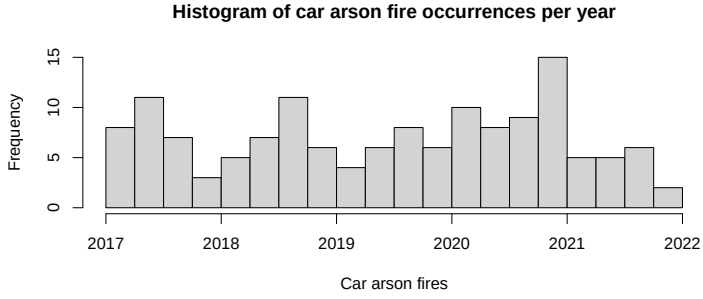


Figure 5.6 Histogram of median (in time interval) car arson fire occurrence in Enschede, separated by year. Each bin represents a time period of three months, roughly representing the seasons.

the side of it. Thirdly, cars in garages or other parking structures can also be set on fire, and are therefore not parked near a road. To account for this, we will assume that all points not located on a street are actually located at the nearest point on the street network to the point, in space.

The underlying road network must also be modelled. To do this, we will use a Euclidean graph structure. In practice, we will separate each street in the road network into discrete line segments, so that points on the graph correspond closely to positions on the actual network. This allows for curved roads to be

estimated by a number of smaller line segments. To perform analysis, we will use the Python package OSMnx [19] and the C++ library MPPLIB [92], which has been extended to allow for complex censoring mechanisms like those developed in Chapters 2 and 3.

5.6.1 Description of covariate data and data cleaning

If covariates are to be considered in the full model, we require a source of data for both the spatial and temporal components. In the spatial dimension, we have decided to use a data set from Statistics Netherlands (CBS, website: `https://www.cbs.nl`). This data set imposes a grid of 500m x 500m squares within the European Dutch borders, and provides statistical information such as average income, number of houses, distance to the nearest train station and many others, per square [84]. This data is provided in table form. We have decided to take data from 2021, since this is the last year in which data was collected. Since these statistics do not differ significantly on a year-to-year basis, the exact year chosen is not of much relevance.

Due to low population density in certain squares and the possibility of identification of specific residents inherent to this, certain fields did not have values in certain squares. To rectify this, for any fields counting the number of objects or people in squares with a very low number or residences, we have set all values to 0. In addition, for any fields which denote percentages or closest distances, the global average over values for which data is present has been inserted. As we have picked the most populated city in the region as the study area, the contribution of squares with low population density is minimal.

Provided below is a table of covariate field names (metadata) as seen in Figure 5.7, along with a description. Full details are provided in [84] (in Dutch).

For the temporal data, data from the Royal Netherlands Meteorological Institute (KNMI, website: `https://www.knmi.nl/home`) has been used, specifically data from the *Twenthe* [sic, archaic spelling] weather station, located in the Twente region of the Netherlands and the nearest weather station to where the car arson fire data set was recorded. This data is in table form, containing weather observations such as humidity, wind direction and temperature on a per-day basis. Rarely, defects in the recording equipment led to certain days missing a subset of the recorded data. In these cases, data from the central weather station of De Bilt, located approximately 115 kilometres from Enschede as the crow flies, have replaced the missing observations at the Twenthe weather station.

Label name in plot	Description of label	Label name in plot	Description of label
build_tot	Number of buildings	ave_hhold_size	Average household size (in people)
build_hotel	Number of hotels	num_prop	Number of properties
build_hlth	Number of healthcare buildings	num_multifam_prop	Number of multi-family properties
build_indu	Number of industrial buildings	p_sale_prop	Owner-occupied property percentage
build_off	Number of office buildings	p_rent_prop	Rental property percentage
build_logs	Number of warehouses	num_housing_corp	Number of properties owned by housing corporations
build_edu	Number of educational buildings	num_prop_vacant	Number of vacant properties
build_sport	Number of sport buildings	ave_prop_worth	Average property value
build_shop	Number of shops	p_low_income	Low-income percentage
build_homes	Number of homes	p_high_income	High-income percentage
build_oth	Number of other buildings	med_income_hhold	Median household income
tot_num_res	Total number of residents	num_benefits	Number of residents on non-pension government benefits
num_res_015	Number of residents between 0-15	d_cafe_km	Distance to nearest cafe
num_res_1525	Number of residents between 15-25	d_establ_km	Distance to nearest eating establishment
num_res_2545	Number of residents between 25-45	d_fire_station_km	Distance to nearest fire station
num_res_4565	Number of residents between 45-65	d_mainroad_km	Distance to nearest main road
num_res_65	Number of residents above 65	d_trans_tstation_km	Distance to nearest train station with transfer
num_hholds	Number of households	d_tstation_km	Distance to nearest train station
num_sin_res_hholds	Number of single resident households	d_themepark_km	Distance to nearest theme park
num_mult_res_hholds	Number of multiple resident households	d_cinema_km	Distance to nearest cinema
num_sinpar_hholds	Number of single parent households	add_per_kms	Number of addresses per square kilometre
num_multipar_hholds	Number of multi parent households	urbanity	Urbanity index

Table 5.3 Spatial covariate metadata table.

5.6.2 Model selection

We now tackle the problem of selecting a small subset of covariates from the large set of possible covariates available to us. The arson fire counts per day and statistical area, depending on whether or not the temporal or spatial covariates are to be analysed, will from this point on be referred to as the response variable. We have used a conditional random forest technique [66] to perform model selection based on covariates.

The random forest, introduced in [21], is a classification method that uses a large number of decision trees on subsets of the data to predict response variables based on a number of predictors, often also referred to as explanatory variables [136]. Our predictors are a number of potential covariates, and the response variable is the number of arson fires. The output of such an analysis is an importance score for each variable, which measures the mean increase in prediction error over all trees, compared to when this variable is excluded [21, 136]. While random forests can be very effective, they often struggle when explanatory variables are strongly correlated. Therefore we will be using conditional random forests [66] in combination with permutation importance techniques [35, 135], as implemented in the R package `party` [137]. These conditional inference techniques have been shown to reduce importance measure bias [136]. Similar techniques have also been

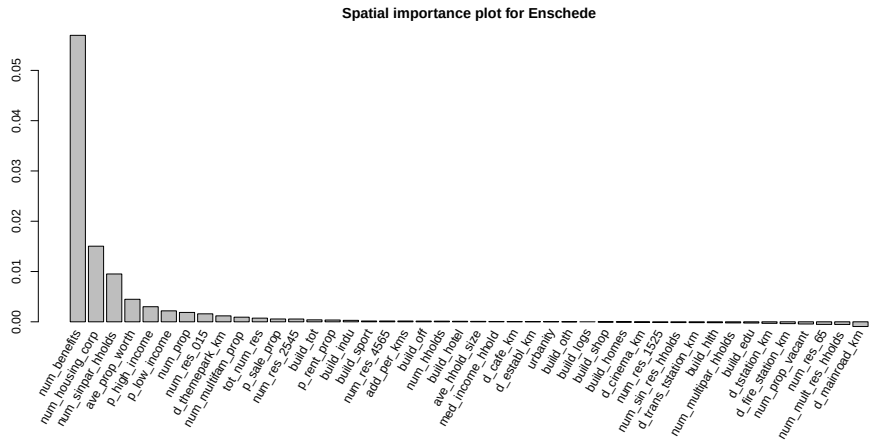


Figure 5.7 A plot detailing the importance values (y-axis) in the spatial domain generated by a conditional random forest analysis using the R `party` package.

used in [93] in to identify relevant covariates for chimney fire risk prediction.

When fitting such a model using the `party` package, two parameters play an active role in how the classification is performed: the number of trees used, and the number of input variables that are randomly sampled [137]. We analyse spatial and temporal explanatory variables separately, and will therefore fit two separate conditional random forest models on the covariate data. In the spatial case, we set the number of trees to 5000, since this was the point at which we observed no significant change in the output. In the temporal case, it was observed that 2000 trees were enough for convergence. For both the spatial and temporal covariate data, we have set the number of randomly sampled input variabes to 1/3 of the total number of input variables.

After collation of spatial and temporal covariate data, we ran conditional random forests on both data sets independently to determine the most relevant explanatory variables. The results can be viewed in the form of R plots in Figures 5.7 and 5.9. For an overview on the abbreviated axis labels in the spatial case, we direct the reader to Table 5.3.

Figure 5.7 shows the importance scores for all spatial explanatory variables, taken from the CBS data set outlined in Chapter 5.6 [84], against the response variable, the spatial location of recorded arson fires.

We observe that three explanatory variables have an importance score higher than 0.01. These are:

1. the number of people on government benefits excluding pension, functional form $\zeta_{1,\sigma}(z)$,

2. the number of residences owned by housing corporations, functional form $\zeta_{2,\sigma}(z)$, and

3. the number of single-parent households, functional form $\zeta_{3,\sigma}(z)$.

This means that these three variables are relatively very important for prediction of the response variable, compared to other explanatory variables. Note that this does not mean that the perpetrators are necessarily likely to fall in any of these categories, but rather that arson fires tend to occur in areas with these indicators. As a result, we have decided to include these three variables in the covariate analysis.

To determine the exact model form for the spatial component of $\beta(z,t)$, we proceed by plotting covariate function values against the response variable (in this case, the number of arson fires) to deduce the general trend. To do this, we took the average of the response variable over a range of values in the domain of $\zeta_{1,\sigma}$, $\zeta_{2,\sigma}$ and $\zeta_{3,\sigma}$, and plotted these against each other in Figure 5.8. Based on these plots, we decide to go for polynomial forms of the covariate functions.

