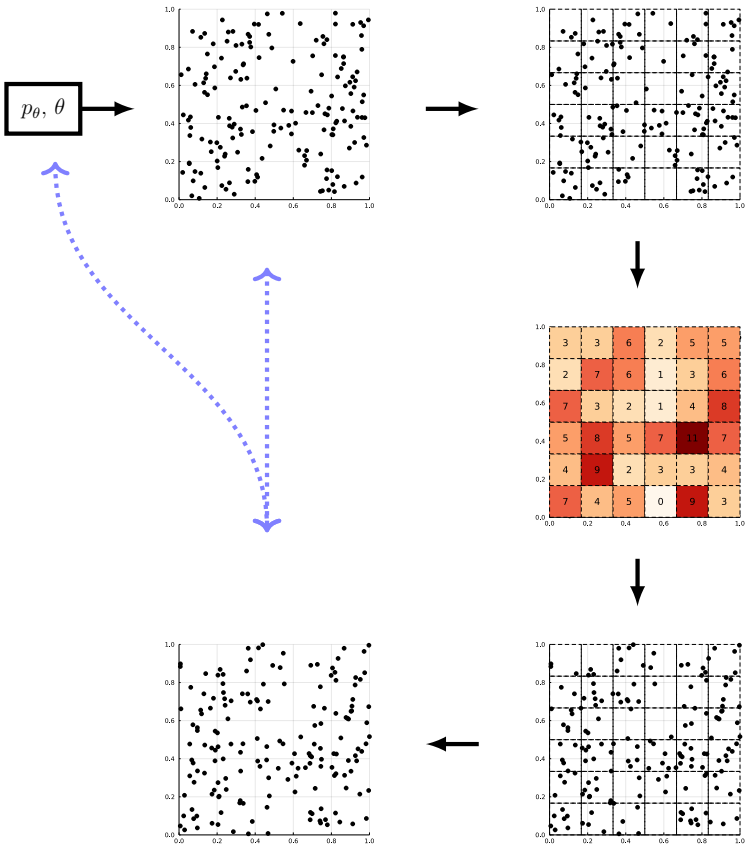
CHALMERS
UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG



Reconstructing point patterns from spatially aggregated data

With an eye towards epidemiological modelling

Master's thesis
Master's Programme in Mathematical Statistics, University of Gothenburg

JENS MICHELSEN

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY & UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024
www.chalmers.se

Master's thesis 2024

Reconstructing point patterns from spatially aggregated data

With an eye towards epidemiological modelling

Jens Michelsen

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematics
Chalmers University of Technology
Gothenburg, Sweden 2024

Reconstructing point patterns from spatially aggregated data
With an eye towards epidemiological modelling
JENS MICHELSEN

Supervisor: Ottmar Cronie, Company or Department
Examiner: Moritz Schauer, Department

Master's Thesis 2024
Department of Mathematics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Illustration of a reproduction of a point configuration.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2024

Reconstructing point patterns from spatially aggregated data
With an eye towards epidemiological modelling
JENS MICHELSEN
Department of Mathematics
Chalmers University of Technology

Abstract

In this thesis we explore the ability to reconstruct samples (point configurations) from a point process, based only on the information contained in spatially aggregated data, namely the number of points in the partitions of a larger region. The ability to reconstruct a point configuration, in such a way that it retains most of it's statistical properties, could be useful in cases where one is faced with a mixed dataset; some regions containing the full point configuration data, while other regions only contain aggregated data, i.e. the counts of subregions. Our main motivation in this thesis however, concerns epidemic modelling, where the locations of individual infections are represented by point(-configuration)s, drawn from a hypothetical point process model, and typically, data is only available in spatially aggregated form.
Here we present a scheme for reconstructing point configurations, as well as a collection of dissimilarity measures to assess the quality of reproduction. These are then analysed (in part) theoretically and verified using simulation studies. We obtain constraints regarding the size of the partitions/subregions in order for the reconstructed point configuration to retain important statistical properties of the original point process.

Keywords: lorem, ipsum, dolor, sit, amet, consectetur, adipisicing, elit, sed, do.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Ottmar Cronie for both inspiring and supporting me throughout this thesis. I would also like to thank my examiner Moritz Schauer for being patient while I slowly finalised the thesis. Finally I would like to thank my friends and family for supporting me and believing in me.

Jens Michelsen, Gothenburg, March 2023

Contents

1
Introduction

Following the outbreak of Covid-19, there has been renewed interest in epidemic modelling and forecasting. Traditional methods focus on modelling the time evolution of spatially aggregated statistics, such as the number of infected, $N_i(t) = N(A_i, t)$, in a certain geographical region, A_i, at time t. The time evolution can be modelled either through deterministic differential or difference equations, such as the SIR-model and various extensions and variations thereof, or through stochastic propagator models, taking into account regional differences as well as infection rates in neighbouring regions (see eg. [1, 2] for recent applications to Covid-19). Inference about model parameters in each region may thus provide insight into regional differences in behaviours and/or rate of spread between regions while also allowing for prediction of regional rates of infection. Additionally, one might imagine also modelling the parameters themselves by using covariates such as population density, data from public transportation, cell phone data from radio masts, etc.

While certainly useful and, to a certain degree, successful at predicting the aggregated rates of infection, ultimately these methods describe the spread of disease (which at its core is a very local process) at a spatially aggregated level, disregarding spatial inhomogeneities of key predictors (such as population density, levels of movement, etc) within the regions themselves. Furthermore, the localized nature of typical methods of infection suggest that infection events should appear in localized clusters, typically at much smaller scales than the extent of typical regions. A superspreader event in close proximity of a border between two regions would lead to a local cluster affecting the aggregate infection rates of both regions, leading to a strong correlation between said regions, even though the aggregate levels of movement between the regions may not be particularly high.

An alternative approach could be to model the epidemic as a spatio-temporal point process, $\boldsymbol{X} = \{(X_j, T_j)\}_{j=1,\dots,N_X}$, with each element of the process (X_j, T_j) representing an infection event, i.e. a person being infected at a random time T_j, and location X_j. Such a localized model could depend on local spatio-temporal predictors based on e.g. data on population density, proximity to major roads or public transportation hubs, local (in space and time) regulatory interventions, etc. Spatio-temporal data on individual infections could provide valuable details about local characteristics of the spread such as clustering, the velocity of growth of such clusters, and could potentially allow one to pinpoint common spatio-temporal pathways of the spread, which could lead to more efficient policies directed at stopping the spread. It could also provide a better representation of a persons individual risk, as opposed to the community level risk of infection.

Unfortunately, such spatio-temporal data sets on individual infection events is excep-

tionally rare (see for example Varicella data [3]), primarily due to privacy concerns; While in principle the Health Departments may have data on time and location (e.g. home adress) of individuals who have received positive Covid tests, privacy regulations would clearly prohibit the use (and certainly publication) of such data in any meaningful way. Thus, more than likely, we are left with data aggregated spatially at a regional level, and temporally at the level of reporting intervals (daily, weekly, etc), i.e. counts of new infections, $N_i(\tau_j) = N(A_i \times \tau_j)$, in region A_i during the time interval $\tau_j = [t_j, t_j + \Delta t]$. Nevertheless, it would seem reasonable that if one could make the regions and reporting intervals (i.e. spatio-temporal intervals $A_i \times \tau_j$) small enough, most of the information contained in the point process could be retained in the aggregated statistics. In this extreme case, one could faithfully reproduce the underlying point process from knowledge of the aggregate statistics. It would also stand to reason that there exists an intermediate regime of sizes for the spatio-temporal intervals where one could approximately reproduce the underlying point process from the aggregate statistics, at least in the sense that it preserves most of its core properties.

It is the aim of this study to examine the conditions under which this is possible. To make this endeavour as manageable and concrete as possible, we shall restrict ourselves to point processes on two dimensional surfaces, as well as simple and regular partitions of these. The main interest of this study is to explore how well these reconstructions hold up as a function of size of the subregions, which we shall refer to as "quadrats". The basic approach of the study is discussed in more detail in the next section.

1.1 Main approach

To investigate the ability to adequately reproduce the core properties of a point process from aggregate statistics, we propose the following basic approach:

1. Specify a family of point process models, $p_\theta(\boldsymbol{x})$, and associated parameters, θ.
2. Generate a sample realization (point pattern), $\boldsymbol{x}$, of the point process over a window, W.
3. (a) Given a partition of the window into disjoint regions, $A_i \subset W$, (b) compute aggregated counts $N_i = N_{\boldsymbol{x}}(A_i) = \sum_{x \in \boldsymbol{x}} \mathbb{I}_{x \in A_i}$. The number of regions (inversely proportional to the size of each region) will be referred to as the *resolution* of the partition.
4. (a) Given these aggregated counts, simulate new patterns, $\tilde{\boldsymbol{x}}_i$ consisting of N_i (e.g. homogeneously distributed) events over corresponding region A_i. (b) Then combine these into a point pattern $\tilde{\boldsymbol{x}} = \cup_i \tilde{\boldsymbol{x}}_i$ over the entire window W, which we shall call the *reconstructed point pattern*.
5. (a) Compare the reconstructed pattern $\tilde{\boldsymbol{x}}$ to the original pattern $\boldsymbol{x}$, and (b) try to ascertain whether the reconstructed pattern $\tilde{\boldsymbol{x}}$ could be considered a reasonable approximation of a "typical" realization of the original point process model $p_\theta(\boldsymbol{x})$. In order to evaluate (a) and (b) we introduce different measures of dissimilarity of point patterns.

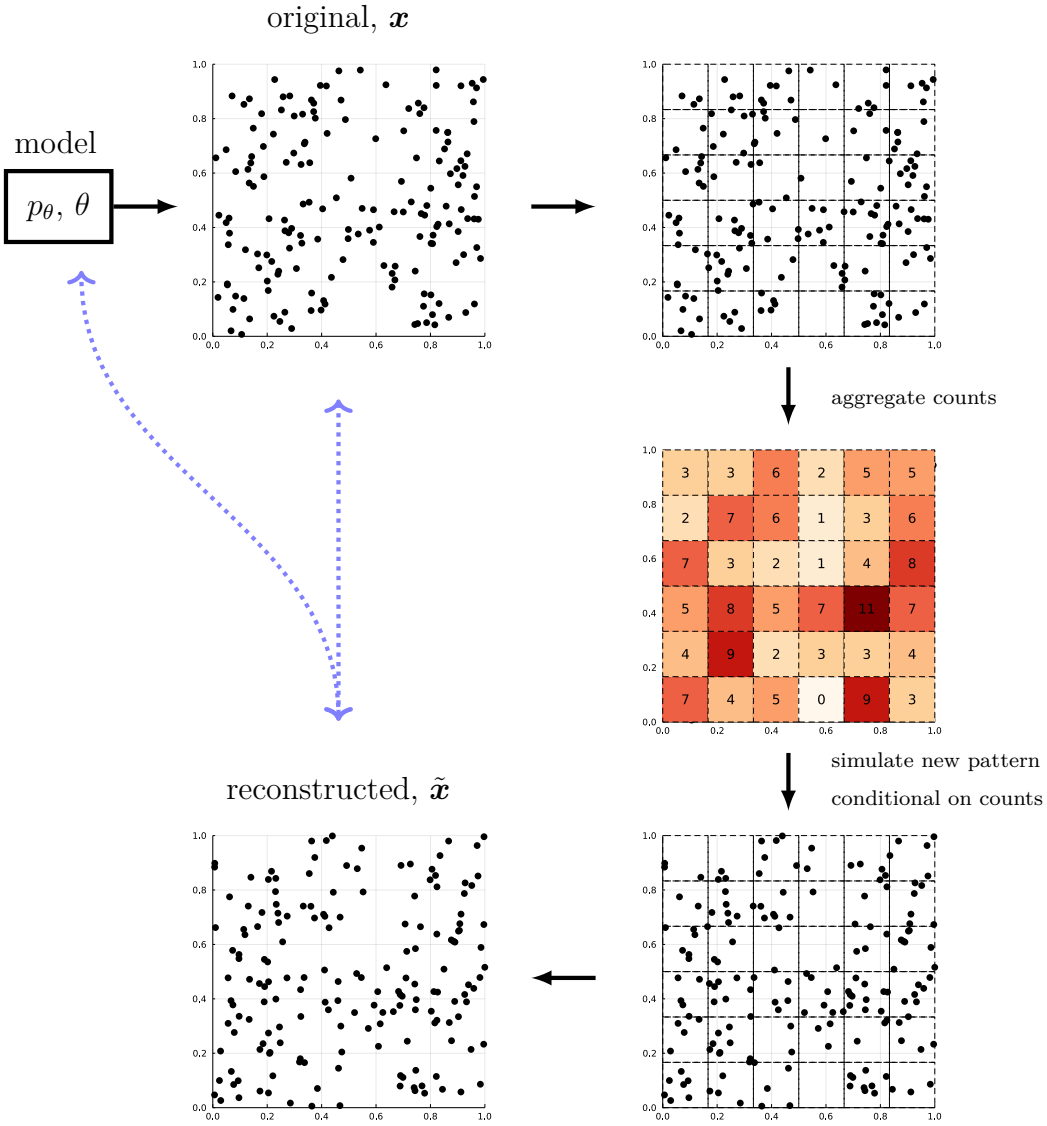


Figure 1.1: Illustration of the main approach.

1.2 Outline of the thesis

In chapter 2 we introduce the theory of finite point processes along with most of the concepts used in the thesis. In particular, we present the the main types of point process models that shall be used, as well as methods of generating point patterns, parameter inference and important properties that shall later be used when constructing dissimilarity measures and theoretical analysis. The intention is for the thesis to be as self-contained as possible.

In chapter 3 we discuss in detail the methods and models that shall be used in the study. Specifically, we present the reconstruction process in more detail, and give a couple of variations of these. We specify the particular point process models that we shall use and discuss their advantages and limitations. We also present the dissimilarity measures that shall be used to evaluate the quality of reconstruction.

In chapter 4 we present the results of our study. This includes a theoretical analysis of the reconstruction process itself, and the results of simulations. Finally, our results are summarised and discussed in chapter 5.

Code used in this thesis

All numerical computations as well as figures presented in this thesis were generated using a package in the Julia programming language written by the author specifically for this thesis. Samples from point processes were generated according to the algorithms described in chapter 2, while reconstruction and numerical analysis was performed according to the algorithms in chapter 2-3. The code is available on GitHub: `https://github.com/jens81/PointProcessReconstruction`

2

Theory

In this chapter, we present the theoretical background underlying the methods and models used in the thesis. The aim is for this section to be as self contained as possible, but for more details we recommend [4],[5],[6] and [7] on which most of this introduction is based.

2.1 Theory of finite point processes

A *point pattern* is an unordered collections of points, $\boldsymbol{x} = \{x_1, x_2, \ldots\}$, each in the same space $x_i \in \mathcal{X}$, and as such, it can be seen as a subset $\boldsymbol{x} \subset \mathcal{X}$. If the number of points in the set, $|\boldsymbol{x}|$, is fixed and/or known, we will sometimes suffix the symbol of the set by the number of points, e.g. $\boldsymbol{x}_n = \{x_1, \ldots, x_n\}$. A *point process*, is a stochastic process for generating random point patterns, $\boldsymbol{X} = \{X_1, X_2, \ldots\}$, where, in general, both the locations $X_i \in \mathcal{X}$, $i = 1, 2, \ldots$, and the number of points $|\boldsymbol{X}|$ are random. In the next two sections, I will follow Illian, et al. [6] and first introduce point processes with a fixed number of points in section 2.1.1, and then point processes with a random number of points in section 2.1.2. This way, one can gradually introduce some of the intricacies of point processes, as well as the notation which will be used in this thesis.

2.1.1 Point processes with fixed number of points

When working with a set of points $x_i \in \mathcal{X}$, $i = 1, \ldots, n$, it is convenient to work in the space $\mathcal{X}^n$. However, since a point pattern $\boldsymbol{x}_n = \{x_1, \ldots, x_n\}$, or *point configuration* as it is sometimes called, is an unordered set, it refers to the equivalence class of the $n!$ possible points $(x_{\phi(1)}, \ldots, x_{\phi(n)}) \in \mathcal{X}^n$, where ϕ is an element of the permutation group, P_n of n items. The space, $S^{(n)}$, of point configurations with n-points is given by the quotient space $\mathcal{X}^n / P_n$. There are several ways to deal with this issue, as laid out in [7]. In this thesis we shall consider only functions $h(\boldsymbol{x}_n) = h(x_1, \ldots, x_n)$ which are symmetric with respect to their arguments, and the redundancy this brings about will be taken care of explicitly. The expectation of such a function, represented abstractly as an integral with respect to some probability measure $\mathbb{P}$ over the configuration space, can be rewritten in terms of an integral over the space $\mathcal{X}^n$:

$$\mathbb{E}[h(\boldsymbol{X}_n)] = \int_{S^{(n)}} h(\boldsymbol{x}_n)\, \mathrm{d}\mathbb{P}(\boldsymbol{x}_n) = \int_{\mathcal{X}^n} h(\boldsymbol{x}_n) f_n(\boldsymbol{x}_n) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_n}{n!},$$

where the n-point *probability density* $f_n(\boldsymbol{x}_n)$ is also symmetric with respect to its arguments and, in this sense, puts equal weight on all of the equivalent points $(x_{\phi(1)}, \ldots, x_{\phi(n)}) \in \mathcal{X}^n$. The factor $1/n!$ is to account for the double counting that occurs due to treating these $n!$ points as separate.
The probability density $f_n(\boldsymbol{x}_n)$ is normalized to unity in the sense that

$$\int_{\mathcal{X}^n} f_n(\boldsymbol{x}_n) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_n}{n!} = 1,$$

while in some literature one uses instead the Janossy density $j_n(\boldsymbol{x}_n) = f_n(\boldsymbol{x}_n)/n!$ which has the simpler normalization

$$\int_{\mathcal{X}^n} j_n(\boldsymbol{x}_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n = 1.$$

2.1.1.1 Binomial point process

If the points are distributed independently, the Janossy density can be factorized $j_n(\boldsymbol{x}_n) = \prod_{i=1}^n j(x_i)$, where $j(x_i)$ is the individual probability density of each point, x_i, $i = 1, \ldots, n$. Note that the interchangeability of the points necessitates that they have the same distribution. When the points are uniformly distributed, i.e. $j(x_i) = 1/|\mathcal{X}|$, the process is called a *Binomial point process*, with corresponding densities:

$$j_n(\boldsymbol{x}_n) = \frac{1}{|\mathcal{X}|^n}, \qquad f_n(\boldsymbol{x}_n) = \frac{n!}{|\mathcal{X}|^n}.$$

We shall later use the more general case where the points are independently distributed, but $j(x_i) \neq 1/|\mathcal{X}|$, in which case we refer to this as an *inhomogeneous Binomial point process*.

2.1.1.2 Gibbs process with fixed number of points

Dependence can be introduced in a systematic fashion by rewriting[1]

$$f_n(\boldsymbol{x}_n) = \frac{1}{Z} e^{-V(\boldsymbol{x}_n)}, \qquad V(\boldsymbol{x}_n) = \sum_{i=1}^n \phi_1(x_i) + \sum_{i,j=1}^n \phi_2(x_i, x_j) + \ldots,$$

where Z is a normalization coefficient, chosen such that $\int_{\mathcal{X}^n} f_n(\boldsymbol{x}_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n/n! = 1$, and $V(\boldsymbol{x}_n)$ is the "potential" of the point configuration, expanded in a series with ϕ_1 being a single point potential, ϕ_2 being a pair-potential, etc. Keeping only the single point potential ϕ_1 reduces to the inhomogeneous Binomial point process:

$$f_n(\boldsymbol{x}_n) = \frac{1}{Z} \prod_{i=1}^n e^{-\phi_1(x_i)} = \prod_{i=1}^n j(x_i),$$

while keeping only the pair-potential, and requiring it to be homogeneous and isotropic, $\phi_2(x_i, x_j) = \phi_2(|x_i - x_j|)$ leads us to what is known as a Gibbs process

$$f_n(\boldsymbol{x}_n) = \frac{1}{Z} e^{-\sum_{i \neq j} \phi_2(\|x_i - x_j\|)}.$$

[1]The exponential form originates from statistical mechanics where it is obtained by maximising the entropy $-\mathbb{E}[\ln f_n(\boldsymbol{X}_n)]$ under the constraint of constant energy (here potential) $\mathbb{E}[V(\boldsymbol{X}_n)] =$ const.

For example, a hard core Strauss model [8] assigns lower probability to point configurations which have many points in close vicinity by setting

$$\phi_2(\|x_i - x_j\|) = \gamma \mathbb{I}_{\|x_i - x_j\| < R}.$$

where $\gamma, R > 0$ are the parameters of the interaction potential. In other words, the potential in the hard core Strauss model counts the number of pairs of points $x_i, x_j \in \boldsymbol{x}_n$ that are within a distance R from each other:

$$f_{n,\mathrm{S}}(\boldsymbol{x}_n) = \frac{1}{Z} e^{-V_{\mathrm{S}}(\boldsymbol{x}_n)}, \quad V_{\mathrm{S}}(\boldsymbol{x}_n) = \gamma \sum_{i \neq j} \mathbb{I}_{\|x_i - x_j\| < R}.$$

This type of process, which promotes configurations with points far from each other, are called *inhibitive* or *regular* processes. In this thesis, however, we shall be more interested in *attractive* or *clustering* processes, i.e. those that promote clustered point configurations. One might assume that such a process can easily achieved by reversing the sign of γ in the Strauss process. Unfortunately, such a process would not be well defined when moving to the more general case of random number of points. Intuitively, this can be understood from the fact that, for a fixed number of points, the probability density of such a process would be maximized when all points are within a distance R from each other, and when allowing a random number of points, the probability density can be made infinite by putting an infinite amount of points within a distance R from each other. Thus we here consider instead a regularized version, known as the Geyer model[9].
In the *Geyer model*, the probability density is given by

$$f_{n,\mathrm{G}}(\boldsymbol{x}_n) = \frac{1}{Z} e^{-V_{\mathrm{G}}(\boldsymbol{x}_n)}, \quad V_{\mathrm{G}}(\boldsymbol{x}_n) = -\gamma \min\Big(s^*, \sum_{i \neq j} \mathbb{I}_{\|x_i - x_j\| < R}\Big),$$

where apart from changing the sign in front of γ, we also introduced a saturation parameter, s^*, ensuring that, while the process promotes point configurations with points close to each other, it does not do so beyond a threshold, s^*, of points within a distance R from each other.

2.1.1.3 Area interaction process

The potential, $V(\boldsymbol{x}_n)$, can also be modeled without directly specifying k-point interactions, as illustrated by the *area interaction model*[10] which is defined by

$$f_{n,\mathrm{A}}(\boldsymbol{x}_n) = \frac{1}{Z} e^{-V_{\mathrm{A}}(\boldsymbol{x}_n)}, \quad V_{\mathrm{A}}(\boldsymbol{x}_n) = \eta \left| \cup_i b_R(x_i) \right|,$$

where $b_R(x_i) = \{x \in \mathcal{X} : \|x - x_i\| < R\}$, is the ball centered at x_i, of radius R.

2.1.1.4 Simulating point processes with fixed number of points

In the case of independently distributed points, one can sample each point independently from the 1-point distributions $j(x_i)$. For the homogeneous Binomial point

process, this means sampling each point uniformly over $\mathcal{X}$, whereas the inhomogeneous case might require using e.g. rejection sampling if the distribution $j(x_i)$ is not easy to sample from.
Gibbs point processes with fixed number of points can easily be simulated by using a birth-death MCMC sampling scheme as described in [6]. Here one starts with an initial point configuration $\boldsymbol{x}_n^{(0)}$, then in subsequent steps $t = 0, 1, \ldots$, iteratively deleting one of the points, say x_k, from $\boldsymbol{x}_n^{(t)}$ at random with probability $1/n$, leaving you with a temporary set, $\boldsymbol{x}_n^{(t)} \setminus \{x_k\}$. Then the number of points is restored by sampling a new point, u, from the conditional 1-point probability density function:

$$f_1\left(u \mid \boldsymbol{x}_n^{(t)} \setminus \{x_k\}\right) = \frac{f_n\left(\{u\} \cup \left(\boldsymbol{x}_n^{(t)} \setminus \{x_k\}\right)\right)}{\int_{\mathcal{X}} f_n(\boldsymbol{x}_n^{(t)})\,\mathrm{d}x_k},$$

and then setting $\boldsymbol{x}_n^{(t+1)} = \{u\} \cup \left(\boldsymbol{x}_n^{(t)} \setminus \{x_k\}\right)$. Here, $\int_{\mathcal{X}} f_n(\boldsymbol{x}_n^{(t)})\,\mathrm{d}x_k$ is the marginal distribution of the remaining points $\boldsymbol{x}_n^{(t)} \setminus x_k$. Upon convergence the point configurations generated by this method will represent draws from the distribution $f_n(\boldsymbol{x}_n)$. Using some computation, one can show that the 1-point conditional probability density, when keeping terms up to the pair-potentials, reduces to

$$f_1(u \mid \boldsymbol{x} \setminus \{x_i\}) \propto e^{-\tilde{V}(u;\boldsymbol{x}_n \setminus \{x_i\})}, \qquad \tilde{V}(u; \boldsymbol{x}_n \setminus \{x_i\}) = \phi_1(u) + \sum_{x \in \boldsymbol{x}_n \setminus \{x_i\}} \phi_2(\|u - x\|).$$

Sampling from this distribution can be done using, e.g. rejection sampling, provided that one can determine an upper bound $M > \exp(-\tilde{V}(u; \boldsymbol{x}_n \setminus \{x_i\}))$. In the case of the (homogeneous, i.e. $\phi_1(u) = \text{constant}$) Geyer process, this kind of bound is naturally set by the saturation parameter, s^*, in which case $M = e^{-\gamma \cdot s^*}$.

2.1.2 Point processes with random number of points

Allowing for a random number of points, we must extend the space of configurations to $S = \cup_n S^{(n)}$, and the expectation of a function $h(\boldsymbol{X})$, still invariant with respect to its (unspecified number of) arguments can be written

$$\mathbb{E}[h(\boldsymbol{X})] = \sum_{n=0}^{\infty} \int_{S^{(n)}} h(\boldsymbol{x}_n)\,\mathrm{d}\mathbb{P}^{(n)}(\boldsymbol{x}_n) = \sum_{n=0}^{\infty} \int_{\mathcal{X}^n} h(\boldsymbol{x}_n) p(\boldsymbol{x}_n) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_n}{n!}.$$

Here, the *probability density* $p(\boldsymbol{x})$ is normalised to unity in the sense that

$$\sum_{n=0}^{\infty} \int_{\mathcal{X}^n} p(\boldsymbol{x}_n) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_n}{n!} = 1,$$

while upon integration only over a specific slice $S^{(n)}$ with a fixed number of points yields the probability, p_n, of observing exactly n-points within the space $\mathcal{X}$:

$$\int_{\mathcal{X}^n} p(\boldsymbol{x}_n) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_n}{n!} = \mathbb{E}[\mathbb{I}_{\boldsymbol{x} \in S^{(n)}}] = \mathbb{P}\{\boldsymbol{x} \in S^{(n)}\} = p_n.$$

Thus we can define the *location density* $f_n(\boldsymbol{x}_n) = p(\boldsymbol{x}_n)/p_n$, which is normalised to unity over the subset of the configuration space with n points:

$$\int_{\mathcal{X}^n} f_n(\boldsymbol{x}_n) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_n}{n!} = 1,$$

and thus this location density is equivalent to the n-point density function f_n from the previous section. Again, sometimes it may be preferable to work with the Janossy density $j_n(\boldsymbol{x}_n) = f_n(\boldsymbol{x}_n)/n!$, which satisfies the simpler normalisations

$$\int_{\mathcal{X}^n} j_n(\boldsymbol{x}_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n = 1.$$

To help the reader navigate the different types of densities, I provide the following set of equations relating the three to each other:

$$p(\boldsymbol{x}_n) = p_n f_n(\boldsymbol{x}_n) = n! p_n j_n(\boldsymbol{x}_n).$$

We note, for later use, that the probability density $p(\boldsymbol{x}_n)$ can be interpreted as the joint probability density $p(\boldsymbol{x}, n) \equiv p(\boldsymbol{x} \cap \{\boldsymbol{x} \in S^{(n)}\})$ or simply $p(\boldsymbol{x} \cap \{|\boldsymbol{x}| = n\})$. In this sense, the location density $f_n(\boldsymbol{x}_n) = p(\boldsymbol{x}_n)/p_n = p(\boldsymbol{x}, n)/\mathbb{P}\{|\boldsymbol{x}| = n\}$ is formally the probability density conditioned on the number of points:

$$f_n(\boldsymbol{x}_n) = p(\boldsymbol{x} \mid n) = p(\boldsymbol{x} \mid \{|\boldsymbol{x}| = n\}).$$

In the same sense, the Janossy density is related to the conditional probability density of the point configuration through:

$$j_n(\boldsymbol{x}_n) = p(\boldsymbol{x} \mid n)/n!.$$

2.1.2.1 Poisson point process

Following the approach in the section on point processes for fixed number of points, we consider first the case where point locations are independently distributed, in which case the Janossy density factorises $j_n(\boldsymbol{x}_n) = \prod_{i=1}^n j(x_i)$. If we also require the points to uniformly distributed over $\mathcal{X}$ we have $j_n(\boldsymbol{x}_n) = 1/|\mathcal{X}|^n$. Finally, requiring the number of points be Poisson distributed with some parameter Λ, i.e. $p_n = \Lambda^n e^{-\Lambda}/n!$ we get:

$$p(\boldsymbol{x}_n) = n! p_n j_n(\boldsymbol{x}_n) = n! \cdot \frac{\Lambda^n e^{-\Lambda}}{n!} \cdot \frac{1}{|\mathcal{X}|^n} = \left(\frac{\Lambda}{|\mathcal{X}|}\right)^n e^{-\Lambda}.$$

Setting $\Lambda = \lambda|\mathcal{X}|$ we get

$$p(\boldsymbol{x}_n) = \lambda^n e^{-\lambda|\mathcal{X}|},$$

where λ is known as the local intensity of the homogeneous *Poisson point process*. The process can be generalised to an *inhomogeneous Poisson point process* by introducing a location dependent intensity function $\lambda(u)$, in which case the probability density becomes:

$$p(\boldsymbol{x}_n) = e^{-\int_{\mathcal{X}} \lambda(u)\,\mathrm{d}u} \prod_{i=1}^n \lambda(x_i),$$

corresponding to an independent process with 1-point densities $j(x_i) \propto \lambda(x_i)$, i.e. $j(x_i) = \lambda(x_i)/\int_{\mathcal{X}} \lambda(u)\,\mathrm{d}u$.