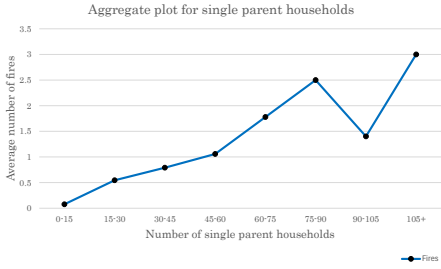
The plots in Figure 5.8 seem to show a quasi-linear trend. Additionally, the covariates are highly correlated, which may lead to issues in parameter estimation. We therefore choose to include only $\zeta_{1,\sigma}(z)$, the most important covariate function based on the conditional random forest analysis. Based on Figures 5.7 and 5.8, we propose a log-linear form for the spatial component of the intensity, which is

$$\lambda(z) = \exp\left(\theta_1 \zeta_{1,\sigma}(z)\right), \tag{5.1}$$

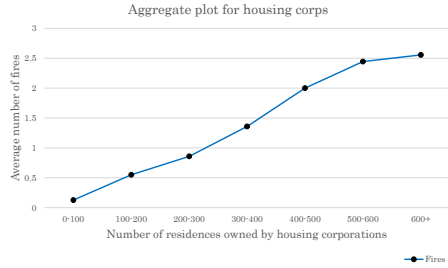
where θ_j is the coefficient corresponding to the jth term. The exponential function is employed to guarantee that $\lambda(z) \geq 0$.

5.6.3 Temporal covariate analysis

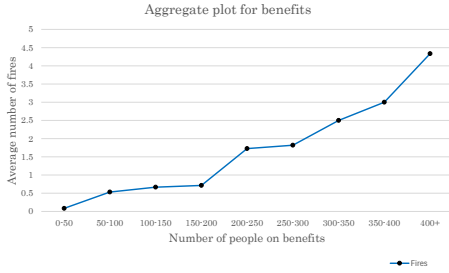
We perform the conditional random forest analysis on the KNMI data outlined in Chapter 5.6, with the response variable being the mean car arson fire occurrence times. In the temporal case, shown in Figure 5.9, the analysis is much less clear-cut. No single explanatory variable seems to be much more important in predicting the response variable than any other, with very low importance scores across the board. Therefore, we have not decided to include any variables based on the random forest analysis. Not entirely satisfied, we decided to look for trends in the data by plotting the number of arson fires by day and by month (Figure 5.10). For an overview of axis labels, please see Table 5.4.



(a) Plotting discrete bins of $\zeta_{1,\sigma}$ against the response variable.



(b) Plotting discrete bins of $\zeta_{2,\sigma}$ against the response variable.



(c) Plotting discrete bins of $\zeta_{3,\sigma}$ against the response variable.

Figure 5.8 Plots for the average number of fires for bins of different covariate values. These show the average trend of the response variable as the values of the covariate function increase.

Label name in plot	Description of label
ave_wind_direction	Average daily wind direction in degrees (360 = north, 90 = east, etc.)
ave_daily_wind_speed	Average daily wind speed (in 0.1 m/s)
max_wind_speed	Maximum hourly mean wind speed (in 0.1 m/s)
min_wind_speed	Minimum hourly mean wind speed (in 0.1 m/s)
max_wind_gust	Maximum wind gust on that day (in 0.1 m/s)
daily_mean_temp	Daily mean temperature (in 0.1 degrees Celsius)
min_temp	Minimum temperature (in 0.1 degrees Celsius)
min_temp_10cm	Minimum temperature measured at 10cm above ground (in 0.1 degrees Celsius)
max_temp	Maximum temperature (in 0.1 degrees Celsius)
pc_of_max_sunshine	Percentage of maximum sunshine duration on that day
global_radiation	Amount of energy irradiated by the sun on that day (in J/cm^2)
precip_duration	Duration of precipitation (rain, snow or hail, in 0.1 hr)
precip_amount	Total amount of precipitation (in 0.1 mm, -1 if less than 0.05 mm)
max_precip_amount	Maximum hourly precipitation amount (in 0.1 mm, -1 if less than 0.05 mm)
daily_mean_pressure	Daily mean sea level pressure (in 0.1 hPa, hectopascals)
max_pressure	Maximum hourly sea level pressure (in 0.1 hPa)
min_pressure	Minimum hourly sea level pressure (in 0.1 hPa)
min_visibility	Minimum visibility on that day*
max_visibility	Maximum visibility on that day*
mean_cloud_cover	Mean daily cloud cover (in octants, 9 if sky is entirely clear)
mean_humidity	Daily mean relative atmospheric humidity (in percent)
max_humidity	Maximum relative atmospheric humidity (in percent)
min_humidity	Minimum relative atmospheric humidity (in percent)
evapotranspiration	Potential evapotranspiration, calculated using the Makkink equation [95]

Table 5.4 A table of temporal covariate field names (metadata) along with a description.

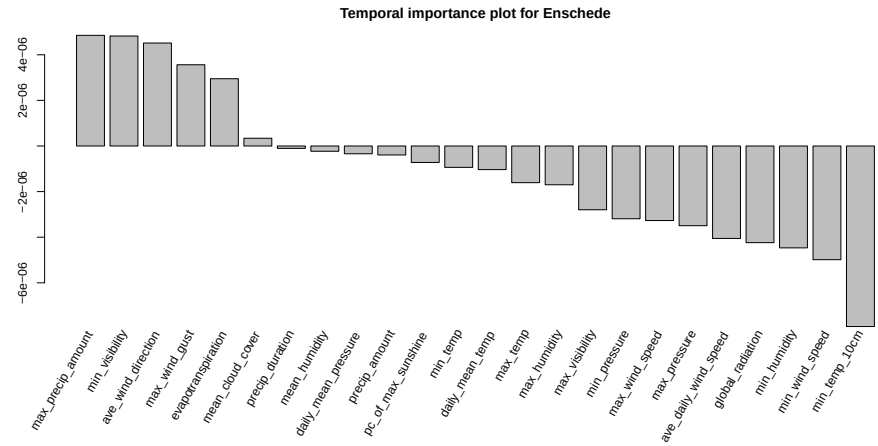
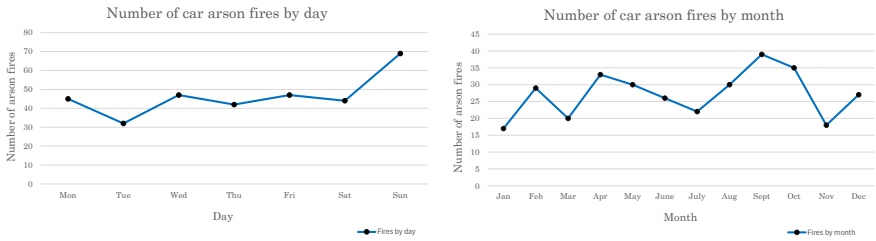


Figure 5.9 A plot detailing the importance values (y-axis) in the temporal domain generated by a conditional random forest analysis using the R `party` package.



(a) The amount of car arson fires by day. **(b)** The amount of car arson fires by month.

Figure 5.10 A breakdown of the number of arson fires per time period.

*Scale for visibility: 0: $<$ 100 m, 1: 100-200 m, 2: 200-300 m, ..., 49: 4900-5000 m, 50: 5-6 km, 56: 6-7 km, 57: 7-8 km,..., 79: 29-30 km, 80: 30-35 km, 81: 35-40 km, ..., 89: $>$ 70 km.

The monthly data seem to show that there is a somewhat increased prevalence in the warmer months, but not enough that anything concrete can be concluded. This can also be observed in Figure 5.6, which shows some seasonality in 2017, 2018 and 2019, with different behaviour occurring during the Covid-19 pandemic. What is more evident is that there seems to be a significant increase in arson fires on Sunday, independent of month and year. We therefore propose the form

$$Z(t) = \exp(\theta_2 \zeta_{T,\tau}(t))$$

for the temporal component of the intensity, where $\zeta_{T,\tau}(t)$ is an indicator function taking the value 1 if it is Sunday at a given point in time t, and 0 otherwise.

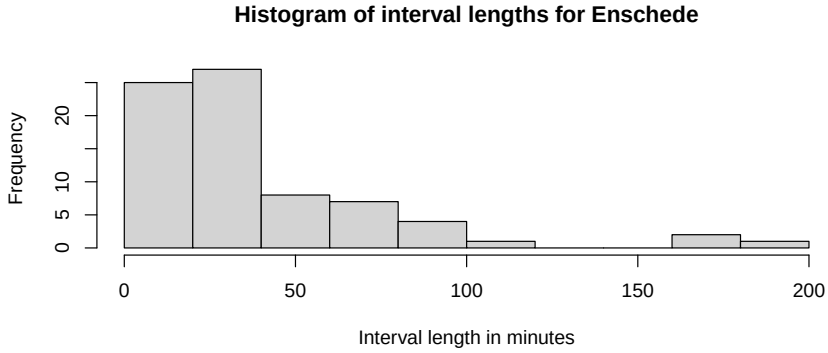


Figure 5.11 A histogram of interval lengths from the Enschede car arson data set.

There may be some interaction between the spatial and temporal paradigms. Therefore, we also add an interaction term between the two chosen covariate functions. This becomes

$$I(z,t) = \exp\left(\theta_3 \zeta_{1,\sigma}(z)\zeta_{T,t}(t)\right).$$

We hence propose the form $\beta(z,t) = \exp(\theta_0)\lambda(z)Z(t)I(z,t)$, where θ_0 is an intercept term. Combining Equation 5.1, the temporal component $Z(t)$ and the interaction $I(z,t)$, we obtain the full covariate functional form

$$\beta(z,t) = \exp\left(\theta_0 + \theta_1 \zeta_{1,\sigma}(z) + \theta_2 \zeta_{T,\tau}(t) + \theta_3 \zeta_{1,\sigma}(z)\zeta_{T,t}(t)\right). \tag{5.2}$$

5.7 Results for spatio-temporal model

5.7.1 Forward model parameter estimation

Having verified that the generation procedure works correctly, we move on to parameter estimation for the forward model in the data example. This involves estimating the parameter vector $\boldsymbol{\xi} = (\xi_S, \xi_R)$ from Chapter 4.

We assume that the time intervals present in the data have been generated by a semi-Markov process. In this case, we must estimate parameters for two functions. The first part of the parameter vector is $\xi_S = (\alpha, \kappa, b, c)$, and refers to parameters contained in the function g_Y, the derivative of the semi-Markov kernel. For more information, see [91] and/or Table 4.1. Represent a time interval by $[a, a+l]$, where a is the left-most point and l is the interval length.

Figure 5.11 plots the interval lengths of the data. Of the 142 data points, 20 have a length of longer than 200 minutes and thus been omitted from this plot.

Based on this, for this application we assume that g_Y follows a non-homogeneous Weibull distribution with rate parameter $\lambda_r(a) = \alpha\,(b + \sin(ca))$. Hence

$$g_Y(a, l) = \kappa\lambda_r(a)[l\lambda_r(a)]^{\kappa-1}e^{[-l\lambda_r(a)]^{\kappa}}.$$

In addition, we also must estimate the parameters $\xi_R = (\delta_1, \ldots, \delta_J)$ for the function m, which is the derivative of the Markov renewal function, see Chapter 3. This takes the form of a step function

$$m(t) = \sum_{j=1}^{J} \delta_j \mathbf{1}_{A_j}(t),$$

where $t \in \mathbb{R}^+$, $A_1, \ldots, A_J$ split up the observation window W_T and the δ_j are parameters (see row 5 of Table 4.1 for more information). This approach is particularly powerful if data is not evenly spread out over time [91]. In this application, we have decided to set $J = 1$, meaning that $W_T = A_1$ and only δ_1 must be estimated.

As explained in Chapter 4.4.1, $\xi_S = (\alpha, \kappa, b, c)$ refers to the parameters associated with the function g_Y, and $\xi_R = (\delta_1)$ refers to the parameter associated with the function m. The values of b and c, which are typically not estimated and set in advance, have been set to $b = 1.3$ and $c = 1/2\pi$ respectively. Therefore, the parameters $(\alpha, \kappa, \delta_1)$ remain to be estimated. After running the maximum likelihood-based procedure from Chapter 3.6.1 on the time intervals from the data set, we obtain the values $\alpha = 2.40, \kappa = 1.00$ and $\delta_1 = 0.48$, to two decimal places. Since $\kappa = 1$, the form of g_Y becomes exponential. These parameter values are subsequently used when generating intervals for the dummy process D from Chapter 4.4.3.

5.7.2 Results for forward model

Having verified that the generation procedure works correctly, we move on to parameter estimation for the forward model in the data example. This involves estimating the parameter vector $\boldsymbol{\xi} = (\xi_S, \xi_R)$. As explained in Chapter 4.4.1, $\xi_S = (\alpha, \kappa, b, c)$ refers to the parameters associated with the function g_Y, and $\xi_R = (\delta_1)$ refers to the parameter associated with the function m. The values of b, c and δ_1 have been set to 1.0, 0 and 1.0 respectively due to the lack of periodicity. Therefore, the Weibull parameters (α, κ) remain to be estimated. We obtain the values $\alpha = 1.00e\text{-}6$ and $\kappa = 3.40e\text{-}1$, to two decimal places. The final parameter vector $\boldsymbol{\xi}$ becomes $\boldsymbol{\xi} = (1.00e\text{-}6, 3.40e\text{-}1, 1.00, 0.00, 1.00)$.

5.7.3 Prior parameter estimation results

We must first decide on suitable values of the "irregular" parameters r and h. As mentioned in row 3 of Table 4.1, these parameters dictate the size of the interaction area. Note that in this application, we consider a time-based metric for the

interaction distance to take speed limits into account. We apply a profile likelihood method, in which we run the model with a variety of possible values of r and h and choose the value that gives the highest value of the log-pseudolikelihood. We observe that beyond the parameter values $h = 100$, calculation becomes intractable. For any value of r above approximately 0.007, the interaction area in time is more than twice the length of any individual interval, making analysis less meaningful. In addition, beyond these values of h and r, estimated parameter values remain reasonably constant and the log-pseudolikelihood does not change significantly. We choose $h = 100$ seconds and $r = 0.006$ as the irregular parameters. This means that all points reachable by driving 100 seconds at the maximum road speed are included in the interaction area, as well as all events occurring within a time span of roughly 5 days before or after each event. In practice, the maximum speed limit is likely an upper bound on average driving speeds, meaning that travel times are likely overestimates.

In addition, due to the point pattern consisting of 142 points, the unnormalised density quickly blows up when sufficiently large spatial interaction distances are used. Therefore, in this application, the choice has been made to use the log-unnormalised densities $\log h_\psi(\mathbf{x})$ and $\log h_\theta(\mathbf{x})$ for the weights. Using $\boldsymbol{\xi}$ from Chapter 5.7.2, we generate an instance $\mathbf{d}$ of the dummy process D with $r_D = 16$ empirically.

After some exploratory analysis, we decide to set the initial parameter vector to $\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3, \psi_4) = (3.0, 2.3, 2.6, -3.7, -0.065)$. We then run the Metropolis-Hastings algorithm described in Chapter 4.3.3 under parameters $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ for 100,000 time steps, at which point we know that convergence has occurred. We subsequently obtain $K = 500$ samples $\{\mathbf{y}_1, \ldots, \mathbf{y}_{500}\}$ by sampling from the converged Markov chain every 1000 time steps. These samples and the realisation of the dummy process allow us to evaluate and solve Equation 4.19 for $\boldsymbol{\theta}$ using the Newton-Raphson method from [134]. Results for the parameter vector $\boldsymbol{\theta}$ are provided in Table 5.5.

A value of $\eta = 0.000845$ implies that the behaviour of arsonists is near random. There may be multiple reasons for this. Random interactions imply that car arson fires are not affected by other car arson fires occurring in the area. As shown earlier, certain areas and times have a higher incidence of car arson fires. Car arson can therefore be seen as a fairly spontaneous and opportunistic crime with heightened incidence at certain times and certain spatial locations. Additionally, no observed intervals overlap in time and only a small number of intervals overlap in spatial interaction area, meaning that points that can interact only make up a small proportion of the data set. An increased incidence of fires may lead to the detection of interactions, but based on the available data there is simply not enough evidence to suggest any form of clustering or regularity. As discussed earlier, increasing the interaction area may lead to the detection of spurious

Par	Value		g_{i0}	g_{i1}	g_{i2}	g_{i3}	g_{i4}
θ_0	4.055e0	g_{0j}	6.971e-1	2.189e-1	1.014e-1	1.079e-1	1.044e1
θ_1	3.078e0	g_{1j}	1.759e0	1.590e0	9.236e-1	9.331e-1	3.698e1
θ_2	1.916e0	g_{2j}	1.344e0	4.082e-1	1.749e0	3.297e-1	2.008e1
θ_3	-2.638e0	g_{3j}	0.884e0	5.692e-1	1.056e0	6.141e-1	1.769e1
θ_4	8.450e-3	g_{4j}	1.963e1	1.545e1	6.017e0	5.455e0	3.739e2

Table 5.5 Estimated parameter and Godambe information values from Chapter 5.7.3.

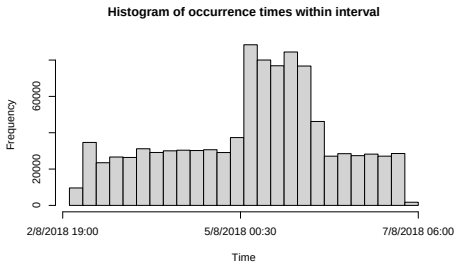
interaction and is computationally infeasible. Therefore, we conclude that while car arson fires are spatially and temporally heterogeneous, there is not enough evidence to suggest anything but random behaviour between individual fires.

5.7.4 Stability of solution

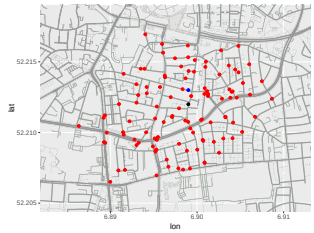
To verify the stability of the solution in Chapter 5.7.3, we use the estimated parameter vector $\hat{\boldsymbol{\theta}}$ to calculate the matrices $H(\hat{\boldsymbol{\theta}})$ and $V(\hat{\boldsymbol{\theta}})$. To do this, Algorithm 4.10 is used to obtain the estimators $\hat{H}(\hat{\boldsymbol{\theta}})$ and $\hat{V}(\hat{\boldsymbol{\theta}})$. We generate 200 realisations of U from the full model using coupling from the past, and for each sample pattern $\mathbf{u}$ we generate 500 realisations of the conditional process using the Metropolis-Hastings sampler. The same dummy realisation $\mathbf{d}$ that was used for the parameter estimation step is also used to evaluate the corresponding integrals. This gives us an estimator for the Godambe information matrix $\hat{G}(\boldsymbol{\theta})$. The elements g_{ij} of $\hat{G}(\boldsymbol{\theta})$, where $i, j = 0, \ldots, 4$ correspond to the indices of the parameter vector $\boldsymbol{\theta}$, can be found in Table 5.5. The low values of the diagonals g_{ii} for $i = 0, \ldots, 3$ imply that the maximum is smooth, meaning that a similar log-likelihood is attained for alternate values of $\boldsymbol{\theta}_i$ in the neighbourhood of $\hat{\boldsymbol{\theta}}_i$. The value of g_{44} is relatively high, which implies that the model is particularly sensitive to the choice of η. This makes sense, since the area calculations have a large effect on transition and/or birth-death probabilities in the simulation algorithms.