2.1.2.2 Interacting point processes

Similar to section 2.1.1.2 we can specify:

$$p(\boldsymbol{x}_n) = \frac{1}{Z} e^{-V(\boldsymbol{x}_n)}, \qquad V(\boldsymbol{x}_n) = \phi_0\, n + \sum_{i=1}^{n} \phi_1(x_i) + \sum_{i,j=1}^{n} \phi_2(x_i, x_j) + \dots$$

While the constant ϕ_0 could be absorbed, e.g. by the 1-point potential $\phi_1(x_i)$, we keep it separate to highlight its role as a "chemical activity". Sometimes one uses instead $z = e^{-\phi_0}$ to denote the "activation". If we turn the potentials off, we get $p(\boldsymbol{x}_n) = z^n Z^{-1}$, and $p_n = \int_{\mathcal{X}^n} p(\boldsymbol{x}_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n / n! = (z|\mathcal{X}|)^n\, Z^{-1}$. Normalising such that $\sum_{n=0}^{\infty} p_n = 1$, i.e. $Z = \sum_{n=0}^{\infty} (z|\mathcal{X}|)^n / n! = e^{z|\mathcal{X}|}$ yields finally $p_n = (z|\mathcal{X}|)^n\, e^{-z|\mathcal{X}|}$, and $p(\boldsymbol{x}_n) = z^n e^{-z|\mathcal{X}|}$. Hence, in this limit we recover the homogeneous Poisson point process with local intensity $\lambda = z$. In general, however, the distribution p_n depends on both the activation $z = e^{-\phi_0}$, as well as the interactions. Of course, in the more general case the normalisation Z is also intractable, so we can not obtain a closed form expression for p_n. This implies that in general the likelihood $p(\boldsymbol{x}_n)$ is also intractable and therefore maximum likelihood estimation as well.
We also define the area interaction point process for random number of points similar to 2.1.1.3:

$$p_{\mathrm{A}}(\boldsymbol{x}_n) = \frac{1}{Z} e^{-V_{\mathrm{A}}(\boldsymbol{x}_n)}, \quad V_{\mathrm{A}}(\boldsymbol{x}_n) = \phi_0 n + \eta\, \left|\cup_i b_R(x_i)\right|.$$

In the same sense as before, the process reduces to the Poisson point process when $\eta = 0$.

2.1.2.3 Papangelou conditional intensity

While the exact form of $p(\boldsymbol{x})$ is intractable due to the normalization factor, there is a very useful quantity known as the *Papangelou conditional intensity*[11], here defined as

$$\lambda(u; \boldsymbol{x}) = \frac{p(\{u\} \cup \boldsymbol{x})}{p(\boldsymbol{x})}.$$

It can be used in an MCMC scheme to sample patterns from any distribution (see section 2.1.2.5) for which the Papangelou conditional intensity has a tractable form, it also facilitates a class of relations for expectations that can be used for model checking (section 2.1.2.4) and parameter inference (section 2.2 and specifically 2.2.1). Therefore, we shall present the Papangelou conditional intensities for the point processes considered so far (note that the Papangelou conditional intensity can only be specified for point processes with random number of points):

$$\begin{aligned}
\text{Poisson:} \quad & \lambda(u; \boldsymbol{x}) = \lambda = e^{-\phi_0} \\
\text{Geyer:} \quad & \lambda(u; \boldsymbol{x}_n) = e^{-(\phi_0 + \phi_1(u) + \gamma\, t_R(u; \boldsymbol{x}_n))} \\
\text{Area int.:} \quad & \lambda(u; \boldsymbol{x}_n) = e^{-(\phi_0 + \phi_1(u) + \eta\, C_R(u, \boldsymbol{x}_n))}
\end{aligned}$$

where $C_R(u; \boldsymbol{x}_n) = \left|\bigcup_{x \in \boldsymbol{x} \cup \{u\}} b_R(x_i)\right| - \left|\bigcup_{x \in \boldsymbol{x}} b_R(x_i)\right|$ and $t_R(u; \boldsymbol{x}) = \sum_{x \in \boldsymbol{x}} \mathbb{I}_{\|u - x\|}$.

2.1.2.4 GNZ formula

Many observables can be expressed as a sum over the individual points x, in relation to the other points in the set $\boldsymbol{X}$, i.e. of the form $\sum_{x\in\boldsymbol{X}} h(x;\boldsymbol{X}\setminus\{x\})$. A result from Georgii-Nguyen-Zessin [12][13] contends that the expectation of such a function satisfies the following relation (often called the GNZ formula)

$$\mathbb{E}\left[\sum_{x\in\boldsymbol{x}} h(x;\boldsymbol{X}\setminus\{x\})\right] = \mathbb{E}\left[\int_{\mathcal{X}} h(u;\boldsymbol{X})\lambda(u;\boldsymbol{X})\,\mathrm{d}u\right]$$

The relation can be proven pretty straight-forwardly by starting from the RHS expectation:

$$\begin{aligned}\mathbb{E}\left[\int_{\mathcal{X}} h(u;\boldsymbol{X})\lambda(u;\boldsymbol{X})\,\mathrm{d}u\right] &= \sum_{n=0}^{\infty}\int_{\mathcal{X}^n}\left(\int_{\mathcal{X}} h(u;\boldsymbol{x}_n)\lambda(u;\boldsymbol{x}_n)\,\mathrm{d}u\right)p(\boldsymbol{x}_n)\frac{\mathrm{d}x_1\cdots\mathrm{d}x_n}{n!}\\ &= \sum_{n=0}^{\infty}\int_{\mathcal{X}^{n+1}} h(u;\boldsymbol{x}_n)\underbrace{\lambda(u;\boldsymbol{x}_n)p(\boldsymbol{x}_n)}_{p(\boldsymbol{x}_n\cup\{u\})}\frac{\mathrm{d}x_1\cdots\mathrm{d}x_n\,\mathrm{d}u}{n!}\end{aligned}$$

Defining now the set $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n\cup\{u\}$, and relabeling the points $\{x_1,\ldots,x_{n+1}\}$, while also utilising that the points of the set $\boldsymbol{x}_{n+1}$ are interchangeable, so that we can replace the integral over u by an integral over any of the points $x_1,\ldots,x_{n+1}$, or, as is done below, by a sum of integrals over each of them, and dividing by $n+1$:

$$\begin{aligned}&\mathbb{E}\left[\int_{\mathcal{X}} h(u;\boldsymbol{X})\lambda(u;\boldsymbol{X})\,\mathrm{d}u\right] = \ldots\\ &= \sum_{n=0}^{\infty}\int_{\mathcal{X}^{n+1}}\frac{1}{n+1}\sum_{i=1}^{n+1} h(x_i;\boldsymbol{x}_{n+1}\setminus\{x_i\})\,p(\boldsymbol{x}_{n+1})\frac{\mathrm{d}x_1\cdots\mathrm{d}x_i\cdots\mathrm{d}x_{n+1}}{n!}\\ &= \sum_{m=1}^{\infty}\int_{\mathcal{X}^m}\sum_{i=1}^{m} h(x_i;\boldsymbol{x}_m\setminus\{x_i\})\,p(\boldsymbol{x}_m)\frac{\mathrm{d}x_1\cdots\mathrm{d}x_i\cdots\mathrm{d}x_m}{m!}\\ &= \mathbb{E}\left[\sum_{x\in\boldsymbol{X}} h(x;\boldsymbol{X}\setminus\{x\})\right]\end{aligned}$$

where in going to the third line we set $m = n+1$ and in the last step utilised the fact that $\sum_{x\in\emptyset}\ldots = 0$ by definition, such that the $m=0$ case can be included.
As a special case of the GNZ-formula, we note the case when $h(x;\boldsymbol{X}\setminus\{x\}) = h(x)$:

$$\mathbb{E}\left[\sum_{x\in\boldsymbol{X}} h(x)\right] = \mathbb{E}\left[\int_{\mathcal{X}} h(u)\lambda(u;\boldsymbol{X})\,\mathrm{d}u\right] = \int_{\mathcal{X}} h(u)\rho(u)\,\mathrm{d}u, \qquad \rho(u) = \mathbb{E}[\lambda(u;\boldsymbol{X})]$$

known as the Campbell-Mecke theorem. Here $\rho(u)$ is the 1-point intensity function. In particular, using the number of points $|\boldsymbol{X}\cap A| = \sum_{x\in\boldsymbol{X}}\mathbb{I}_{x\in A}$ in a given subset $A\subseteq\mathcal{X}$ we get

$$\mathbb{E}\Big[|\boldsymbol{X}\cap A|\Big] = \mathbb{E}\left[\sum_{x\in\boldsymbol{X}}\mathbb{I}_{x\in A}\right] = \mathbb{E}\left[\int_{\mathcal{X}}\mathbb{I}_{u\in A}\lambda(u;\boldsymbol{X})\,\mathrm{d}u\right] = \int_A\rho(u)\,\mathrm{d}u$$

i.e. the expected number of points in A is given by the integral of the 1-point intensity over A.

2.1.2.5 Generating point patterns through Metropolis-Hastings MCMC

Point processes can be generated through iterative MCMC algorithms[4],[6]. In this section we present the Metropolis-Hastings birth-death algorithm. It starts at $t = 0$ with an initial configuration $\boldsymbol{x}^{(0)}$, that does not violate the basic conditions of the point process. Then in iterative steps, one proposes randomly a "birth", i.e. addition of a point to the point configuration $\boldsymbol{x}^{(t)}$, or a "death", i.e. a deletion of a point from the point configuration, $\boldsymbol{x}^{(t)}$. Denote the probability of a "birth" to be $q(\boldsymbol{x}^{(t)})$ and the probability of a "death" to be $1 - q(\boldsymbol{x}^{(t)})$. Let's also denote the probability density for the location of the new point as $b(u; \boldsymbol{x}^{(t)})$, and the probability of deleting point $x_i \in \boldsymbol{x}^{(t)}$ by $d(x_i; \boldsymbol{x}^{(t)})$. Suppose also the proposals are accepted with probability $A(\boldsymbol{x}_{\text{old}}; \boldsymbol{x}_{\text{new}})$. The detailed-balance condition is then given by (here using the shorthand $\boldsymbol{x} = \boldsymbol{x}^{(t)}$ for notational convenience):

$$q(\boldsymbol{x})b(u;\boldsymbol{x})A(\boldsymbol{x},\boldsymbol{x}\cup\{u\})p(\boldsymbol{x}) = (1-q(\boldsymbol{x}\cup\{u\}))d(u;\boldsymbol{x}\cup\{u\})A(\boldsymbol{x}\cup\{u\},\boldsymbol{x})p(\boldsymbol{x}\cup\{u\})$$

The Hastings-ratio is then given by

$$\frac{A(\boldsymbol{x},\boldsymbol{x}\cup\{u\})}{A(\boldsymbol{x}\cup\{u\},\boldsymbol{x})} = \frac{(1-q(\boldsymbol{x}\cup\{u\}))}{q(\boldsymbol{x})}\frac{d(u;\boldsymbol{x}\cup\{u\})}{b(u;\boldsymbol{x})}\cdot\frac{p(\boldsymbol{x}\cup\{u\})}{p(\boldsymbol{x})} = r(u;\boldsymbol{x})$$

where we immediately recognise the Papangelou conditional intensity $\lambda(u;\boldsymbol{x}) = p(\boldsymbol{x}\cup\{u\})/p(\boldsymbol{x})$.
Convenient choices for the probabilities are

$$q(\boldsymbol{x}) = \frac{1}{2}, \qquad b(u;\boldsymbol{x}) = \frac{1}{|\mathcal{X}|}, \qquad d(x_i;\boldsymbol{x}) = \frac{1}{|\boldsymbol{x}|}$$

i.e. equal probability of "birth" or "death" proposals, uniform proposal of new point, and uniform proposal of point to be deleted. With these choices, the Hastings ratio becomes

$$r(u;\boldsymbol{x}) = \frac{|\mathcal{X}|}{|\boldsymbol{x}|+1}\lambda(u;\boldsymbol{x})$$

It can be easily verified that the choice

$$A(\boldsymbol{x},\boldsymbol{x}\cup\{u\}) = \min\Big(1, r(u;\boldsymbol{x})\Big), \qquad A(\boldsymbol{x}\cup\{u\},\boldsymbol{x}) = \min\left(1, \frac{1}{r(u,\boldsymbol{x})}\right)$$

satisfies the detailed balance condition. Thus, a "birth" with new point u added to $\boldsymbol{x}$, is accepted with probability $\min(1, r(u;\boldsymbol{x})$, whereas a "death" with a deleted point $x_i \in \boldsymbol{x}$, is accepted with probability $\min(1, 1/r(x_i,\boldsymbol{x}))$. The following pseudocode provides the Metropolis-Hastings MCMC birth-death algorithm:

Choose $q \in (0,1)$ for probability of "birth" proposal.
Set $t = 0$ and initialise a point configuration $\boldsymbol{x}^{(0)}$.
for $t = 1 : T$ **do**
    Draw $r \sim \text{Uniform}([0,1])$
    if $r < q$ **then**
        Draw point to be added: $u \sim \text{Uniform}(\mathcal{X})$, and $r_b \sim \text{Uniform}([0,1])$

7: **if** $r_b < \min(1, r(u; \boldsymbol{x}^{(t-1)}))$ **then**
8: Accept "birth" and set $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} \cup \{u\}$.
9: **else**
10: Set $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)}$.
11: **end if**
12: **else**
13: Draw point to be deleted: $x_i \sim \text{Uniform}(\boldsymbol{x}^{(t)})$, and $r_d \sim \text{Uniform}([0, 1])$
14: **if** $r_d < \min(1, 1/r(x_i; \boldsymbol{x}^{(t-1)}))$ **then**
15: Accept "death" and set $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} \setminus \{x_i\}$.
16: **else**
17: Set $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)}$.
18: **end if**
19: **end if**
20: **end for**

Assuming the MCMC chain has reached convergence by $t = T$, the resulting point configuration $\boldsymbol{x}^{(T)}$ will be a sample from $p(\boldsymbol{x})$. Convergence can be monitored by analysing trace plots of some point configuration statistics, such as the number of points for instance. It should be noted that there exists also methods for "perfect simulations" where the Markov process itself indicates when its equilibrium is reached.

In the implementation used in this thesis, we also added a third option, besides "birth" and "death", denoted "move", where a point is deleted and subsequently a new point is added at a random location in $\mathcal{X}$, all in one step, effectively moving the point to a new, random location. It has been observed that the addition of this step improves mixing and thus convergence of the MCMC chain.

2.1.3 Point processes on linear networks

A linear network, L, is a finite set of line segments, ℓ, that end at vertices [10]. Furthermore, line segments are only allowed to intersect at vertices. As such, we can view a linear network as embedded in a space $\mathcal{X}$, i.e. $L \subset \mathcal{X}$. The length, or measure, of L is given by the sum of lengths of line segments ℓ:

$$|L| = \sum_{\ell \in L} |\ell|$$

A point $u \sim \text{Uniform}(L)$ can be drawn uniformly from L by first selecting a specific line segment, ℓ, randomly with probabilities $\ell/|L|$, and then drawing a point from the selected line segment ℓ randomly $u \sim \text{Uniform}(L)$. The last step is achieved by first parametrising the line segment ℓ by $t \to \ell$, $t \in [0, 1]$, and drawing $t \sim \text{Uniform}([0, 1])$.

Gibbs processes as described in sections 2.1.1 and 2.1.2 can be defined also on a linear network, although care needs to be taken with respect to the choice of distance function. Often, the appropriate distance is computed on the linear network, e.g. shortest path, or something similar. In this thesis, however, the distance used will be the euclidean distance in the embedded space $\mathcal{X}$, i.e. the distance is of the birds-view-type.

Point processes on linear networks will, in this thesis, be generated using the same Metropolis-Hastings MCMC birth-death algorithm as described in 2.1.2.5, but replacing $\mathcal{X}$ by L, together with the sampling scheme described above to sample uniformly form L, although the matter can be somewhat delicate, see [14].

2.2 Parameter inference

Given a point pattern $\boldsymbol{x}$, over an observation window W, and a model with a probability density $p_\theta(\boldsymbol{x}) = p(\boldsymbol{x} \mid \theta)$, with parameters $\theta = (\theta_1, \theta_2, \ldots)$, there are different ways to estimate these parameters. In this thesis, we shall primarily be interested in families of exponential Gibbs models, i.e. Gibbs models with probability densities that can be written in the form

$$p_\theta(\boldsymbol{x}) = \frac{1}{Z_\theta} e^{\sum_i \theta_i \phi_i(\boldsymbol{x})}$$

with corresponding Papangelou conditional intensities

$$\lambda_\theta(u; \boldsymbol{x}) = e^{\sum_i \theta_i S_i(u;\boldsymbol{x})}, \qquad S_i(u; \boldsymbol{x}) = \phi_i(\boldsymbol{x} \cup \{u\}) - \phi_i(\boldsymbol{x})$$

The most straight forward approach uses maximum likelihood:

$$\hat{\theta} = \arg\max_\theta p_\theta(\boldsymbol{x})$$

or equivalently, maximising the log likelihood:

$$\hat{\theta} = \arg\max_\theta \log p_\theta(\boldsymbol{x}) = \arg\max_\theta \left(\log Z_\theta + \sum_i \theta_i \phi_i(\boldsymbol{x}) \right)$$

However, given that the normalisation Z_θ is generally intractable, this is usually not a feasible approach, although there exist Monte-Carlo methods (including Bayesian approaches with priors) [15].
A separate approach determines the parameters θ by instead maximising the pseudolikelihood [16, 17, 18],

$$\mathrm{PL}(\theta) = e^{-\int_W \lambda_\theta(u;\boldsymbol{x})\,\mathrm{d}u} \prod_{x_i \in \boldsymbol{x}} \lambda_\theta(x_i; \boldsymbol{x} \setminus \{x_i\})$$

which is equivalent to the likelihood for (inhomogeneous) Poisson models, i.e. Gibbs models in the absence of interactions, if we replace the intensity function by the Papangelou conditional intensity. Maximising the log pseudolikelihood

$$\hat{\theta} = \arg\max_\theta \log \mathrm{PL}(\theta) = \arg\max_\theta \left(-\int_W \lambda_\theta(u; \boldsymbol{x})\,\mathrm{d}u + \sum_{x_i \in \boldsymbol{x}} \log \lambda_\theta(x_i; \boldsymbol{x} \setminus \{x_i\}) \right)$$

for an exponential model yields the equations

$$\int_W S_j(u; \boldsymbol{x})\,\mathrm{d}u = \sum_{x_i \in \boldsymbol{x}} \frac{S_j(x_i; \boldsymbol{x} \setminus \{x_i\})}{\lambda_\theta(u; \boldsymbol{x} \setminus \{x_i\}}, \quad j = 1, 2, \ldots$$

The integral can be difficult to compute, and usually one uses a quadrature scheme, by replacing the integral by a sum over points $v \in \boldsymbol{D}$, weighting each term by $w(u) = |B(v)|/|W|$, i.e. the relative measure of the region, B_v, the point should represent,

$$\sum_{v \in \boldsymbol{D}} S_j(v; \boldsymbol{x}) w(v) = \sum_{x_i \in \boldsymbol{x}} \frac{S_j(x_i; \boldsymbol{x} \setminus \{x_i\})}{\lambda_\theta(u; \boldsymbol{x} \setminus \{x_i\}}, \quad j = 1, 2, \ldots$$

Extending the number of points used in the sum to also include $x_i \in \boldsymbol{x}$, i.e. summing over the points $u \in \boldsymbol{D} \cup \boldsymbol{x}$, is referred to as the Berman-Turner device [19]. Here the weights $w(v) = |\boldsymbol{x} \cap B(v)| \cdot |B(v)|/|W|$ are multiplied by the number of points $x \in \boldsymbol{x} \cap B(v)$. In this approach the log pseudolikelihood can be written as one sum:

$$\log \mathrm{PL}(\theta) = \sum_{u \in \boldsymbol{D} \cup \boldsymbol{x}} \left[\frac{\mathbb{I}_{u \in \boldsymbol{x}}}{w(u)} \log \lambda_\theta(u; \boldsymbol{x} \setminus \{u\}) - \lambda_\theta(u; \boldsymbol{x}) \right] w(u)$$

which can be mapped onto the log likelihood of weighted Poisson regression [20]:

$$\sum_{i=1}^{|\boldsymbol{x} \cup \boldsymbol{D}|} \left[y_i \log \lambda_i - \lambda_i \right] w_i, \qquad y_i = \mathbb{I}_{u_i \in \boldsymbol{x}} / w_i$$

where $\lambda_i = \lambda(u_i; \boldsymbol{x} \setminus \{u_i\})$ and $w_i = w(u_i)$. This has the advantage that one can use standard regression software to estimate θ. Unfortunately, this method requires quite a lot of dummy points (far more than the number of data points $x \in \boldsymbol{x}$), and that it is only asymptotically unbiased and typically underestimates the interaction parameters if the number of dummy points is not large enough.

2.2.1 Estimation by mapping to logistic regression

Baddeley, et al. [21] proposed an estimating function

$$e_\theta(\boldsymbol{X}, \boldsymbol{D}) = \sum_{u \in \boldsymbol{X}} \frac{\rho(u) S(u; \boldsymbol{X} \setminus \{u\})}{\lambda_\theta(u; \boldsymbol{X} \setminus \{u\}) + \rho(u)} - \sum_{u \in \boldsymbol{D}} \frac{S(u; \boldsymbol{X}) \lambda_\theta(u; \boldsymbol{X})}{\lambda_\theta(u; \boldsymbol{X}) + \rho(u)}$$

where $\boldsymbol{D}$ is an independent "dummy" point process (for example a binomial point process), and $\rho(u)$ its 1-point intensity function. An estimating function here is one which has an expected value of 0, when θ is chosen correctly (i.e. the true parameter values). To show that this function has this property, we use the GNZ formula for the point process $\boldsymbol{X}$ on the first sum, while using the Campbell-Mecke formula for the point process $\boldsymbol{D}$ given $\boldsymbol{X}$, for the second sum:

$$\mathbb{E}\left[\sum_{u \in \boldsymbol{X}} \frac{\rho(u) S(u; \boldsymbol{X} \setminus \{u\})}{\lambda_\theta(u; \boldsymbol{X} \setminus \{u\}) + \rho(u)} \right] = \mathbb{E}\left[\int_W \frac{\rho(u) S(u; \boldsymbol{X}) \lambda_\theta(u; \boldsymbol{X})}{\lambda_\theta(u; \boldsymbol{X}) + \rho(u)} \, \mathrm{d}u \right]$$

$$\mathbb{E}\left[\sum_{u \in \boldsymbol{D}} \frac{S(u; \boldsymbol{X}) \lambda_\theta(u; \boldsymbol{X})}{\lambda_\theta(u; \boldsymbol{X}) + \rho(u)} \mid \boldsymbol{X} \right] = \mathbb{E}\left[\int_W \frac{\rho(u) S(u; \boldsymbol{X}) \lambda_\theta(u; \boldsymbol{X})}{\lambda_\theta(u; \boldsymbol{X}) + \rho(u)} \, \mathrm{d}u \right],$$

thus showing that $\mathbb{E}[e_\theta(\boldsymbol{X}, \boldsymbol{D})] = 0$, when the expectation is taken with respect to both $\boldsymbol{X}$ and $\boldsymbol{D}$. The equation $e_\theta(\boldsymbol{x}, \boldsymbol{D}) = 0$ can be obtained by maximising a

function:

$$\begin{aligned}\log L(\theta) &= \sum_{u\in\boldsymbol{x}} \log\left(\frac{\lambda_\theta(u;\boldsymbol{x}\setminus\{u\})}{\lambda_\theta(u;\boldsymbol{x}\setminus\{u\})+\rho(u)}\right) + \sum_{u\in\boldsymbol{D}} \log\left(\frac{\rho(u)}{\lambda_\theta(u;\boldsymbol{x})+\rho(u)}\right)\\ &= \sum_{u\in\boldsymbol{x}\cup\boldsymbol{D}} \Big(Y(u)\log p(u) + (1-Y(u))\log(1-p(u))\Big)\end{aligned}$$

which is formally equivalent to the log likelihood of a set of Bernoulli trials with data $Y(u) = \mathbb{I}_{u\in\boldsymbol{x}}$ and probability

$$\mathbb{P}[Y(u)=1] = p(u) = \frac{\lambda_\theta(u;\boldsymbol{x}\setminus\{u\})}{\lambda_\theta(u;\boldsymbol{x}\setminus\{u\})+\rho(u)} = \frac{\exp\left(\theta^T S(u;\boldsymbol{x}\setminus\{u\}) - \log\rho(u)\right)}{1+\exp\left(\theta^T S(u;\boldsymbol{x}\setminus\{u\}) - \log\rho(u)\right)}$$

which where is parametrised by a logit-link with offset $-\log\rho(u)$. As such, one can again use standard statistical software that can handle logistic regression to infer the parameters. Importantly, this provides an unbiased estimation, independent on the number of dummy points used (though of course the accuracy is better the larger the number of dummy points).
This is the method primarily used in this report when estimating parameters for exponential Gibbs models.

3
Methods

This chapter outlines the specific methods and models used to explore the feasibility of reconstructing point patterns from spatially aggregated counts. To limit the scope of the simulation study, we shall limit ourselves to a few point process models, a simple and tractable scheme to partition the observation window into smaller subregions, as well as a few methods of reconstruction. We also outline the methods used in this thesis to evaluate the quality of reconstruction.

3.1 Point process models

Given that at least some of the motivation behind this simulation study is based on the possibility of using point processes to model the spread of disease, we shall limit ourselves to the class of attractive or clustering point process models. Disease is usually transmitted locally between people in close proximity to each other resulting in local clusters. We can imagine associating with each infected individual a region of exposure to others, based on the mechanisms of transmission but more importantly on the movement of the individual during the infectious stage. A very simple model could stipulate a spatial radius of infection, R, i.e. a maximum distance in either direction from the infected persons location of residence. This associates with each infected individual a disc within which the risk of infection of other individuals is increased.

Of course, the observation window, which itself represents a subregion (a country, state or county) is not an isolated system, so one also needs to take into account an influx of infected, possibly due to residents coming back from a trip outside of this subregion. This could be modeled by a background intensity, or risk of infection, causing infected individuals to pop up randomly around the observation window, even without having any infected nearby.

At the same time, population density is typically not constant, but organised into city centers, suburbs, etc, and a significant portion of the region might in fact be unpopulated. A higher population density typically implies more interactions between people, and thus a higher risk of infection. Some regions (like shopping malls, etc) may not be highly populated, but still have a large rate of interactions and thus represent regions of higher risk of infection. This increase in risk may be modeled by allowing the background intensity (or risk of infection) to be spatially inhomogeneous, with a larger risk in highly populated or highly visited regions.

Furthermore, people tend to live in close proximity to roads and other transportation networks and hubs. These can be represented as linear networks, and thus we can

define point processes directly on these, as we have seen in section 2.1.3.
At the same time, we wish to use a point process model which is highly interpretable and easy to work with. In the theory chapter 2, we have introduced the theory of finite point processes and focused mainly on the Gibbs point processes. This class of point process naturally includes the notions of a background intensity (homogeneous or inhomogeneous) as well as interactions between points/events. We also outlined methods to generate point configurations from these models, to perform parameter inference as well as diagnostic tools that could be used to compare an original point configuration with a reconstructed point configuration.
Given these specifications and constraints, we choose to work with the exponential family of Gibbs models with Papangelou intensities of the form

$$\lambda_{\boldsymbol{\theta}}(\boldsymbol{x}) = \exp\left(\boldsymbol{\theta}^T \boldsymbol{S}(u)\right) = \exp\left(\sum_{i=0}^{2} \theta_i S_i(u; \boldsymbol{x})\right).$$

Here $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$ are the parameters of the model, which together with

$$\boldsymbol{S}(u; \boldsymbol{X}) = (1, S_1(u), S_2(u; \boldsymbol{x}))$$

specifies the model in the following sense: $\theta_0 S_0(u; \boldsymbol{x}) = \theta_0$ corresponds to the constant background intensity, $\theta_1 S_1(u)$ models the spatial inhomogeneity of this background intensity, where $S_1(u)$ is a covariate density function, e.g. the population density. Finally $\theta_2 S_2(u; \boldsymbol{x})$ models the interaction between points, or events. In other words, it models the propensity of events to appear close to each other, i.e. it controls the clustering of events.
To interpret the model in more detail, consider the interpretation of the Papangelou conditional intensity $\lambda(u; \boldsymbol{x})$ as the intensity (or risk of infection) at a location u given the locations of the points/events in $\boldsymbol{x}$. In this sense, a larger value of $\theta_1 S_1(u)$ implies a larger risk at this location, due to e.g. a higher population density (if this is what $S_1(u)$ describes). A larger value of $\theta_2 S_2(u; \boldsymbol{x})$ implies a larger intensity at location u due to the presence of other points close by. In an epidemiological context, this would represent a higher risk of infection at location u due to many other infected in close proximity.
These three parameters, along with specification of $S_1(u)$ and $S_2(u; \boldsymbol{x})$ allows us to cover most of the considerations in the paragraphs above outlining how one might model the spread of disease using point processes. Of course, in this report we do not intend to accurately model an epidemic, nor do we work with real data or a real geographical region. Rather, the aim of this thesis is to explore the feasibility to reconstruct a point process based on spatially aggregated data. In light of this, we shall consider two specific, simple, models for the interaction, described below.

Geyer process

The Geyer process is briefly discussed in section 2.1.1, and it's Papangelou conditional intensity is presented in section 2.1.2.3. In the context of exponential Gibbs models, this corresponds to the interaction specification:

$$S_2(u; \boldsymbol{x}) = \min\left(s^*, \sum_{x \in \boldsymbol{x} \setminus \{u\}} \mathbb{I}_{\|u-x\|<R}\right).$$

In other words, $S_2(u; \boldsymbol{x})$ returns either the number of points within a distance R of the location u, or a saturation constant s^*, whichever is smallest. In an epidemiological context, this would imply that the risk of infection at a location u increases exponentially with every other infection within a radius R of that location, up until a specific count s^*, when the region is deemed saturated. If there are more infections than s^* within this radius, the risk no longer increases with the number of infected.