5.7.5 Supplementary diagrams

We now show an example of the distribution of occurrence times within a concrete time interval. The interval we consider in this section refers to a car arson fire that was reported to the police and fire department in early August of 2018, 4 days, 10 hours and 50 minutes after the car had last been observed. It occurred in a residential area located to the south of the Enschede city centre. In a similar vein to the simulated example in Chapter 4.6, we show the effect of the prior by plotting a histogram of Metropolis-Hastings samples in a time interval. We re-run the Metropolis-Hastings algorithm 1,000,000 times, this time with the correctly estimated parameter vector $\boldsymbol{\theta}$, throwing out the first 10,000 due to burn-in. The histogram of estimated occurrence times can be found in Figure 5.12. One would expect random behaviour in time for a value of η close to 0, that being precisely



(a) Histogram for specific interval in car arson fire data set.



(b) A view of the interaction area (red) around a car arson fire.

Figure 5.12 Supplementary figures for car arson fire analysis. In **(b)**, the black point signifies the location of the car arson fire. The blue point is an interacting fire close in space and time.

what we see in Figure 5.12a. Note that temporal covariates will still affect the distribution of occurrence times: the increased incidence after 00:30 on the 5th of August is due to that day being a Sunday. Figure 5.12b shows the location of two fires: the desired car arson fire introduced earlier in this subsection in red and a fire occurring on the 25th of August, 2018. This fire is close enough to the red car arson fire in time for their interaction areas to intersect. Since the point process is near-Poisson in time, one would expect no interaction between the two points. This is confirmed in Figure 5.12a, since there is no visible increase or decrease the closer in time one gets to the black car arson fire within the interval of the red car arson fire.

With all of this information, we can conclude that there seem to be little or no interactions between car arson fires. We do however see that certain covariates affect the distribution of occurrence times of the car arson fires within their observed intervals.

5.7.6 Model validation

To make sure that the results obtained in this section are plausible, we run a number of simulations and calculate a network variant of the scaled K-function [122]. For a point process $X = \{(z_1, s_1), \ldots, (z_n, s_n)\}$ defined on $L \times W_T$, the K-function is given by

$$K((r,h) \mid X) = \sum_{i=1}^{n} \frac{\mathbb{E}N(L_{r,h}(z_i, s_i))}{\lambda_{\hat{\theta}}((z_i, s_i) \mid X)}$$

where d and r are the irregular parameters used for the interaction area, $N(A)$ is the number of points in the set A and $L_{r,h}(\cdot), \lambda_{\hat{\theta}}$ are as before. We estimate

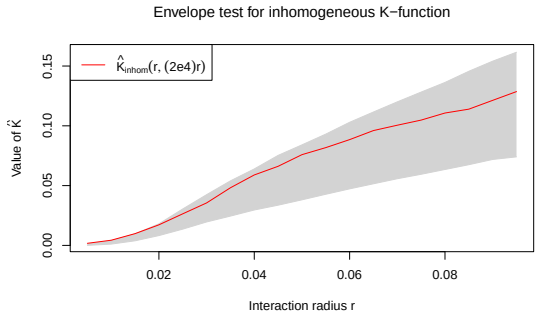


Figure 5.13 A plot of the inhomogeneous K-function estimator $\hat{K}((r,h)\,|\,\mathbf{x})$ against the interaction radius r, together with the min-max envelopes based on 99 simulations from the model under $\hat{\boldsymbol{\theta}}$.

this property by calculating

$$\hat{K}((r,h)\,|\,X) = \sum_{i=1}^{n}\sum_{j\neq i}\frac{\mathbf{1}\{(z_j,s_j)\in L_{r,h}(z_i,s_i)\}}{\lambda_{\hat{\theta}}((z_i,s_i)\,|\,X)\lambda_{\hat{\theta}}((z_j,s_j)\,|\,X)}.$$

We generate 99 simulations $\{\mathbf{y}_1,\ldots,\mathbf{y}_{99}\}$ of the model under $\hat{\boldsymbol{\theta}}$ using coupling from the past. We subsequently calculate $\hat{K}(d\,|\,\mathbf{y})$ for each simulation, where $d = (r,(2e4)r)$, for $r = 0.001, 0.002, \ldots, 0.099, 0.1$. For each value of r, we compute the 95% quantile interval. We then calculate $\hat{K}(d\,|\,\mathbf{x})$ for the same values, where $\mathbf{x}$ is the data pattern. This is then compared to the 95% quantile interval taken from the simulations. As can be seen in Figure 5.13, the model fit is generally very good, especially for larger values of r and h.

5.8 Discussion

Proactive policing methods play a role in reducing crime when applied effectively [107, 125]. Based on sociological and criminological theories such as routine activity theory [28] and temporal factors of community structure [20, 62], proactive policing methods [80, 132] have been developed to improve police effectiveness, strengthening institutions. To facilitate this, predictive policing methods, especially as they relate to hot spots policing, have gained significant traction [107]. With the underlying criminological theory in mind, the models introduced in this thesis are able to address crucial questions and effectively model situations often encountered in criminology.

Aoristic crime data have been modelled in numerous different ways over the years, with approaches ranging from easily calculable summary statistics methods to sampling from probability mass functions. The models introduced in this thesis

model the burglar and the victim separately, do not assume that crimes are equally likely to occur at all times, and takes into account that victims might be away from their places of residence for different lengths of time depending on the time at which they leave and allows for data sets to be interconnected, as crimes taking place in the same neighbourhood almost certainly are. Models are applied to a data set of residential burglaries occurring in Washington D.C. in February 2016, for which a table of most likely burglary times has been provided, along with a visualisation in the form of a graphic showing intervals and estimated times. In addition, a full statistical application on a set of 142 car arson fires in the city of Enschede in the Netherlands shows that there is no evidence to suggest clustered or regular behaviour. To do this, covariate analysis and model selection were performed to posit a plausible prior form. Parameter estimation methods for the network area-interaction point process with partially observed realisations, developed in Chapter 4, were applied. Stability was estimated by way of the Godambe information matrix. Application to the data set resulted in stable estimates, the conclusion being that there are little or no interactions between car arson fires.

CHAPTER 6

Conclusion

6.1 Brief summary of research

In this thesis, we developed numerous approaches to facilitate the modelling of interval-censored data within a Bayesian and point process framework. The goal of this modelling approach was to develop robust models that take prior information and knowledge about the distribution of intervals into account when estimating the exact temporal location of an event within its observed time interval. To facilitate this, an approach combining two statistical models was developed. A "ground truth" point process model serves as a naive approximation of the problem. This is then augmented by a model for the censoring mechanism, assigning higher likelihood to certain regions within intervals depending on the overall distribution of intervals. We allowed for points to be either fully observed, meaning that they lie in an interval of length 0, or partially observed. In the latter case, they lie within an interval of non-zero length.

In Chapter 2, we focused mainly on the censoring mechanism, deriving equations and verifying measure-theoretic aspects to ensure that both the point process model and the censoring mechanism are well-defined. We also introduced the full point process model in a Bayesian context. These steps allow for a base model to be defined, with a number of assumptions that are subsequently loosened in later chapters to allow for more accurate and flexible modelling. Finally, we showed how Markov chain Monte Carlo methods can be used to sample from this model. To demonstrate this, we provided several toy examples to show how the choice of point process prior affects the distribution of occurrence times.

In Chapter 3, we considered a non-homogeneous interval censoring mechanism based on semi-Markov theory and derive distributions for time intervals under this assumption. We provided a number of simulations from the full point process model with the updated censoring mechanism, which show the added power and flexibility that this statistical model is able to provide over the base model introduced in Chapter 2.

In Chapter 4, we provided an extension of the statistical models developed in earlier chapters to spatio-temporal data. Specifically, to allow for seamless interpretation in criminology, we considered Markov point process on Euclidean graphs, which are often used to model street networks. We showed that such an extension is feasible and well-defined within the framework of Markov point processes with our chosen distance metric. Since these extensions are to the "ground truth" point process, we subsequently outlined simulation methods for both the prior point process and the full model, encapsulating the temporal censoring mechanisms developed in Chapters 2 and 3. In addition, we developed parameter estimation methods for the hidden data case when the intensity of the prior point process is affected by covariate functions, as well as adapting methods to verify the stability of solutions to estimation equations for this and similar

problems.

In Chapter 5, we looked at possible applications for the statistical models developed in this thesis in the field of criminology. We showed how the specifics of policing methods and findings from qualitative research make the models developed in this thesis suitable for application to residential burglary and car arson fire data. We compared previous approaches with our approaches and showed how our models are able to take factors such as socioeconomic status and interactions between crimes into account. We applied the model to a number of datasets, with a full statistical application being performed on a set of car arson fires in the city of Enschede in the Netherlands. We found that there is no reason to suggest that there is any interaction between car arson fires.