Area interaction process

While the Geyer process utilises pair-wise interactions, in the sense that it depends on the distances between u and other points of $\boldsymbol{x}$, the area interaction process depends on the total area covered by discs around the points in $\boldsymbol{x}$. Specifically, in the exponential model, the interaction is specified by

$$S_2(u; \boldsymbol{x}) = -\left(\left| \bigcup_{x \in \boldsymbol{x} \cup \{u\}} b_R(x) \right| - \left| \bigcup_{x \in \boldsymbol{x}} b_R(x) \right| \right).$$

Here $b_R(x)$ is a disc of radius R and center x. Thus $|\bigcup_{x \in \boldsymbol{x}} b_R(x)|$ represents the total area covered by the discs with centers at $x \in \boldsymbol{x}$. Thus the interaction $S_2(u; \boldsymbol{x})$ yields the negative of the increase in total area caused by adding a disc with center u. This way, the intensity at a location u is larger if it's associated disc overlaps significantly with the discs with centers in $\boldsymbol{x}$. In the generation of point configurations this promotes points in close proximity with each other, i.e. clustering.
A different, but equivalent up to renormalised parameters, way to express the area interaction is given below:

$$S_2(u; \boldsymbol{x}) = -\left| b_R(u) \setminus \bigcup_{x \in \boldsymbol{x}} b_R(x) \right| / \pi r^2.$$

In this form $S_2(u; \boldsymbol{x})$ is given by (the negative of) the fraction of the area of the disc around u which does not overlap with discs centered around $\boldsymbol{x}$. Due to being somewhat easier to compute, this is the representation of the area interaction process that we shall use in this thesis.
It should be noted, that since the interaction function $S_2(u; \boldsymbol{x})$ is negative, it's actual function is to suppress the intensity (when the location u is far from other points in $\boldsymbol{x}$) rather than to increase it. Therefore, one typically needs a larger background intensity to compensate.

3.1.1 Limitations of chosen point process models

Gibbs models are known to be somewhat deficient when it comes to modelling clustered point processes. In particular, it is difficult to obtain large isolated clusters. To understand why, we can look at the maximum and minimum of the Papangelou conditional intensities. In the following we shall consider statistically homogeneous models, i.e. the case where $\theta_1 S_1(u) = 0$.
For the Geyer process, the conditional intensity is bounded between the values:

$$e^{\theta_0} \leq \lambda_{\boldsymbol{\theta}}(u; \boldsymbol{x}) \leq e^{\theta_0 + \theta_2 s^*}.$$

Recalling that expected number of points in a given subregion A is given by

$$\mathbb{E}N(A) = \mathbb{E}|\boldsymbol{X} \cap A| = \mathbb{E}\left[\int_A \lambda_{\boldsymbol{\theta}}(u; \boldsymbol{X})\,\mathrm{d}u\right],$$

we can see that the expected number of points in a subregion A is bounded by

$$e^{\theta_0}|A| \leq \mathbb{E}N(A) \leq e^{\theta_0+\theta_2 s^*}|A|.$$

If we wish to simulate patterns with isolated clusters, we want there to be regions with very few points, and some regions with a very high number of points. This would correspond to a very low background intensity, e^{θ_0}, and large intensity due to interactions, $e^{\theta_2 s^*}$. At the same time, in simulation studies like this one, it is important to keep the samples generated from a specific model as similar as possible, so we might also want to control the total number of points in the observation window W, say within a window of $N(W) \in [N_{\min}, N_{\max}]$. In other words, it is undesirable if half of the generated point patterns have only a few points, while the other half have 2000 points. However, since $N_{\min} = e^{\theta_0}|W|$ and $N_{\max} = e^{\theta_2 s^*}|W|$, this condition also limits the ability to create isolated clusters. In particular, $N_{\min}$, already determines the background intensity, and thus the number of points in regions which are not clusters. While $N_{\max}$ limits the number the number points in regions within a cluster. To really allow large isolated clusters we need to also allow large fluctuations in the total number of points in the observation window, W.
This behaviour is not limited to the Geyer process, as we have similar bounds for the area interaction process:

$$e^{\theta_0-\theta_2} \leq \lambda_{\boldsymbol{\theta}}(u; \boldsymbol{x}) \leq e^{\theta_0}.$$

While this problem can be mitigated by considering instead Gibbs processes with a fixed number of points, these point processes instead have the problem that the Papangelou conditional intensity can not be defined, and thus the GNZ formula can not be applied, meaning the inference method outlined in section 2.2 can not be used, as well as several of the diagnostic tools used to compare the original point configuration with the reconstructed one (see next sections).

3.2 Partition and reconstruction process

Being interested mainly in the principle of whether we can faithfully reconstruct a point pattern, the shape and dimensions of our observation window is not of immediate importance, and we anticipate that the results and conclusions can be scaled appropriately and applied to arbitrary sizes. Therefore, we will usually use an observation window $W = [0, 1]^2$, i.e. a square with length 1, and area $|W| = 1$ (arbitrary units).
As for the partition into subregions, the relative size and shape of these subregions are likely important factors, but it would be difficult to explore and analyse both of them in a systematic fashion. Therefore, in this thesis we limit ourselves to explore the ability to reconstruct point processes using partitions consisting of equally sized subregions. Thus, the general scheme we adopt here is by partitioning the

observation window into k^2 equally sized quadrats, A_i, $i = 1, \ldots, k^2$, each of size $|W|/k^2 = 1/k^2$ and side length $1/k^2$. By varying the number of partiotions, k^2, we are able to investigate the quality of reconstruction as a function of the size of the quadrats.

3.2.1 Reconstruction process

Starting with an "original" point configuration $\boldsymbol{x}$ over an observation window W, and a partition $W = \bigcup_{i=1}^{k^2} A_i$, this point configuration can then also be split into k^2 separate sub-point configurations, $\boldsymbol{x}_i = \boldsymbol{x} \cap A_i$, $i = 1, \ldots, k^2$. After counting and recording the number of points in each quadrat, $n_i = |\boldsymbol{x}_i|$, the point configuration is then reconstructed by first distributing n_i points randomly within the corresponding subregion A_i, creating the local point configurations $\tilde{\boldsymbol{x}}_i$ over A_i, which are then combined into the full reconstructed point process $\tilde{\boldsymbol{x}}$ over W. The process is summarised in the diagram below.

$$
\begin{array}{rcl}
W & \longrightarrow & A_1 \cup A_2 \cup \cdots \cup A_{k^2} \\
\hline
\boldsymbol{x} & \longrightarrow & \boldsymbol{x}_1 \cup \boldsymbol{x}_2 \cup \cdots \cup \boldsymbol{x}_{k^2} \\
 & & \downarrow \quad \downarrow \qquad \downarrow \\
 & & n_1 \quad n_2 \quad \cdots \quad n_{k^2} \\
 & & \downarrow \quad \downarrow \qquad \downarrow \\
\tilde{\boldsymbol{x}} & \longleftarrow & \tilde{\boldsymbol{x}}_1 \cup \tilde{\boldsymbol{x}}_2 \cup \cdots \cup \tilde{\boldsymbol{x}}_{k^2}
\end{array}
$$

It remains to discuss the manner in which the n_i points are distributed over A_i to produce $\tilde{\boldsymbol{x}}_i$. The most ignorant, or non-informative, manner is to distribute the points using a (homogeneous) binomial point process,

$$\tilde{\boldsymbol{x}}_i \sim \text{BinomialPointProcess}(n_i, A_i).$$

On the other hand, if we have reason to suspect that the point process intensity depends on some known covariate density, e.g. population density, we could use this information to obtain a better reconstruction. In this case we might use an inhomogeneous binomial point process,

$$\tilde{\boldsymbol{x}}_i \sim \text{BinomialPointProcess}(n_i, A_i; f),$$

where $f(u)$ is the covariate density on which the intensity is assumed to depend on. Finally, we might have some knowledge about the interaction of the point process. In the context of disease spread we might have a rough idea about the radius of infection, or we might have plausible models (and parameters) from a previous analysis in some other region or of a disease with similar transmission properties. If the model is assumed to be a Geyer process, or a an area interaction process, or any other type of Gibbs process, one might get better results by reconstructing the point process over each region using a Gibbs process with fixed number of points:

$$\tilde{\boldsymbol{x}}_i \sim \text{GibbsProcess}(n_i, A_i; \tilde{\theta}).$$

However, care should be taken when reconstructing an interacting point process over each subregion individually. Since the probability of points being in a given location depends on the location of the other points (within an interaction radius) the points in the different subregions need to be aware of the points in neighbouring subregions. This can be achieved by involving all the subregions in an MCMC simulation. In this thesis this is achieved by first initialising each subregion using a binomial point process, and then in each step of the birth+death MCMC scheme described in section 2.1.1.4 first choosing the subregion to be updated randomly, and then to delete (death) and add (birth) a point within the subregion, conditional on the points in the entire observation window.
Of course, including too much information (and assumptions) in the reconstruction process might distort point configuration more than a non-informative reconstruction would. Even worse, if the goal is to use the reconstructed point configuration for parameter inference and predictions, these may be influenced by the information provided in the reconstruction process and thus introduce an artificial bias.

3.3 Measures of quality of reconstruction

To evaluate whether the point configuration $\tilde{\boldsymbol{x}}$ is a faithful reconstruction of the original point configuration $\boldsymbol{x}$, i.e. if they are sufficiently similar, we must first specify in which sense this is meant. In the following we shall distinguish two notions of similarity, or rather dissimilarity.
The first notion, which we shall call *direct* or *non-statistical* dissimilarity, concerns the locations of the points, in the absence of any underlying statistical model. Here we simply ask how much the points have been moved around. This approach does not require any knowledge of the model that was used to generate the point configuration, but also does not really compare the statistical properties of the two point processes. Depending on the purpose of the reconstruction process this may not necessarily be what we want. Instead one might be more interested in producing a reconstructed point configuration that shares the same statistical properties but does not necessarily have point locations close to the original point configuration. Using a specific example, consider an original point pattern $\boldsymbol{x}$ which is a realisation of a homogeneous Poisson point process. Reconstructing this point configuration using a binomial point process $\tilde{\boldsymbol{x}}_i \sim \text{BinomialPointProcess}(n_i, A_i)$ for each quadrat A_i, the resulting point configuration $\tilde{\boldsymbol{x}} = \bigcup_{i=1}^{k^2} \tilde{\boldsymbol{x}}_i$ will again be a Poisson point process, with the same intensity. Then, in a statistical sense, the point configurations $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ can both be considered samples from the same point process, and thus share the same statistical properties. Nevertheless, their point locations may not necessarily be close to each other, and thus the two point configurations may have a large direct dissimilarity.
Therefore, we also consider a notion of *statistical dissimilarity*, where the interest lies rather in whether the two point configuration share the same statistical properties. The diagnostic tools we shall use for this notion of dissimilarity typically utilise some knowledge of the model used to generate the original point configuration, typically in the form of the Papangelou conditional intensity. Most of these measures have previously been used for diagnostics, model selection and sometimes even for

parameter inference. In this thesis however, we intend to use them as a means to compare the original and reconstructed point configurations.
In the following sections we describe the different measures used in the two types of dissimilarity.

3.3.1 Direct (non-statistical) dissimilarity

3.3.1.1 Distance between point configurations

To determine how close, or similar, two point configurations are, we need a mesure of similarity, or disimilarity. One such measure, is the Earth Mover Distance (EMD) which, given two point configurations $\boldsymbol{x}_n$, $\boldsymbol{y}_n$, consisting of an equal number of points, n, is defined by

$$\mathrm{EMD}(\boldsymbol{x}_n, \boldsymbol{y}_n) = \min_{\phi \in P_n} \sum_{i=1}^{n} \|x_i - y_{\phi(i)}\|,$$

where the bijections $\phi : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$ are elements of the permutation group P_n of n elements. In other words, it is the minimum sum of pairwise euclidean distances over all possible complete pairings between two point configurations. One may think about it as the minimum total distance, or cost, required to move the points in $\boldsymbol{x}_n$ to the locations of the points in $\boldsymbol{y}_n$. Indeed, the earth mover's distance is a solution to the so called *assignment problem*, where in it's simplest form, we have a set of n tasks that can be performed by n workers. The workers may be differently skilled at different tasks, and thus there is a cost $C_{i,j}$ when worker i performs task j. Every task must be completed and each worker can only complete one task. The problem then consists in matching the task to the workers in an optimal way, i.e. at the lowest cost. This lowest cost is given by $\min_{\phi \in P_n} \sum_{i=1}^{n} C_{i,\phi(i)}$. While for large n it would be a daunting task to go through all $n!$ permutations, fortunately, there exist more clever algorithms. In this thesis we shall use an implementation of the *Hungarian algorithm*, which solves this problem in polynomial time [22, 23].

3.3.2 Statistical dissimilarity

To compare the statistical properties of the two point processes we shall make use of three different methods.
The first, *Stoyan-Grabarnik residuals* [24], is a type of *innovation*, or *point process residual* [25]. It makes use of the GNZ-formula to obtain an expression which has zero expectation, provided that the point configuration is sampled from the original point process. In this sense, one might say that it measures whether the point configuration corresponds to a typical realisation of the original point process. The next section is devoted to describing this approach in more detail.
In the second approach, we use the parameter inference method outlined in 2.2.1 for both point configurations, and compare estimates between the two. If the purpose of reconstruction is to create a point configuration based on aggregated counts which can then be used to infer the parameters of the underlying point process, then this could be seen as the most relevant method. If the parameter estimates, $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}$ are

sufficiently similar when using $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ one could conclude that the reconstructed point configuration is sufficient in terms of parameter estimation.
On the other hand, we might not necessarily be interested in the parameters themselves, but rather in the conditional intensity, λ, which characterises the point process. Two quite distinct sets of parameters $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ could potentially result in very similar conditional densities, $\lambda_{\boldsymbol{\theta}}$ and $\lambda_{\tilde{\boldsymbol{\theta}}}$. For this reason, we also consider a third approach, the *Mean Integrated Square Error* (MISE):

$$\mathrm{MISE}_{\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}(\boldsymbol{x},\tilde{\boldsymbol{x}}) = \int_W \left(\lambda_{\boldsymbol{\theta}}(u;\boldsymbol{x}) - \lambda_{\tilde{\boldsymbol{\theta}}}(u;\tilde{\boldsymbol{x}})\right)^2 \mathrm{d}u.$$

To conclude this section, we go into some more detail about the Stoyan-Grabranik residual.

3.3.2.1 Stoyan-Grabarnik residuals

The GNZ-formula can be used to define innovations [25]:

$$I_h(A;\boldsymbol{X}) = \sum_{x\in\boldsymbol{X}\cap A} h(x,\boldsymbol{X}\setminus\{x\}) - \int_A h(u;\boldsymbol{X})\lambda(u;\boldsymbol{X})\,\mathrm{d}u,$$

the expectation of which is 0 if $\boldsymbol{X}$ is a point process with Papangelou conditional intensity $\lambda(u;\boldsymbol{X})$, which we may sometimes write symbolically as $\boldsymbol{X}\sim\lambda$. A particularly simple form is obtained when using $h = 1/\lambda$:

$$R(A;\boldsymbol{X}) = I_{\frac{1}{\lambda}}(A;\boldsymbol{X}) = \sum_{x\in\boldsymbol{X}\cap A} \frac{1}{\lambda(x;\boldsymbol{X}\setminus\{x\})} - |A|,$$

provided A is the support of λ. This particular choice of innovation was first considered by Stoyan and Grabarnik [24] as a diagnostic tool, which is why we shall refer to this innovation as the Stoyan-Grabarnik residual.
The Stoyan-Grabarnik residuals has a nice interpretation if one defines the "mass" function $m(x;\boldsymbol{X}) = 1/\lambda(x;\boldsymbol{X}\setminus\{x\})$. Then we the sum of masses in a region A is given by $M_{\boldsymbol{X}}(A) = \sum_{x\in\boldsymbol{X}\cap A} m(x;\boldsymbol{X})$ and has expectation $|A|$. Thus we can write the Stoyan-Grabarnik residual as:

$$R(A;\boldsymbol{X}) = M_{\boldsymbol{X}}(A) - \mathbb{E}[M_{\boldsymbol{X}}(A)]$$

The interpretation of $M_{\boldsymbol{x}}(A)$ is that of a weighted counting measure. Each point in A is weighted according to the inverse of the conditional intensity. Points with high conditional intensity (under the model and conditioned on $\boldsymbol{x}$) will have low weights, while points with very low conditional intensity will have large weights. Thus if A contains a large number of points with unlikely positions, the weighted counting measure will return a large value. On the other hand, too small a number of points in A will automatically lead to a low value of $M_{\boldsymbol{x}}(A)$, thus we subtract the expected weighted point count $\mathbb{E}[M_{\boldsymbol{X}}(A)] = |A|$. This way, (strongly) negative values of $R_\theta(A;\boldsymbol{x})$ is likely due to there being not enough points (according to the model) in A, whereas (strongly) positive values are likely due to there being too many points and/or that they have unlikely positions.

Furthermore, while the diagnostic can be evaluated over the whole window W to obtain an overall score, one could also evaluate it separately over the different regions A_i, thus allowing to identify outlier regions. Even more local, one can mark the points $x \in \boldsymbol{x}$ of the pattern by their weights $m(x; \boldsymbol{x})$ and display these in a plot to identify outlier points.

4

Results

Before presenting the results from the simulation studies, we shall present some theoretical considerations that will help us to better interpret and analyse the results. In the following section, we recall the reconstruction process presented in section 3.2, and analyse it theoretically. In section 4.2 we consider the special case when the original point process is a Poisson point process, to be used as a benchmark, while in section 4.3 we then present the results from the simulation studies for interacting (i.e. clustering) point processes. We then add inhomogeneity in section 4.4, and then point processes on linear networks in section 4.5.

4.1 Theoretical analysis of reconstruction process

Following the discussion in 3.2 we shall only consider the restricted point process $\boldsymbol{X}_W = \boldsymbol{X} \cap W$ over the observation window $W = [0,1]^2$. The observation window is then partitioned into disjoint quadrats: $W = \bigcup_{i=1}^{k^2} A_i$, where each quadrat has the measure $|A_i| = 1/k^2$. The partition also splits the point process

$$\boldsymbol{X}_W = \bigcup_{i=1}^{k^2} \boldsymbol{X}_i, \qquad \text{where } \boldsymbol{X}_i = \boldsymbol{X} \cap A_i.$$

We shall denote the number of points in each subregion A_i, by $N_i = |\boldsymbol{X}_i|$. The multivariate distribution

$$p_{n_1,\dots,n_{k^2}} = \mathbb{P}(N_1 = n_1, N_2 = n_2, \dots, N_{k^2} = n_{k^2}),$$

is typically intractable analytically. Indeed, even the marginal distribution $p_{n_i} = \mathbb{P}(N_i = n_i)$ can not typically be obtained analytically. Nevertheless, in section 4.3 we produce some arguments that the distribution may factorise for large quadrats, and in section 4.3.1.1 we investigate these claims via simulations.
Based on the quadrat counts $(N_1, \dots, N_{k^2})$, we generate new point processes

$$\tilde{\boldsymbol{X}}_i \mid N_i \sim \text{BinomialProcess}(N_i, A_i)$$

over the quadrats A_i. The **reconstructed point process** over W is given by

$$\tilde{\boldsymbol{X}}_W = \bigcup_{i=1}^{k^2} \tilde{\boldsymbol{X}}_i.$$

Given a Janossy density $j_{n_i}(\tilde{\boldsymbol{x}}_i)$, the conditional probability density (i.e. the location density), given $N_i = n_i$, is given by

$$p(\tilde{\boldsymbol{x}}_i \mid n_i) = n_i! j_{n_i}(\tilde{\boldsymbol{x}}_i),$$

and the joint pdf $p(\tilde{\boldsymbol{x}}_i, n_i) \equiv p(\tilde{\boldsymbol{x}}_i \cap \{N_i = n_i\})$ is given by

$$p(\tilde{\boldsymbol{x}}_i, n_i) = p(\tilde{\boldsymbol{x}}_i \mid n_i)\mathbb{P}(N_i = n_i) = n_i! p_{n_i} j_{n_i}(\tilde{\boldsymbol{x}}_i).$$

The full reconstructed point process $\tilde{\boldsymbol{X}}_W$ is determined by $\{\tilde{\boldsymbol{X}}_i\}_{i=1}^{k^2}$ which are independent conditional on $(N_i)_{i=1}^{k^2}$. Thus

$$\tilde{\boldsymbol{X}}_W \mid N_1, \ldots, N_{k^2} \overset{d}{=} (\tilde{\boldsymbol{X}}_i \mid N_i)_{i=1}^{k^2},$$

with conditional pdf

$$p(\tilde{\boldsymbol{x}}_W \mid N_1, \ldots, N_{k^2}) = \prod_{i=1}^{k^2} p(\tilde{\boldsymbol{x}}_i \mid N_i),$$

and the joint pdf is then given by

$$\begin{aligned} p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) &= \left(\prod_{i=1}^{k^2} p(\tilde{\boldsymbol{x}}_i \mid n_i) \right) \mathbb{P}(N_1 = n_1, \ldots, N_{k^2} = n_{k^2}) \\ &= \left(\prod_{i=1}^{k^2} p(\tilde{\boldsymbol{x}}_i \mid n_i) \right) \cdot p_{n_1, \ldots, n_{k^2}} \end{aligned} .$$

Inserting the Janossy density of a binomial point process, $j_{n_i}(\tilde{\boldsymbol{x}}_i) = 1/|A_i|^{n_i}$, by recalling that $p(\tilde{\boldsymbol{x}}_i \mid n_i) = n_i! j_{n_i}(\tilde{\boldsymbol{x}}_i)$ we get

$$p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_k) = \left(\prod_{i=1}^{k^2} \frac{n_i!}{|A_i|^{n_i}} \right) p_{n_1, \ldots, n_{k^2}}.$$

While the probability density appears to be independent of the locations of the points, this is not quite true, since the counts n_i determine the number of points in each quadrat, A_i.

Small quadrat limit

As the quadrat size becomes smaller, the part of the configuration space of $\boldsymbol{X}_i$ and $\tilde{\boldsymbol{X}}_i$ with any significant probability mass becomes smaller. This is both due to the reduction of the target space A_i, but also due to the fact that the small size reduces the probability that the quadrat contains multiple points. At some point, most quadrats will be empty, while others contain 1 point. In section 4.2 we show that the expected distance between a randomly displaced point, and its randomly positioned original, is proportional to the side length a of the quadrat A_i. Thus in the limit when $|A_i| \to 0$, we expect that $p(\tilde{\boldsymbol{x}}_i \mid n_i) \to p(\boldsymbol{x}_i \mid n_i)$, and we trivially have $p(\tilde{\boldsymbol{x}}_W) = p(\boldsymbol{x}_W)$, i.e. the original- and reconstructed point processes have the same statistical distribution.

Large quadrat limit: Poisson approximation

If $N_1, \ldots, N_{k^2}$ are **approximately independent**, such that $p_{n_1,\ldots,n_{k^2}} \approx p_{n_1} \cdots p_{n_{k^2}}$, then

$$p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) = \prod_{i=1}^{k^2} \Big[p(\tilde{\boldsymbol{x}}_i \mid n_i) p_{n_i} \Big] = \prod_{i=1}^{k^2} p(\tilde{\boldsymbol{x}}_i, n_i).$$

If furthermore the quadrat counts approximately follow a Poisson distribution, with some intensity Λ_i, such that $p_{n_i} = \Lambda_i^{n_i} e^{-\Lambda_i}/n_i!$, then

$$p(\tilde{\boldsymbol{x}}_i, n_i) = n_i! p_{n_i} j_{n_i}(\tilde{\boldsymbol{x}}_i) = n_i! \cdot \Lambda_i^{n_i} \frac{e^{-\lambda_i}}{n_i!} \cdot \frac{1}{|A_i|^{n_i}} = \left(\frac{\Lambda_i}{|A_i|} \right)^{n_i} e^{-\lambda_i},$$

setting $\Lambda_i = \lambda_i |A_i|$, we have

$$p(\tilde{\boldsymbol{x}}_i, n_i) = \lambda_i^{n_i} e^{-\lambda_i |A_i|},$$

and finally obtain

$$p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) = \prod_{i=1}^{k^2} \lambda_i^{n_i} e^{-\lambda_i |A_i|}, \tag{4.1}$$

which is simply a set of independent Poisson point processes for each quadrat (not surprisingly). The Papangelou conditional intensity is simply given by

$$\lambda(u; \tilde{\boldsymbol{x}}_W) = \sum_{i=1}^{k^2} \mathbb{I}_{u \in A_i} \frac{p(\tilde{\boldsymbol{x}}_W \cup \{u\}, n_1, \ldots, n_i + 1, \ldots, n_{k^2})}{p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_i, \ldots, n_{k^2})} = \sum_{i=1}^{k^2} \mathbb{I}_{u \in A_i} \lambda_i, \tag{4.2}$$

i.e. a piecewise constant function.
If the intensities are also equal (which should be the case if the original point process was homogeneous), then

$$p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) = \lambda^{\sum_{i=1}^{k^2} n_i} e^{-\lambda \sum_{i=1}^{k^2} |A_i|} = \lambda^{n_W} e^{-\lambda |W|},$$

where $n_W = n_1 + \ldots + n_{k^2} = |\boldsymbol{x}_W|$ is the total number of points of $\boldsymbol{X}$ in W. Thus $p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) = p(\tilde{\boldsymbol{x}}_W, n_W)$ is the pdf of the usual Poisson point process, i.e. $\tilde{\boldsymbol{x}}_W \sim \text{PoissonProcess}(\lambda)$.
The same results as above can be obtained by approximating $(N_1, \ldots, N_{k^2}) \mid (N_1 + \ldots + N_{k^2} = N_W) \sim \text{Multinomial}(q_1, \ldots, q_{k^2}; N_W)$, i.e. given a fixed total number of points in W, each point is independently distributed among the k^2 quadrats with probabilities q_i, $i = 1, \ldots, k^2$. The probability mass function of $N_1, \ldots, N_{k^2}$ then becomes

$$p_{n_1,\ldots,n_{k^2}} = \frac{n_W!}{n_1! \cdots n_{k^2}!} \left(\prod_{i=1}^{k^2} q_i^{n_i} \right) \cdot \mathbb{P}(N_W = n_W),$$

and the probability density of $\tilde{\boldsymbol{X}}_W$ becomes:

$$p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) = n_W! \, \mathbb{P}(N_W = n_W) \prod_{i=1}^{k^2} \left(\frac{q_i}{|A_i|} \right)^{n_i}.$$

If the total number of points, N_W, in W, is approximately Poisson distributed, $\mathbb{P}(N_W = n_W) \approx \alpha^{n_W} e^{-\alpha}/n_W!$, and we define $\lambda_i = \alpha\, q_i/|A_i|$ we get again

$$p(\tilde{\boldsymbol{x}}_W, n_1, \ldots, n_{k^2}) = \prod_{i=1}^{k^2} \lambda_i e^{-\lambda_i |A_i|},$$

where we used that $\sum_i q_i = 1$ and thus $\alpha = \sum_i \alpha\, q_i = \sum_i \lambda_i |A_i|$.

4.2 Reconstructing a Poisson point process

To start off our analysis of, we can choose to explore the properties of a Poisson point process under reconstruction. Although this is an exceptionally simple problem, and therefore may seem uninteresting, it may act as a benchmark for our analysis of more interesting point processes. The problem becomes simple since the reconstructed point process $\tilde{\boldsymbol{X}}$ obtained via the reconstruction process described in 3.2.1 from the original point process $\boldsymbol{X} \sim \text{PoissonPointProcess}(\lambda)$, will also be a poisson point process, with the same intensity λ, i.e. $\tilde{\boldsymbol{X}} \sim \text{PoissonPointProcess}(\lambda)$. This can be seen as a result of the displacement theorem [26], but is also shown explicitly in this setting in section 4.1.
When both the original and reconstructed point configurations $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ are samples from the same distribution we expect them to have the same statistical properties. Indeed, the Stoyan-Grabarnik residual for a Poisson point process becomes simply

$$R(W; \boldsymbol{x}) = \sum_{x \in \boldsymbol{x} \cap W} \frac{1}{\lambda} - |W| = \frac{|\boldsymbol{x} \cap W|}{\lambda} - |W|,$$

and can easily be checked to have expectation 0 (using $\mathbb{E}|X \cap W| = \lambda |W|$). More importantly, since $|\boldsymbol{x} \cap W| = |\tilde{\boldsymbol{x}} \cap W| = N_W$, the original and reconstructed point patterns will always have the same value for the Stoyan-Grabarnik residual.
Furthermore, for a Poisson point process, parameter inference is simple, and one can directly use maximum likelihood estimates: $\lambda_{\text{ML}} = |\boldsymbol{x} \cap W|/|W|$, which again will lead to the same estimates for the original and reconstructed point configurations. Similarly, the MISE scores will also automatically be the same for both point configurations, and thus all of our measures of statistical dissimilarity will be 0.