6.2 Novel contributions and findings

This thesis provides a number of significant novel contributions to the fields of point process theory, applied probability, statistical methodology and applied criminology. Additionally, the full statistical application of the most complex model, introduced in Chapter 4, shows that there is not enough evidence to suggest that car arson fires display significant interaction. Contributions and findings from this thesis can be utilised and expanded in the aforementioned fields in the future.

To our knowledge, no model has used point process methods to develop an inference scheme for the estimation of points within interval-censored data. While state estimation methods have been developed for similar models such as in [88], this thesis and the related research articles combine these techniques with a mechanism by which the uncertainty that creates the hidden data can be modelled, in the form of a stochastic model for the interval censoring mechanism. This technique has the potential to be very powerful for interval-censored data. While the censoring mechanisms themselves are based on existing stochastic processes, they are applied in novel ways in this thesis, using distributions of age and excess to derive marginal and conditional distributions of occurrence within intervals. Chapters 2 and 3 outline the theory behind these models and show their effectiveness through simulated examples.

The next main novel contribution comes in Chapter 4. As far as we are aware, there is no literature on inference for Markov point processes on Euclidean graphs, with or without hidden data. This constitutes a valuable contribution to point process theory. The extension of the standard area-interaction point process to Euclidean graphs, performing all of the theoretical groundwork to ensure that existence conditions are met, constitutes a novel contribution. In Chapter 2.7, we introduce a parameter estimation method based on estimating equations for

hidden data. This is an extension of commonly used methods in point process theory to partially observed point processes. This is also a novel contribution to the statistical literature, allowing estimators to be calculated for uncertain states.

The application to car arson fire data is unique, thorough and easily interpretable. This application is the culmination of several research articles outlining novel theoretical approaches in point process theory. By utilising the full spatio-temporal model, we find an absence of spatial and temporal interaction between car arson fire events. Using this information, we can conclude that arson in the city of Enschede is likely a spontaneous and opportunistic crime that varies depending on covariate information relating to socioeconomic factors. However, the presence or lack of fires in a certain area does not have any effect on other fires within that same area. This information can potentially assist law enforcement and fire services with patrolling and prevention measures.

6.3 Potential future research directions

The contributions and findings contained in this thesis can be taken in a number of directions. An outline of these future research directions is provided below.

One potential extension of the work performed in this thesis is to develop an asymptotic regime for measuring standard errors for the estimating equation approach to parameter estimation in Chapter 4. In our application, we used the Godambe information matrix to assess the stability of the score equation and hence also the parameter estimates. However, this approach could be made more mathematically rigorous. Due to the lack of stationarity in networks, the approach that is most likely to be successful is to look at asymptotic regimes in time. This would potentially lead to a central limit theorem for the network estimating equation.

Additionally, there may be fields where interval-censored data of the type outlined in Chapter 1 are also prevalent. In this case, they would be amenable to statistical analysis as outlined in Chapter 5. One possible example is estimating the most likely time of death of patients when they are found dead by hospital workers, police officers or members of the public, after being reported missing at a certain time. Naturally, different phenomena are also affected by different underlying processes, meaning that the model structure may need to be adapted somewhat to most accurately model the behaviour in question. However, the models discussed in this thesis do provide significant flexibility for modelling interval-censored data in general.

In the process of researching these topics, a significant amount of software was

written in `C++, R` and `Python` for the purpose of data extraction and cleaning, statistical calculation, simulation, estimation and creating interpretable graphics. In the future, this software could be combined into one all-encompassing `R` package for point process analysis for interval-censored data. While significant progress has been made towards this goal, it has not been realised yet. Publishing a package of this nature would allow for increased reproducibility, bug-fixing and conform to open science practices.

[1] Ambler, G. *Dominated coupling from the past and some extensions of the area-interaction process*. PhD thesis, University of Bristol, 2002.

[2] Ambler, G., and Silverman, B. Perfect simulation using dominated coupling from the past with application to area-interaction point processes and wavelet thresholding. In *Probability and mathematical genetics*, N. H. Bingham and C. Goldie, Eds. Cambridge University Press, 2010.

[3] Anderes, E., Møller, J., and Rasmussen, J. G. Isotropic covariance functions on graphs and their edges. *The Annals of Statistics 48*, 4 (2020), 2478–2503.

[4] Ang, Q., Baddeley, A., and Nair, G. Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics 39*, 4 (2012), 591–617.

[5] Ashby, M., and Bowers, K. A comparison of methods for temporal analysis of aoristic crime. *Crime Science 2* (2013).

[6] Asmussen, S. *Applied probability and queues*. Springer, 2003.

[7] Baddeley, A., Coeurjolly, J.-F., Rubak, E., and Waagepetersen, R. A logistic regression estimating function for spatial Gibbs point processes. *CSGB Research Reports 2*, 1 (2013).

[8] Baddeley, A., Coeurjolly, J.-F., Rubak, E. G. E., and Waagepetersen, R. Logistic regression for spatial Gibbs point processes. *Biometrika 101*, 2 (2014), 377–392.

[9] Baddeley, A., and Lieshout, M. N. M. v. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics 47* (1995), 601–619.

[10] Baddeley, A., and Lieshout, M. N. M. v. Extrapolating and interpolating spatial patterns. *In Spatial Cluster Modelling* (2002).

[11] Baddeley, A., and Møller, J. Nearest-Neighbour Markov point processes and random sets. *International Statistical Review / Revue Internationale de Statistique 57*, 2 (1989), 89–121.

[12] Baddeley, A., Møller, J., and Pakes, A. G. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics 60*, 3 (2008), 627–649.

[13] Baddeley, A., Nair, G., Rakshit, S., and McSwiggan, G. "Stationary" point processes are uncommon on linear networks. *Stat 6*, 1 (2017), 68–78.

[14] Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., and Davies, T. M. Analysing point patterns on networks — A review. *Spatial Statistics 42* (2021), 100435.

[15] Baddeley, A., and Turner, R. Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics 42*, 3 (2000), 283–322.

[16] Baddeley, A., Turner, R., and Rubak, E. Adjusted composite likelihood ratio test for spatial Gibbs point processes. *Journal of Statistical Computation and Simulation 86*, 5 (2016), 922–941.

[17] Bernasco, W. Burglary. In *Oxford handbook of crime and public policy*, M. Tonry, Ed. Oxford University Press, 2009, pp. 165–190.

[18] Besag, J. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician) 24*, 3 (1975), 179–195.

[19] Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems 65* (2017), 126–139.

[20] Brantingham, P., and Brantingham, P. Criminality of place: crime generators and crime attractors. *European Journal on Criminal Policy and Research 13* (1995), 5–26.

[21] Breiman, L. Random forests. *Machine Learning 45*, 1 (2001), 5–32.

[22] Briz-Redón, Á. A Bayesian aoristic logistic regression to model spatio-temporal crime risk under the presence of interval-censored event times. *Journal of Quantitative Criminology 40*, 3 (2024), 621–644.

[23] Brooks, S., Gelman, A., Jones, G., and Meng, X. L. *Handbook of Markov chain Monte Carlo*. CRC Press, 2011.

[24] Bursik Jr, R. J. Social disorganization and theories of crime and delinquency: Problems and prospects. *Criminology 26*, 4 (1988), 519–552.

[25] Çinlar, E. Markov Renewal Theory. *Advances in Applied Probability 1*, 2 (1969), 123–187.

[26] Coeurjolly, J.-F., Guan, Y., Khanmohammadi, M., and Waagepetersen, R. Towards optimal Takacs–Fiksel estimation. *Spatial Statistics 18* (2016), 396–411.

[27] Coeurjolly, J.-F., and Rubak, E. Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scandinavian Journal of Statistics 40*, 4 (2013), 669–684.

[28] Cohen, L. E., and Felson, M. Social change and crime rate trends: a routine activity approach. *American Sociological Review 44*, 4 (1979), 588–608.

[29] Cohn, E. G., and Rotton, J. Even criminals take a holiday: Instrumental and expressive crimes on major and minor holidays. *Journal of Criminal Justice 31*, 4 (2003), 351–360.

[30] Cox, D. R. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological) 17*, 2 (1955), 129–157.

[31] Cox, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological) 34*, 2 (1972), 187–220.

[32] Crema, E. R. A Bayesian alternative for aoristic analyses in archaeology. *Archaeometry 61*, S1 (2024), 7–30.

[33] Cressie, N., Frey, J., Harch, B., and Smith, M. Spatial prediction on a river network. *Journal of Agricultural, Biological, and Environmental Statistics 11*, 2 (2006), 127–150.

[34] Daley, D., and Vere-Jones, D. *An introduction to the theory of point processes. Vol. I: Elementary theory and methods. 2nd ed*, vol. Vol. 1. Springer, 2003.

[35] Debeer, D., and Strobl, C. Conditional permutation importance revisited. *BMC Bioinformatics 21*, 1 (2020), 307.

[36] Derek Tucker, J., Shand, L., and Lewis, J. R. Handling missing data in self-exciting point process models. *Spatial Statistics 29* (2019), 160–176.

[37] Diggle, P. A kernel method for smoothing point process data. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 34*, 2 (1985), 138–147.

[38] Diggle, P., Rowlingson, B., and Su, T.-l. Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics 16*, 5 (2005), 423–434.

[39] Diggle, P. J. *Statistical analysis of spatial point patterns*. Arnold, 2003.

[40] Diggle, P. J. Spatio-temporal point processes: methods and applications. *Monographs on Statistics and Applied Probability 107* (2006), 1.

[41] Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D., and Tanemura, M. On Parameter Estimation for Pairwise Interaction Point Processes. *International Statistical Review / Revue Internationale de Statistique 62*, 1 (1994), 99–117.

[42] Diggle, P. J., Moraga, P., Rowlingson, B., and Taylor, B. M. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statistical Science 28*, 4 (2013), 542–563.

[43] Eck, J., and Weisburd, D. Crime places in crime theory. *Crime and Place, Crime Prevention Studies 4* (1995).

[44] Eckardt, M., and Moradi, M. Marked spatial point processes: current State and extensions to point processes on linear networks. *Journal of Agricultural, Biological and Environmental Statistics 29*, 2 (2024), 346–378.

[45] Felson, M. *Crime and everyday life*. Pine Forge Press Thousand Oaks, California, 1998.

[46] Felson, M., and Cohen, L. E. Human ecology and crime: a routine activity approach. *Human Ecology 8*, 4 (1980), 389–406.

[47] Felson, M., and European Institute for Crime Prevention and Control, a. w. t. U. N. *The ecosystem for organized crime*. The European Institute for Crime Prevention and Control, affiliated with the United Nations Helsinki, 2006.

[48] Felson, M., and Poulsen, E. Simple indicators of crime by time of day. *International Journal of Forecasting 19*, 4 (2003), 595–601.

[49] Gaines, L. K., Miller, R. L., and Bassi, L. *Criminal justice in action*. Chicago University Press, 2003.

[50] Gamerman, D., and Lopes, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

[51] Gau, J. M., and Brunson, R. K. Procedural justice and order maintenance policing: a study of inner-city young men's perceptions of police legitimacy. *Justice Quarterly 27*, 2 (2010), 255–279.

[52] Georgii, H.-O. Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics 48*, 1 (1976), 31–51.

[53] Geyer, C. J. Likelihood inference for spatial point processes. In *Stochastic geometry, likelihood, and computation*, O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. v. Lieshout, Eds. CRC Press, 1999, pp. 141–172.

[54] Geyer, C. J., and Møller, J. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian journal of statistics* (1994), 359–373.

[55] Geyer, C. J., and Thompson, E. A. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological) 54*, 3 (1992), 657–699.

[56] Godambe, V. P., and Heyde, C. C. Quasi-likelihood and optimal estimation, correspondent paper. *International Statistical Review / Revue Internationale de Statistique 55*, 3 (1987), 231–244.

[57] González, J. A., Rodríguez-Cortés, F. J., Cronie, O., and Mateu, J. Spatio-temporal point process statistics: A review. *Spatial Statistics 18* (2016), 505–544.

[58] Gregori, P., M, L., and Mateu, J. Modelización de procesos area-interacción generalizados. *Electronic Notes in Theoretical Computer Science - ENTCS* (2003).

[59] Haezendonck, J., and De Vylder, F. A comparison criterion for explosions in point processes. *Journal of Applied Probability 17*, 4 (1980), 1102–1107.

[60] Harcourt, B. E., and Ludwig, J. Broken windows: new evidence from New York City and a five-city social experiment : Vol. 73: Iss. 1, Article 14. *University of Chicago Law Review 73*, 1 (2006), 271–350.

[61] Hawkes, A. G., and Oakes, D. A cluster process representation of a self-exciting process. *Journal of applied probability 11*, 3 (1974), 493–503.

[62] Hawley, A. H. *Human ecology - a theory of community structure*, vol. 26. The Ronald Press Company, 1950.

[63] Heitgerd, J. L., and Bursik, R. J. Extracommunity dynamics and the ecology of delinquency. *American Journal of Sociology 92*, 4 (1987), 775–787.

[64] Helms, D. Temporal analysis. In *Exploring Crime Analysis: readings on essential skills, second edition*, S. L. Gwinn, C. Bruce, J. Cooper, and S. Hick, Eds. International Association of Crime Analysts, Overland Park, 2008, pp. 214–257.

[65] Hoef, J. M. V., Peterson, E., and Theobald, D. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics 13*, 4 (2006), 449–464.

[66] Hothorn, T., Hornik, K., and Zeileis, A. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics 15*, 3 (2006), 651–674.

[67] Huang, J. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics 24*, 2 (1996), 540–568.

[68] Huang, J., and Wellner, J. A. Interval censored survival data: A review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics* (New York, NY, 1997), D. Y. Lin and T. R. Fleming, Eds., Springer US, pp. 123–169.

[69] Huber, M. L. *Perfect simulation*. Apple Academic Press Inc., SE - 250 sidor ; 16.4 cm, 2015.

[70] Iftimi, A., Lieshout, M. N. M. v., and Montes, F. A multi-scale area-interaction model for spatio-temporal point patterns. *Spatial Statistics 26* (2018), 38–55.

[71] Iftimi, A., van Lieshout, M. N. M., and Montes, F. A multi-scale area-interaction model for spatio-temporal point patterns, 2017.

[72] Janssen, J. *Semi-Markov models: theory and applications*. Springer US, 2013.

[73] Janssen, J., and Manca, R. *Applied semi-Markov processes*. Springer, 2006.

[74] Jensen, E. B. V., and Nielsen, L. S. A review on inhomogeneous Markov point processes. *Lecture Notes-Monograph Series 37* (2001), 297–318.

[75] Jensen, J. L., and Moller, J. Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability 1*, 3 (1991), 445–461.

[76] Kaplan, E. L., and Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association 53*, 282 (1958), 457–481.

[77] Karr, A. *Point processes and their statistical inference*, vol. 7. CRC press, 1991.

[78] Kendall, W. S. Perfect simulation for the area-interaction point process. *Probability Towards 2000 128* (1998), 218–234.

[79] Kendall, W. S., and Møller, J. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability 32*, 3 (2000), 844–865.

[80] Kennedy, D. M., Piehl, A. M., and Braga, A. A. Youth violence in Boston: gun markets, serious youth offenders, and a use-reduction strategy. *Law and Contemporary Problems 59*, 1 (1996), 147.

[81] Korolyuk, V., and Swishchuk, A. *Semi-Markov Random Evolutions*. Springer Netherlands, Dordrecht, 1995, pp. 59–91.

[82] Kubrin, C. E., and Weitzer, R. New directions in social disorganization theory. *Journal of Research in Crime and Delinquency 40*, 4 (2003), 374–402.

[83] Lamb, D. S., Downs, J. A., and Lee, C. The network K-function in context: examining the effects of network structure on the network K-function. *Transactions in GIS 20*, 3 (2016), 448–460.

[84] Leeuwen, N. v., and Venema, J. Statistische gegevens per vierkant en postcode 2022-2021-2020-2019. Tech. rep., Centraal Bureau van de Statistiek, 2023.

[85] Lieshout, M. N. M. v. *Stochastic geometry models in image analysis and spatial statistics.*, vol. 253 of *CWI tracts*. Centrum voor Wiskunde en Informatica, 1995.

[86] Lieshout, M. N. M. v. *Markov Point Processes and Their Applications*. World Scientific, 2000.

[87] Lieshout, M. N. M. v. On estimation of the intensity function of a point process. *Methodology and Computing in Applied Probability 14*, 3 (2012), 567–578.

[88] Lieshout, M. N. M. v. *State estimation for temporal point processes*. No. 2048 in Memorandum of the Department of Applied Mathematics. University of Twente, Department of Applied Mathematics, 2015.

[89] Lieshout, M. N. M. v. Nearest-neighbour Markov point processes on graphs with Euclidean edges. *Advances in Applied Probability 50*, 4 (2018), 1275–1293.

[90] Lieshout, M. N. M. v., and Markwitz, R. L. State estimation for aoristic models. *Scandinavian Journal of Statistics 50*, 3 (2023), 1068–1089.

[91] Lieshout, M. N. M. v., and Markwitz, R. L. A non-homogeneous alternating renewal process model for interval censoring. *Journal of Applied Probability 62*, 2 (2024), 494–515.

[92] Lieshout, M. N. M. v., and Stoica, R. S. Exact Metropolis-Hastings sampling for marked point processes using a C++ library. *PNA-R : probability, networks and algorithms* (2004), 22.

[93] Lu, C., Lieshout, M. N. M. v., de Graaf, M., and Visscher, P. Data-driven chimney fire risk prediction using machine learning and point process tools. *The Annals of Applied Statistics 17*, 4 (2023), 3088–3111.

[94] Macchi, O. The coincidence approach to stochastic point processes. *Advances in Applied Probability 7*, 1 (1975), 83–122.

[95] Makkink, G. *Limitations and perspectives of lysimeter research.* No. no. 48, 49 in Publication / Association Internationale d'Hydrologie Scientifique. AIHS, 1959, pp. 13–25.

[96] Markwitz, R. L. A likelihood-based approach to developing effective proactive police methods. In *Sustainable development goals of the UN, safety, security and evidence based response to crime and other security threats*, G. Meško, S. Kutnjak Ivkovich, and R. Hacin, Eds. Maribor University Press, 2024.

[97] Marsan, D., and Lengliné, O. Extending earthquakes' reach through cascading. *Science 319*, 5866 (2008), 1076–1079.

[98] McSwiggan, G., Baddeley, A., and Nair, G. Kernel density estimation on a linear network. *Scandinavian Journal of Statistics 44*, 2 (2017), 324–345.

[99] Mengersen, K. L., and Tweedie, R. L. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics 24*, 1 (1996), 101–121.

[100] Meyn, S., and Tweedie, R. *Markov chains and stochastic stability.* Springer, 2009.