More interesting is the direct (or non-statistical) dissimilarity, specifically the EMD distance between $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$. While the two point configurations are statistically the same, the points are generally in different locations, but how different? And does the distance depend on the resolution of the reconstruction process, i.e. the number, k^2, of subregions, or quadrats. Equivalently, we might ask if the EMD distance depends on the side length $a = 1/k$ of the quadrats.
To answer this question, it is useful to first consider the case of very small quadrats, $a^2 \ll |W|$. Let's suppose that such a small quadrat, $A = [0, a]^2$, contains only one point, x, which is located at a random (uniform) position in the small quadrat. Then, under the reconstruction method, we redistribute this point again randomly

(uniformly) over the quadrat, resulting in $\tilde{x}$. The question then becomes, what is the expected distance between two points distributed at random (uniform) over a square of side a, i.e.

$$\begin{aligned}\mathbb{E}\|x-\tilde{x}\| &= \int_A\int_A \|x-\tilde{x}\|\,\mathrm{d}^2x\,\mathrm{d}^2\tilde{x}\\ &= \int_0^a\int_0^a\int_0^a\int_0^a \sqrt{(x_1-\tilde{x}_1)^2+(x_2-\tilde{x}_2)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}\tilde{x}_1\,\mathrm{d}\tilde{x}_2\\ &= \frac{2+\sqrt{2}+5\log(\sqrt{2}+1)}{15}\cdot a \approx 0.52\,a.\end{aligned}$$

The computation of the integral is arduous, and the precise computation shall not be presented in this thesis. Instead, we simply note that in this limit, the EMD score becomes

$$\mathrm{EMD}(\boldsymbol{x}_n,\tilde{\boldsymbol{x}}_n) \approx n\cdot 0.52\,a, \qquad a \lesssim \frac{1}{\sqrt{n}}, \tag{4.3}$$

where we added a more informative condition: if the quadrat has side length $a = 1/\sqrt{n}$, then $na^2 = 1$, i.e. there is on average one point per quadrat. Thus we expect this relationship to hold approximately whenever $a \lesssim 1/\sqrt{n}$ (or $a \lesssim \sqrt{|W|/n}$ for arbitrary window size).
In the other limit, when $a \approx 1$, i.e. when the quadrat covers the entire window, the reconstructed point process is a new Poisson point process over the same observation window. Thus the question becomes, what is the expected earth mover's distance between two poisson point processes (or rather binomial point process) with an equal number of points? While we have not found a result about this in the literature, and have not managed to derive an expression ourselves, we can identify a lower bound. In the best case scenario (in the sense of yielding the lowest EMD), each point $x \in \boldsymbol{x}$ has it's optimal partner $\tilde{x}$ as the closest neighbour among the set $\tilde{\boldsymbol{x}}$. Thus, in this optimal case, we want to find the expected distance $\min_{\tilde{x}\in\tilde{\boldsymbol{x}}}\|x-\tilde{x}\|$ for a given points x and it's nearest neighbour $\tilde{x} \in \tilde{\boldsymbol{x}}$. To this end, we can consider the void probability:

$$\mathbb{P}(\tilde{\boldsymbol{x}} \cap b_r(x) = \emptyset) = e^{-\lambda|b_r(x)|} = e^{-\lambda\pi r^2}.$$

That is, the probability that there are no points in $\tilde{\boldsymbol{x}}$ within a radius r. Thus, the probability that there is at least one (the nearest neighbour) is given by

$$1-\mathbb{P}(\tilde{\boldsymbol{x}} \cap b_r(x) = \emptyset) = 1-e^{-\lambda\pi r^2} = F(r),$$

where $F(r)$ is the cumulative distribution function of the random variable corresponding to the nearest neighbour distance. The probability density function is then given by $f(r) = \frac{\mathrm{d}F(r)}{\mathrm{d}r} = 2\pi\lambda r e^{-\lambda\pi r^2}$ and the expected distance is

$$\mathbb{E}\left[\min_{\tilde{x}\in\tilde{\boldsymbol{x}}}\|x-\tilde{x}\|\right] = \int_0^\infty r f(r)\,\mathrm{d}r = \frac{1}{2\sqrt{\lambda}},$$

which can be approximated as $1/2\sqrt{n}$, where $n = |\boldsymbol{x}| = |\tilde{\boldsymbol{x}}|$. Thus, a lower bound for the earth mover's distance is given by $\mathrm{EMD}(\boldsymbol{x}_n,\tilde{\boldsymbol{x}}_n) > n\cdot 1/2\sqrt{n} = \sqrt{n}/2$. Of course, this optimal situation is unlikely to occur. Oftentimes there will be multiple points

in $\boldsymbol{x}$, which have the same closest neighbour in $\tilde{\boldsymbol{x}}$), but this neighbour can only be assigned to one of them. For point processes with a medium intensity $50 < \lambda < 500$ we find that multiplying the lower bound by a factor of 2 yields the approximate value of the EMD. Therefore, we use as an approximation in the large quadrat limit:

$$\mathrm{EMD}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \approx \sqrt{n}, \qquad a \approx 1. \tag{4.4}$$

The results from simulations, along with these approximations, are plotted in figure 4.1.

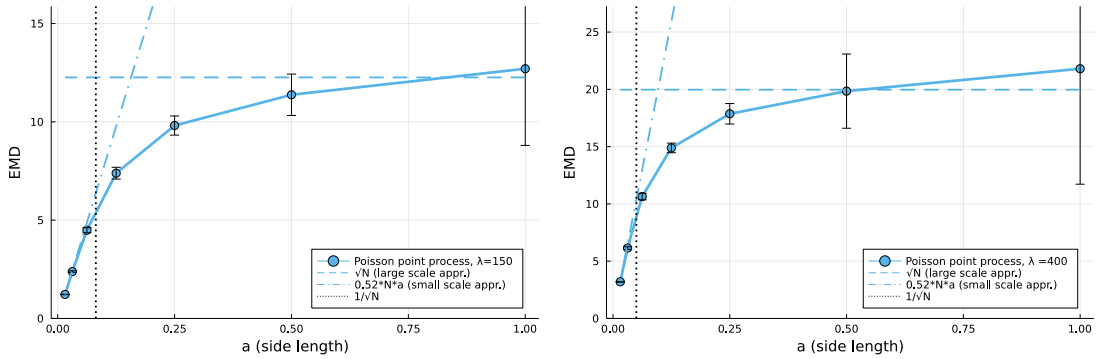


Figure 4.1: The mean (solid lines), and variance (errorbars) of the EMD for two Poisson point processes with $\lambda = 150$ (left) and $\lambda = 400$ (right). The EMD was computed for 200 pairs of original and reconstructed point configurations. Each original was reconstructed at different resolutions k, or equivalently, quadrat side lengths, $a = 1/k$. Also depicted are the small and large quadrat approximations, as well as a vertical dividing line $1/\sqrt{n}$ which determines when the small quadrat approximation is expected to be valid.

4.3 Reconstructing homogeneous interacting point processes

We now turn to our interacting models, more specifically the Geyer model and the area interaction model. We expect the analysis in the previous section on the earth mover distance to still be approximately valid. In the small quadrat size limit, $a \lesssim 1/\sqrt{n}$, the same argument holds, and the argument in the large quadrat limit $a \approx 1$, should also be somewhat valid since the reconstructed point pattern, $\tilde{\boldsymbol{x}}$, in this limit is still a sample from a Poisson point process, with intensity, $\lambda = |\boldsymbol{x}|/|W|$. Thus, for any given point in $\boldsymbol{x}$ (which is not a poisson point process), the nearest neighbour is still determined by the void distribution of the reconstructed (Poisson) point process.
As for the statistical dissimilarity measures, we expect that in the very small quadrat limit $a < 1/\sqrt{n}$, and $a < R$, where R is the interaction distance, the reconstructed point process will retain most of the statistical properties of the original point process. This is simply a consequence of the quadrats being small enough so that the points cannot be moved significantly, in other words the same reason as why the

EMD tends to zero in this limit. Similarly, we expect in the large quadrat limit, $a \approx 1$, the statistical properties will approach those of a Poisson point process.

In the intermediate regime, the deviation from the Poisson approximation depends on the deviation from the independence of the counts $N_1, \ldots, N_{k^2}$, i.e. on $p_{n_1,\ldots,n_{k^2}} = \mathbb{P}(N_1 = n_1, \ldots, N_{k^2} = n_{k^2})$. For pairwise interactions we can split the interaction potential into contributions from *contained interactions* and *exterior interactions*:

$$\sum_{\substack{x,y\in\boldsymbol{x}\\ x\neq y}} \phi_2(x,y) = \underbrace{\sum_{i=1}^{k^2} \sum_{\substack{x,y\in\boldsymbol{x}_i\\ x\neq y}} \phi_2(x,y)}_{\text{contained interactions}} + \underbrace{\sum_{i=1}^{k^2}\sum_{\substack{j=1\\ j\neq i}}^{k^2} \sum_{\substack{x\in\boldsymbol{x}_i\\ y\in\boldsymbol{x}_j}} \phi_2(x,y)}_{\text{exterior interactions}}.$$

Here, the contained interactions are between points within the same quadrat, whereas the exterior interactions are between points in different quadrats. It is then reasonable to expect that the relative strength of these two terms depends on the available area for each of them. For instance, given a interaction radius R, only points within a distance R from the border of a quadrat are capable of interacting with points in other quadrats. Likewise, only points that are at least a distance R away from the borders of the quadrats are certain to interact only with points within the same quadrat.
More generally, we can express the Gibbs probability density, $p(\boldsymbol{x}) = p(\bigcup_{i=1}^{k^2} \boldsymbol{x}_i)$ through

$$p(\boldsymbol{x}) = p(\boldsymbol{x}) = \frac{1}{Z} e^{-V(\boldsymbol{x})} = \frac{1}{Z} e^{-\sum_{i=1}^{k^2} V_c(\boldsymbol{x}_i) - \sum_{i=1}^{k^2} V_e(\boldsymbol{x}_i, \boldsymbol{x}\setminus\boldsymbol{x}_i)} = \frac{1}{Z}\prod_{i=1}^{k^2} e^{-V_c(\boldsymbol{x}_i)} e^{-V_e(\boldsymbol{x}_i,\boldsymbol{x}\setminus\boldsymbol{x}_i)},$$

where $V_c(\boldsymbol{x}_i)$ incorporates the activation and local potentials, as well as all the interactions between points within $\boldsymbol{x}_i$, and $V_e(\boldsymbol{x}_i, \boldsymbol{x}\setminus\boldsymbol{x}_i)$ incorporates all the interactions between points in $\boldsymbol{x}_i$ and those in the *exterior*, $\boldsymbol{x}\setminus\boldsymbol{x}_i$. Defining $p_i(\boldsymbol{x}_i) = \frac{1}{Z_i} e^{-V_c(\boldsymbol{x}_i)}$ and $f_i(\boldsymbol{x}) = 1 - \frac{Z_i}{Z} e^{-V_e(\boldsymbol{x}_i,\boldsymbol{x}\setminus\boldsymbol{x}_i)}$ we can write

$$p(\bigcup_{i=1}^{k^2} \boldsymbol{x}_i) = \prod_{i=1}^{k^2} p_i(\boldsymbol{x}_i)\,(1 - f_i(\boldsymbol{x}))\,.$$

In this form, the limit when $V_e(\boldsymbol{x}_i, \boldsymbol{x}\setminus\boldsymbol{x}_i) \to 0$ implies $f_i(\boldsymbol{x}) \to 0$, and thus the factorisation of the probability density in terms of $\boldsymbol{x}_i$. If the contribution to the Gibbs potential from the exterior interactions, $V_e(\boldsymbol{x}_i, \boldsymbol{x}\setminus\boldsymbol{x}_i)$, can be considered small, $f_i(\boldsymbol{x})$ could be interpreted as the lowest nontrivial term in a power series expansion $e^{-V_e(\boldsymbol{x}_i,\boldsymbol{x}\setminus\boldsymbol{x}_i)} \approx 1 - V_e(\boldsymbol{x}_i, \boldsymbol{x}\setminus\boldsymbol{x}_i) + \mathcal{O}(V_e^2)$, i.e. $f_i(\boldsymbol{x}) \approx V_e(\boldsymbol{x}_i, \boldsymbol{x}\setminus\boldsymbol{x}_i)$. From the discussion above, we expect the relative strength of V_e to depend on the available area for these interactions.
From this discussion, we hypothesise that the interdependence of the quadrat counts $N_1, \ldots, N_{k^2}$ is some function of *the fraction of edge/exterior area*, F_e, or inversely, the independence depends on the *fraction of contained area* (see Figure 4.2):

$$F_c = \frac{(a-2R)^2}{a^2}, \qquad F_e = 1 - F_c. \tag{4.5}$$

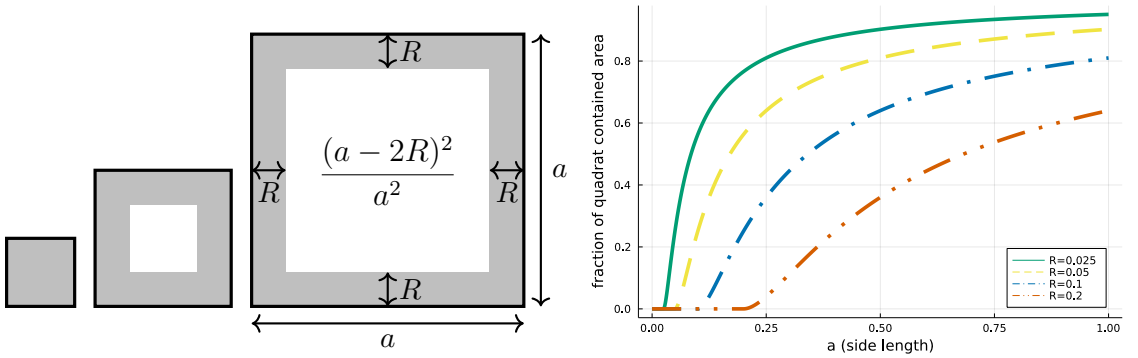


Figure 4.2: Left: Illustration of quadrats with areas within R of its borders marked in gray. It is only within these regions that points are capable of interacting with points in neighbouring quadrats. The three quadrats illustrate the progression from small to large quadrats. Right: Graphs of the fraction of contained area, $F_c = (a-2R)^2/a^2$, with points *not* capable of interacting with neighbouring points as a function of side length a for various values of the interaction range R.

4.3.1 Geyer process

In this subsection we present the results for a Geyer process presented in 3.1. In all figures, 200 original samples of the process was generated, and subsequently reconstructed using the method outlined in 3.2.1, with resolution $k \in \{1, 2, 4, 8, 16, 32, 64\}$. The dissimilarity measures are then plotted against the quadrat side length $a = 1/k$, as well as against the fraction of contained area for the statistical dissimilarities.

Figure 4.3 shows a sample of a Geyer process with $\theta_0 = \log 100$, $\theta_2 = \log 1.25$ as well as the EMD (direct) dissimilarity measure. As expected, the EMD dissimilarity measure between the originals and reconstructions shows a very similar behaviour as the one obtained from reconstructing a Poisson point process. Indeed, the small quadrat ($a \lesssim 1/\sqrt{n_W}$) and large quadrat ($a \approx 1$) approximations appear to be useful even for interacting point processes.

Figures 4.4 and 4.5 show the Stoyan-Grabarnik residuals and parameter estimates for the reconstructed point configurations, $\tilde{\boldsymbol{x}}$, both against the side length $a = 1/k$ for the quadrats, as well as against the fraction of contained area, Eq. 4.5. Again we can clearly identify the two regimes of small quadrats ($a \lesssim 1/\sqrt{n}$) and large quadrats $a \approx 1$. In the limit of very small quadrats, both Stoyan-Grabarnik residuals and parameter estimates approach the values obtained from the original point configurations, whereas in the large quadrat limit they approach values expected for the corresponding Poisson approximation.

As described in the introduction to this section, we expect in the intermediate regime that the deviation from the independent Poisson approximation, is determined by the fraction of contained area. This is verified by the right side figures, in which the Stoyan-Grabarnik residuals as well as the parameter estimates are plotted not against the side length, but rather against the fraction of contained area. The plots reveal an almost linear relationship in this intermediate regime.

In figure 4.6 we also show the same graphs, including a graph of the root mean integrated square error for a more clustered Geyer process, with a lower activation

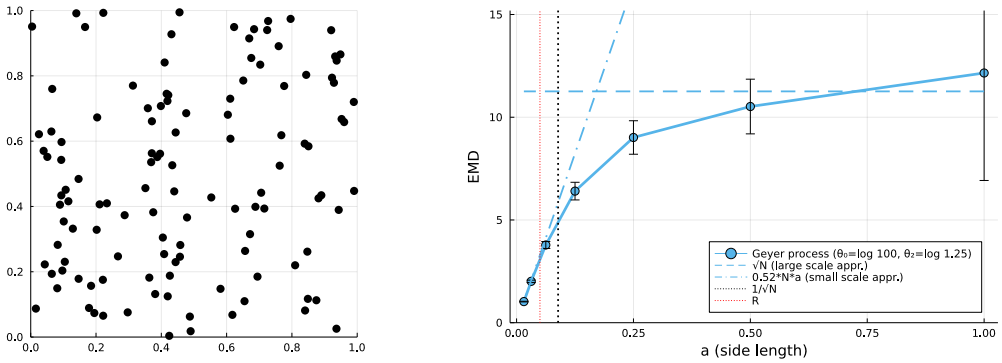


Figure 4.3: Left: Sample point configuration, $\boldsymbol{x}$, of a Geyer process with $\theta_0 = \log 100$, $\theta_2 = \log 1.25$, and nuisance parameters $R = 0.05$ and $s^* = 4$. Right: The mean (solid lines), and variance (errorbars) of the EMD direct dissimilarity measure, each point is computed using 200 original and corresponding reconstructed point configurations. Also displayed are the small quadrat limit, as well as the large quadrat limit (Poisson approximation), as well as two vertical lines indicating possibly important length scales, $a = 1/\sqrt{n}$, and $a = R$.