[101] Møller, J., Syversveen, A. R., and Waagepetersen, R. P. Log Gaussian Cox processes. *Scandinavian Journal of Statistics 25*, 3 (1998), 451–482.

[102] Møller, J., and Waagepetersen, R. *Statistical inference and simulation for spatial point processes*, vol. 100. CRC Press, 2004.

[103] Møller, J., and Waagepetersen, R. Some recent developments in statistics for spatial point patterns. *Annual Review of Statistics and Its Application 4* (2017), 317–342.

[104] Monteiro, A., Smirnov, G. V., and Lucas, A. Nonparametric estimation for non-homogeneous semi-Markov processes: an application to credit risk. Tinbergen Institute Discussion Papers 06-024/2, Tinbergen Institute, 2006.

[105] Moradi, M. M., Rodríguez-Cortés, F. J., and Mateu, J. On Kernel-Based Intensity Estimation of Spatial Point Patterns on Linear Networks. *Journal of Computational and Graphical Statistics 27*, 2 (2018), 302–311.

[106] Nasar, J. L., and Fisher, B. 'Hot spots' of fear and crime: A multi-method investigation. *Journal of Environmental Psychology 13*, 3 (1993), 187–206.

[107] National Academies of Sciences Engineering and Medicine. *Proactive policing: effects on crime and communities*. The National Academies Press, Washington, DC, 2018.

[108] Neyman, J., and Scott, E. L. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological) 20*, 1 (1958), 1–43.

[109] Nguyen, X. X., and Zessin, H. Ergodic theorems for spatial processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 48*, 2 (1979), 133–158.

[110] Ogata, Y., and Katsura, K. Likelihood analysis of spatial inhomogeneity for marked point patterns. *Annals of the Institute of Statistical Mathematics 40*, 1 (1988), 29–39.

[111] Okabe, A., and Sugihara, K. *Spatial analysis along networks : statistical and computational methods*. Statistics in practice. Wiley, Chichester, West Sussex, 2012.

[112] Okabe, A., and Yamada, I. The K-Function method on a network and its computational implementation. *Geographical Analysis 33*, 3 (2001), 271–290.

[113] Open Data DC. Crime Incidents in 2016, 2016.

[114] Papangelou, F. The conditional intensity of general point processes and an application to line processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 28* (1974), 207–226.

[115] Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., and Hollywood, J. S. *Predictive policing: The role of crime forecasting in law enforcement operations*. RAND Corporation, 2013.

[116] Porcu, E., White, P. A., and Genton, M. G. Stationary nonseparable space-time covariance functions on networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*, 5 (2023), 1417–1440.

[117] Preston, C. Spatial birth and death processes. *Advances in Applied Probability 7*, 3 (1975), 465–466.

[118] Propp, J. G., and Wilson, D. B. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Algorithms 9*, 1–2 (1996), 223–252.

[119] Ratcliffe, J. Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology 18* (2002), 23–43.

[120] Ratcliffe, J., and McCullagh, M. Aoristic crime analysis. *International Journal of Geographical Information Science 12* (1998), 751–764.

[121] Reiss, A. J. Police organization in the twentieth century. *Crime and Justice 15* (1992), 51–97.

[122] Ripley, B. D. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological) 39*, 2 (1977), 172–212.

[123] Ripley, B. D., and Kelly, F. P. Markov point processes. *Journal of the London Mathematical Society s2-15*, 1 (1977), 188–192.

[124] Roberts, G. O., and Smith, A. F. M. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications 49* (1994), 207–216.

[125] Rosenbaum, D. P. The limits of hot spots policing. In *Police Innovation: Contrasting Perspectives*, D. Weisburd and A. A. Braga, Eds., Cambridge Studies in Criminology. Cambridge University Press, 2006, pp. 245–264.

[126] Ross, S. M. *Stochastic processes*. Wiley series in probability and statistics: Probability and statistics. Wiley, 1996.

[127] Sampson, R. J., and Groves, W. B. Community structure and crime: testing social-disorganization theory. *American Journal of Sociology 94*, 4 (1989), 774–802.

[128] Santos, R. B. The effectiveness of crime analysis for crime reduction: cure or diagnosis? *Journal of Contemporary Criminal Justice 30*, 2 (2014), 147–168.

[129] Sharma, S. Markov chain Monte Carlo methods for Bayesian data analysis in astronomy. *Annual Review of Astronomy and Astrophysics 55*, 1 (2017), 213–259.

[130] Shaw, C. R., and McKay, H. D. *Juvenile delinquency and urban areas.* University of Chicago Press, 1942.

[131] Sherman, L., and Eck, J. Policing for crime prevention. *Evidence-based crime prevention* (2002), 295–329.

[132] Sherman, L., and Weisburd, D. General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly 12* (1995), 625–648.

[133] Short, M. B., D'Orsogna, M. R., Brantingham, P. J., and Tita, G. E. Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology 25*, 3 (2009), 325–339.

[134] Soetaert, K., and Herman, P. M. *A practical guide to ecological modelling. Using R as a simulation platform.* Springer, 2009.

[135] Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. Conditional variable importance for random forests. *BMC Bioinformatics 9*, 1 (2008), 307.

[136] Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution." bmc bioinformatics, 8(1), 25. *BMC bioinformatics 8* (2007), 25.

[137] Strobl, C., Hothorn, T., and Zeileis, A. Party on! *The R Journal 1*, 2 (2009), 14–17.

[138] Takacs, R. Estimator for the pair–potential of a gibbsian point process. *Statistics 17*, 3 (1986), 429–433.

[139] Takacs, R., and Fiksel, T. Interaction pair-potentials for a system of ant's nests. *Biometrical Journal 28*, 8 (1986), 1007–1013.

[140] Tang, J., and Zimmerman, D. Space-time covariance models on networks. *Electronic Journal of Statistics 18* (2024).

[141] Waagepetersen, R. Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika 95*, 2 (2008), 351–363.

[142] Wilcox, S. *The geography of robbery; the prevention and control of robbery.* Center of Administration of Justice, University of California, 1973.

[143] Wilson, J. Q., and Kelling, G. Broken windows: The police and neighborhood safety, 1982.

[144] Zeng, D., Mao, L., and Lin, D. Y. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika 103*, 2 (2016), 253–271.

Summary

In this thesis, we develop a number of statistical frameworks within which interval-censored data can be modelled in a point process context. By interval-censored data, we refer to data in which events can be partially observed in the form of intervals, instead of being fully observed as simply a point in time and space. As of this date, the research completed within the scope of this PhD thesis has amounted to three accepted research papers in point process theory, applied probability and applied criminology, with a fourth article in the works in the field of spatial statistics.

In Chapter 2, we focus entirely on the temporal aspect of the interval-censored data. The base statistical model consists of two separate stochastic processes. One is responsible for the interval censoring mechanism, whereas the other process models the behaviour of the underlying stochastic process responsible for the event times. We assume that the censoring mechanism generates intervals via an alternating renewal process and the underlying stochastic process takes the form of a Markov point process. In this way, we construct a marked point process model, with the marks being the intervals within which a given point is contained. The measure-theoretic foundations and model-building concepts introduced in this chapter form the base layer and allow for extensions to be added in later chapters. We sample from this model by using a Bayesian framework, where the point process is the prior, the censoring mechanism is the likelihood and the full model is the posterior.

A shortcoming of this model is that the renewal process censoring mechanism leads to an implicit homogeneity assumption. Interval starting points may be more frequent in different parts of an observation window based on external factors. Similarly, the length of an interval may also be dependent on where it is observed. Therefore, we develop an interval censoring scheme based on semi-Markov theory, assuming that intervals are generated non-homogeneously in time. A modelling framework for this censoring mechanism is also developed, showing that functions of certain forms lead to the presence of semi-Markovian behaviour. Simulated comparisons between homogeneous and non-homogeneous censoring mechanisms show that the updated model is able to more accurately account for differences in interval starting point frequency and interval lengths.

In previous chapters, we assume that censoring takes place only along the temporal axis and that the exact location of a point in space is known. However, the spatial location of a point may affect the prior and posterior likelihood due to the presence of covariates. In many cases, the geometry of the spatial component of such data is quite complex. Therefore, we have proposed a novel form of the point process describing the underlying stochastic process of occurrence times by extending it to take values on Euclidean graphs. We retain the Bayesian

framework used throughout the research project. We develop a posterior sampling method to sample from this complex spatio-temporal model and introduce robust parameter estimation methods based on estimating equations to derive estimators for statistical quantities. We test this approach on simulated data.

Chapter 5 provides a link between the mathematics developed thus far in the project and an actual use case. Specifically, the developed statistical theory has significant potential to be applied in the field of criminology. Both burglaries and arson fires are often poorly observed by victims and law enforcement alike, leading to data often being partially interval-censored in nature. After providing background information on criminological theories, outlining previous methods used in the field and describing the statistical models developed in this thesis in a criminological context, the updated model was applied to a residential burglary and car arson fire data set, in the latter case as part of a full statistical application. We find that car arson fires in the city of Enschede in the Netherlands do not exhibit significant interaction, instead varying in location and time based on other external factors.

Samenvatting

In dit proefschrift ontwikkelen we statistische methodologie voor het modelleren van interval-gecensureerde data binnen een puntprocescontext. Met interval-gecensureerde data wordt data bedoeld waarbij gebeurtenissen deels kunnen worden waargenomen in de vorm van intervallen, in plaats van volledig te worden waargenomen als een punt in tijd en ruimte. Tot op heden heeft het onderzoek waarop dit proefschrift gebaseerd is, geresulteerd in drie geaccepteerde wetenschappelijke artikelen op het gebied van puntprocestheorie, toegepaste kansrekening en toegepaste criminologie, met een vierde artikel binnen het vakgebied ruimtelijke statistiek in voorbereiding.

In hoofdstuk 2 richten we ons volledig op het temporele aspect van interval-gecensureerde data. Het basismodel bestaat uit twee verschillende stochastische processen. Het ene process geeft het intervalcensuurmechanisme weer, terwijl het andere proces het gedrag modelleert van het onderliggende stochastische proces dat verantwoordelijk is voor de tijden. We maken doorgaans de aanname dat het censuurmechanisme intervallen genereert via een alternerend vernieuwingsproces (Engels: *alternating renewal process*) en dat het onderliggende stochastische proces de vorm van een Markoviaans puntproces heeft. Op deze manier hebben we een gemarkeerd puntprocesmodel geconstrueerd, waarbij de markeringen tevens de intervallen zijn waarin een bepaald punt valt. De maattheoretische ideeën en modelbouwconcepten die in dit hoofdstuk geïntroduceerd worden, vormen de basislaag en maken het mogelijk om in latere hoofdstukken het model uit te breiden. We nemen steekproeven uit dit model met behulp van een Bayesiaanse aanpak, waarbij het puntproces het a priori process is, het censuurmechanisme de waarschijnlijkheid en het volledige model het a posteriori proces is.

Een tekortkoming van dit model is dat het censuurmechanisme van het vernieuwingsproces homogeneiteit impliceert. In de realiteit is de frequentie waarbij intervallen gegenereerd worden vaak niet uniform over het tijdsbestek dat waargenomen wordt. Op dezelfde manier kan de lengte van een interval ook afhankelijk zijn van waar het waargenomen wordt. Daarom ontwikkelen we een model voor interval-gecensureerde data gebaseerd op de semimarkovtheorie, ervan uitgaande dat intervallen niet-homogeen in de tijd gegenereerd worden. Er wordt ook statistische methodologie voor dit censuurmechanisme ontwikkeld, waaruit blijkt dat bepaalde functievormen als basis voor een semi-Markov model kunnen dienen. Gesimuleerde vergelijkingen tussen homogene en niet-homogene censuurmechanismen tonen aan dat het bijgewerkte model nauwkeuriger rekening kan houden met verschillen in de frequentie van intervalstartpunten en intervallengten.

In voorgaande hoofdstukken hebben we aangenomen dat censuur alleen langs de temporele as plaatsvindt en dat de exacte locatie van een punt in de ruimte er

niet toe doet. De ruimtelijke locatie van een punt kan echter zowel de a priori als de a posteriori kans beïnvloeden door de aanwezigheid van covariaten of features. In veel gevallen is de geometrie van het ruimtelijke component van dergelijke data vrij complex. Daarom hebben we een nieuwe vorm van het puntproces voorgesteld door het onderliggende stochastische proces waarden aan te laten nemen op euclidische grafen. We behouden het Bayesiaanse model dat tot nu toe in het onderzoeksproject gebruikt is. We ontwikkelen methodologie die het mogelijk maakt om steekproeven te nemen uit dit complexe ruimtelijk-temporele model en introduceren robuuste parameterschattingsmethoden op basis van schattingsvergelijkingen (Engels: *estimating equations*) om schatters af te leiden. Deze aanpak wordt vervolgens getoetst op gesimuleerde data.

Hoofdstuk 5 legt een verband tussen de wiskunde die in het project ontwikkeld is, en een concrete toepassing. Daaruit blijkt dat de ontwikkelde statistische theorie aanzienlijk potentieel heeft om toegepast te worden in de criminologie. Zowel inbraken als brandstichtingen worden vaak slecht geobserveerd door slachtoffers, politie en brandweer, waardoor de data deels interval-gecensureerd zijn. We leggen achtergrondinformatie over criminologische theorieën uit, stellen alternatieve methoden voor die eerder in het veld gebruikt zijn, en beschrijven de statistische modellen die in dit proefschrift ontwikkeld zijn binnen een criminologische context. Vervolgens wordt het bijgewerkte model toegepast op datasets van woninginbraken en autobranden, in het laatste geval als onderdeel van een volledige statistische toepassing. We vinden dat autobranden in de stad Enschede geen significante interactie vertonen, maar in plaats daarvan variëren in locatie en tijd op basis van andere externe factoren.

Zusammenfassung

In dieser Arbeit entwickeln wir statistische Methoden zur Modellierung intervallzensierter Daten basiert auf Punktprozessen. Intervallzensierte Daten beziehen sich auf Daten, bei denen Ereignisse teilweise in Form von Intervallen beobachtet werden können, anstatt vollständig als Punkt in Raum und Zeit beobachtet zu werden. Bis jetzt hat das Projekt, auf dem diese Arbeit basiert, zu drei akzeptierten wissenschaftlichen Artikeln in der Punktprozesstheorie, der angewandten Wahrscheinlichkeitstheorie und der angewandten Kriminologie geführt, wobei eine vierte Arbeit im Fachbereich räumlicher Statistik in Vorbereitung ist.

In Kapitel 2 konzentrieren wir uns komplett auf den zeitlichen Aspekt von intervallzensierten Daten. Das Grundmodell besteht aus zwei verschiedenen stochastischen Prozessen. Der eine Prozess modelliert den Mechanismus der Intervallzensierung, während der andere Prozess das Verhalten des zugrunde liegenden stochastischen Prozesses modelliert, der für die Zeiten verantwortlich ist. In der Regel gehen wir davon aus, dass der Zensurmechanismus die Intervalle über einen alternierenden Erneuerungsprozess (englisch: *alternating renewal process*) erzeugt, und dass der zugrunde liegende stochastische Prozess die Form eines markovschen Punktprozesses hat. Auf diese Weise haben wir ein Modell eines markierten Punktprozesses konstruiert, bei dem die Markierungen zugleich die Intervalle sind, in die ein bestimmter Punkt fällt. Die maßtheoretischen Grundlagen und Modellierungskonzepte, die in diesem Kapitel entwickelt werden, bilden die Basisschicht des Modells und ermöglichen uns, in späteren Kapiteln Erweiterungen vorzunehmen. Stichproben werden durch einen bayesschen Ansatz erhoben, bei dem der Punktprozess der a priori Prozess, der Zensierungsmechanismus die Wahrscheinlichkeit und das vollständige Modell der a posteriori Prozess ist.

Ein Nachteil dieses Modelles ist, dass der Zensierungsmechanismus des Erneuerungsprozesses Homogenität voraussetzt. In der Realität ist die Häufigkeit, mit der Intervalle erzeugt werden, oft nicht gleichmäßig über den beobachteten Zeitrahmen verteilt. Ebenso kann die Länge eines Intervalls auch davon abhängen, wo es beobachtet wird. Daher entwickeln wir ein Modell für intervallzensierte Daten, das auf der Semimarkov-Theorie basiert und davon ausgeht, dass Intervalle im Laufe der Zeit inhomogen erzeugt werden. Es werden auch statistische Methoden für diesen Zensierungsmechanismus entwickelt, die zeigen, dass bestimmte Funktionen mit semimarkovschem Verhalten einhergehen. Simulierte Vergleiche zwischen homogenen und inhomogenen Zensurmechanismen zeigen, dass das aktualisierte Modell Unterschiede in der Häufigkeit von Intervallstartpunkten und Intervalllängen genauer berücksichtigen kann.

In den vorigen Kapiteln sind wir davon ausgegangen, dass die Zensierung nur entlang der zeitlichen Achse erfolgt und dass die räumliche Lage eines Punktes keine Rolle spielt. Dies kann jedoch die a priori und a posteriori Wahrschein-

lichkeiten aufgrund des Vorhandenseins von Kovariaten beeinflussen. In vielen Fällen ist die Geometrie der räumlichen Komponente solcher Daten recht komplex. Daher haben wir eine neue Form des Punktprozesses vorgeschlagen, bei der der zugrunde liegende stochastische Prozess Werte auf euklidischen Graphen annimmt. Wir behalten das bayessche Modell bei, das bisher im Forschungsprojekt verwendet wurde. Wir entwickeln Methoden, die es ermöglichen, Stichproben aus diesem komplexen räumlich-zeitlichen Modell zu ziehen, und führen robuste Methoden zur Parameterschätzung ein, die auf Schätzgleichungen (en: *estimating equations*) basieren, um Schätzer abzuleiten. Dieser Ansatz wird dann an simulierten Daten getestet.

Kapitel 5 setzt die bisher im Projekt entwickelte Mathematik in Beziehung zu einer konkreten Anwendung. Es wird gezeigt, dass die entwickelte statistische Theorie erhebliches Anwendungspotenzial in der Kriminologie hat. Sowohl Einbrüche als auch Brandstiftungen werden von Opfern, Polizei und Feuerwehr oft nur unzureichend beobachtet, was zu teilweise intervallzensierten Daten führt. Wir erläutern Hintergrundinformationen zu kriminologischen Theorien, schlagen alternative Methoden vor, die bisher in diesem Bereich verwendet wurden, und beschreiben die in dieser Arbeit entwickelten statistischen Modelle in einem kriminologischen Zusammenhang. Anschließend wenden wir das aktualisierte Modell auf Datensätze zu Wohnungseinbrüchen und Autobränden an, im letzteren Fall als Teil einer vollständigen statistischen Anwendung. Wir stellen fest, dass Autobrände in der Stadt Enschede keine signifikante Interaktion aufweisen, sondern stattdessen in Abhängigkeit von anderen externen Faktoren örtlich und zeitlich variieren.