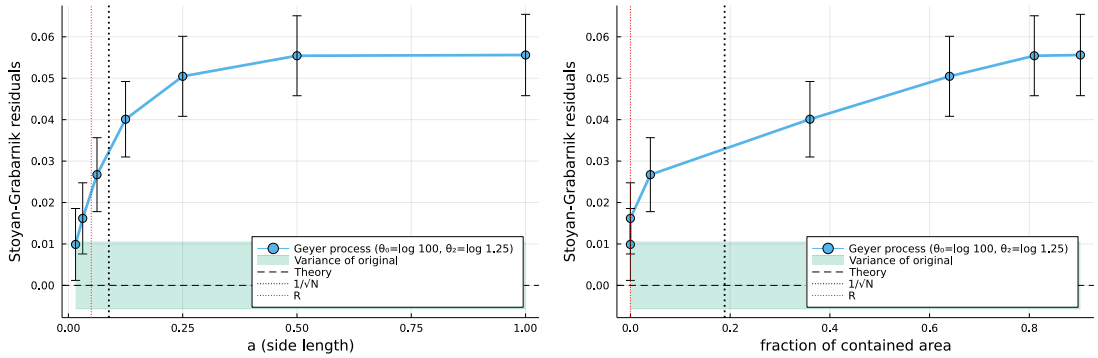


Figure 4.4: The mean (solid lines), and variance (errorbars) of the Stoyan-Grabarnik residuals for a Geyer process with $\theta_0 = \log 100$, $\theta_2 = \log 1.25$, and nuisance parameters $R = 0.05$ and $s^* = 4$. Left: SG-residuals plotted against side length, a of the quadrats. Right: SG-residuals plotted against the fraction of contained area, Eq. 4.5. Also displayed are the theorietical expectation, $= 0$, for the originals (dashed line), as well as a filled region representing the variance obtained form the SG-residuals of the originals. Finally there are two vertical lines indicating possibly important length scales.

$\theta_0 = \log 60$, stronger interaction parameter $\theta_2 = \log 1.35$ as well as larger interaction range, $R = 0.75$ and saturation limit $s^* = 8$. The chosen parameters lead to an increase in the average number of points in the observation window (≈ 500 as compared to ≈ 120). This results primarily in a failure of the large quadrat approximation for the EMD, while the approximations of the statistical dissimilarities still seem to hold up quite well.

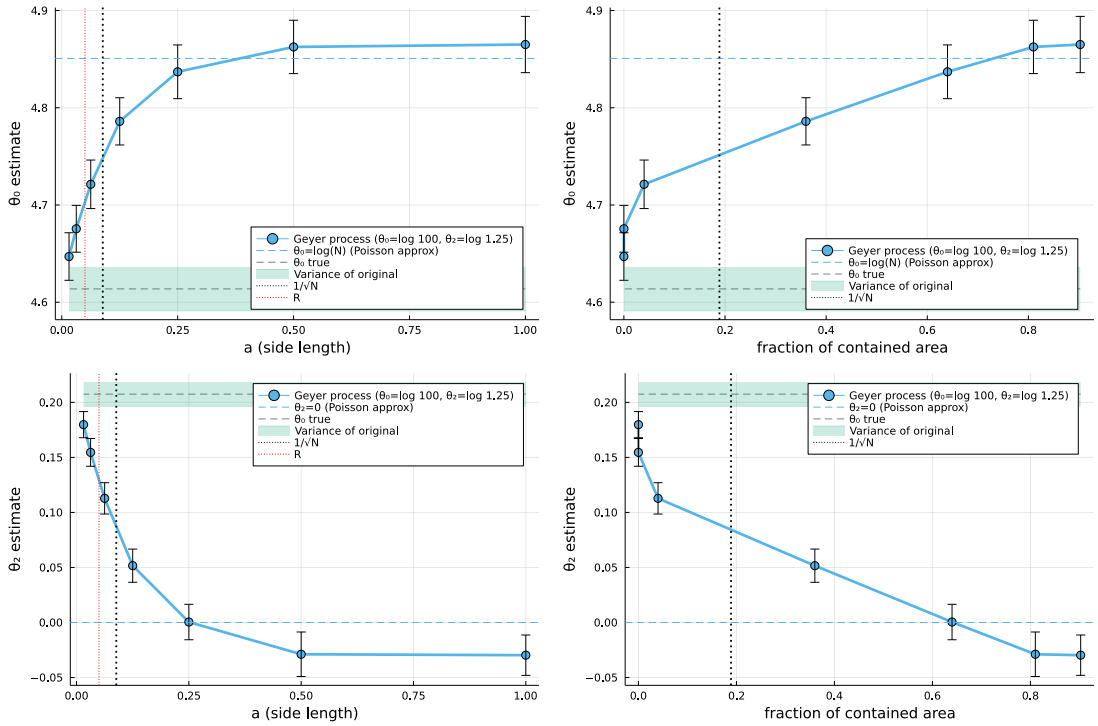


Figure 4.5: Parameter estimates when using the reconstructed point configurations of a Geyer process with $\theta_0 = \log 100$, $\theta_2 = \log 1.25$, and nuisance parameters $R = 0.05$ and $s^* = 4$. Top row: Means (solid lines), and variance (errorbars) of the estimates of parameters θ_0, Bottom row: Means and variance of estimates of θ_2. In the left column the estimates are plotted against side lengths, a, while in the right column the estimates are plotted against the fraction of contained area, Eq. 4.5.

4.3.1.1 Distribution of quadrat counts

In this subsection we want to investigate two of the properties needed for the independent Poisson approximation for large quadrats ($a \approx 1$).

1. Quadrat counts $N_1, \ldots, N_{k^2}$ are approximately independent for large quadrats.
2. The marginal distribution of N_i is approximately Poisson, for large quadrats.

From the discussion in section 4.1 we saw that approximating $p_{n_1,\ldots,n_{k^2}}$ by a multinomial distribution, with equal probabilities for each quadrat, resulted in the Poisson approximation. The independence can then be checked by use of the χ^2-test statistic:

$$X^2 = \sum_{i=1}^{k^2} \frac{(O_i - E_i)^2}{E_i}, \qquad E_i = \frac{1}{k^2}\sum_{i=1}^{k^2} n_i, \qquad O_i = n_i.$$

Under the hypothesis of independent counts, X^2 is a sample of the χ^2 distribution with $k^2 - 1$ degrees of freedom, i.e. $X^2 \sim \chi^2_{k^2-1}$. In figure 4.7 we present the results obtained by drawing 200 samples of a Geyer process, computing the X^2 statistic at different resolutions k, and averaging the results over the sample values, for each k. Since the mean of the χ^2 distribution is $k^2 - 1$, which grows rapidly with k, we divide our test statistic by $k^2 - 1$. Thus, a result of $X^2/(k^2 - 1)$ above 1, means

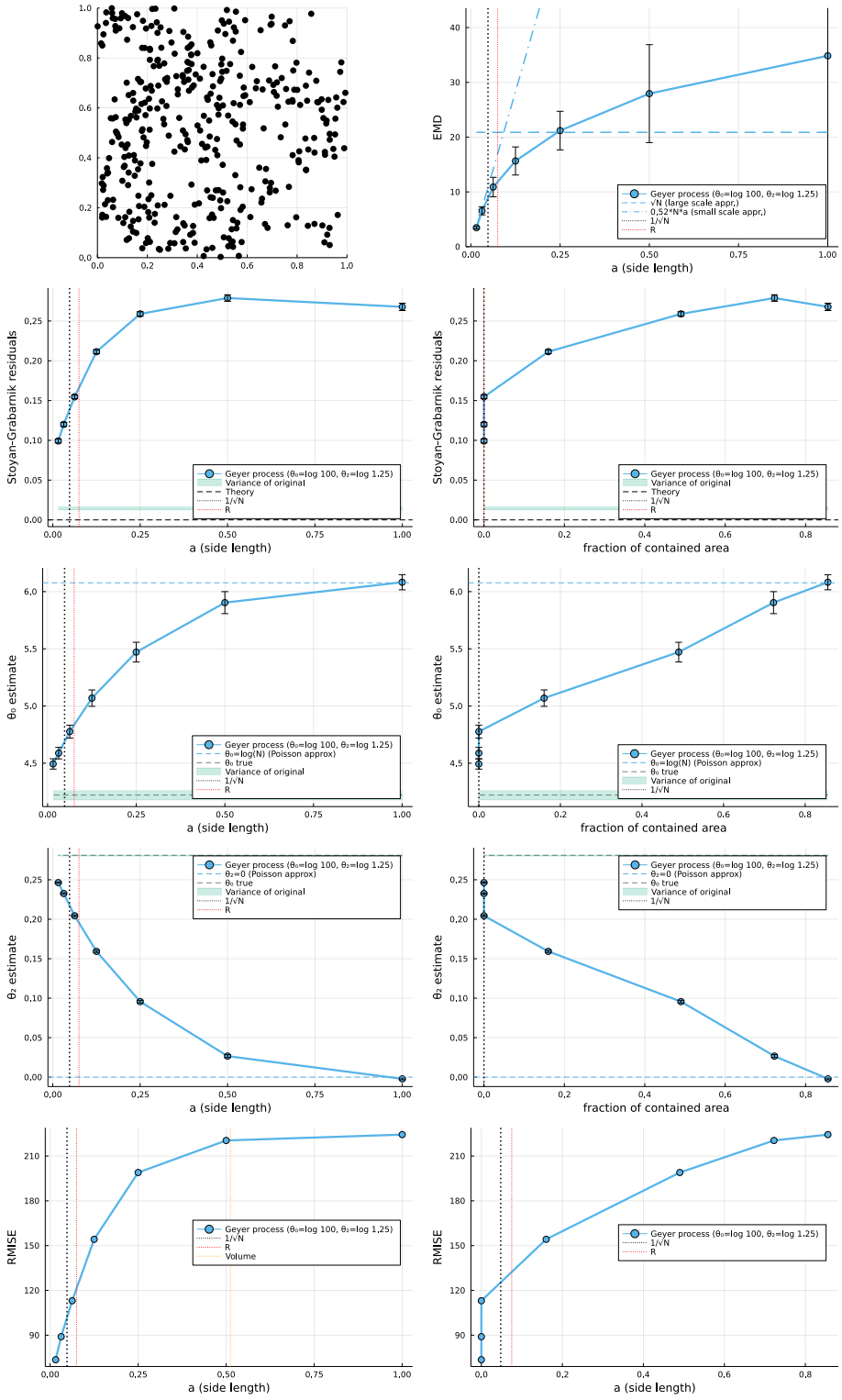


Figure 4.6: Same as in figures 4.3, 4.4, 4.5, but for a Geyer process with $\theta_0 = \log 60$, $\theta_2 = \log 1.35$, $R = 0.075$, $s^* = 8$. Additionally, the root mean integrated square error (RMISE) (last row).

the test statistic is larger than the mean of the χ^2-distribution. To better assess the "extremeness" of our values, we also show the area between the 10th and 90th

quantile of the χ^2 distribution (divided by $k^2 - 1$). These results are then plotted against quadrat side length, a, as well as fraction of contained area F_c, with the expectation that the X^2 values become more and more extreme as the quadrat size is lowered (and thus also the fraction of contained area). Indeed, this what is observed in figure 4.7 as the values of the $X^2/(k^2 - 1)$ statistic goes above the 90th quantile of the χ^2-distribution for small quadrats, and stays solidly inside the 10th and 90th quantile for larger quadrats (although it does not quite approach the mean value, 1, of the χ^2-distribution).

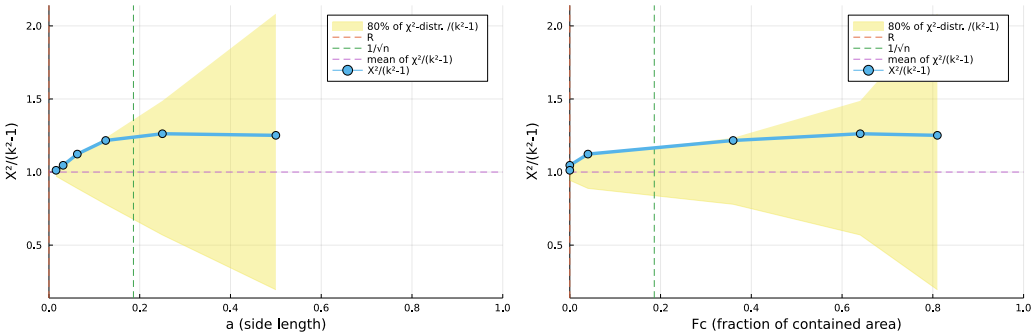


Figure 4.7: The blue line indicates empirical mean of $X^2/(k^2-1)$ for a 200 samples from a Geyer process with $\theta_0 = \log 100$, $\theta_2 = \log 1.25$, and nuisance parameters $R = 0.05$ and $s^* = 4$. Also indicated is the range between the 10th and 90th quantile of the χ^2 distribution (divided by $k^2 - 1$). Left: Plotted against quadrat side length, $a = 1/k$, and Right: plotted against fraction of contained area, F_c. Note that the range between the 10th and 90th quantile divided by the degrees of freedom (and mean), $k^2 - 1$, becomes smaller as the degrees of freedom is increased.

To investigate the validity of the independent Poisson approximation in the large quadrat limit further, we can examine the empirical distribution of quadrat counts over many experiments. In figure 4.8 we display the empirical marginal distributions, $\hat{p}_{n_i}$, of the counts n_i over 200 simulations in the form of a histogram of occurrences. As can be seen, the counts mostly seem to follow a Poisson distribution (even at high resolutions, i.e. high values of k), and in this visualisation it is difficult to observe any major deviation.
To check deviations of the quadrat counts from the Poisson distribution, we can also compute the dispersion index:

$$D = \frac{s^2}{\bar{n}}, \qquad \bar{n} = \frac{1}{k^2}\sum_{i=1}^{k^2} n_i, \qquad s^2 = \frac{1}{k^2 - 1}\sum_{i=1}^{k^2}(n_i - \bar{n})^2,$$

which for a Poisson process should be closes to 1, while for an aggregated/clustering process we expect a value > 1, i over-dispersed, due to a larger variation in quadrat counts (some empty, some containing clusters). However, some algebra readily shows that $D = X^2/(k^2 - 1)$, and thus the precise plot already shown in figure 4.7.
A related measure is given by the Morisita index,

$$M = \sum_{i=1}^{k^2} \frac{n_i(n_i - 1)}{\frac{1}{k^2} n_W(n_W - 1)}, \qquad n_W = \sum_{i=1}^{k^2} n_i, \tag{4.6}$$

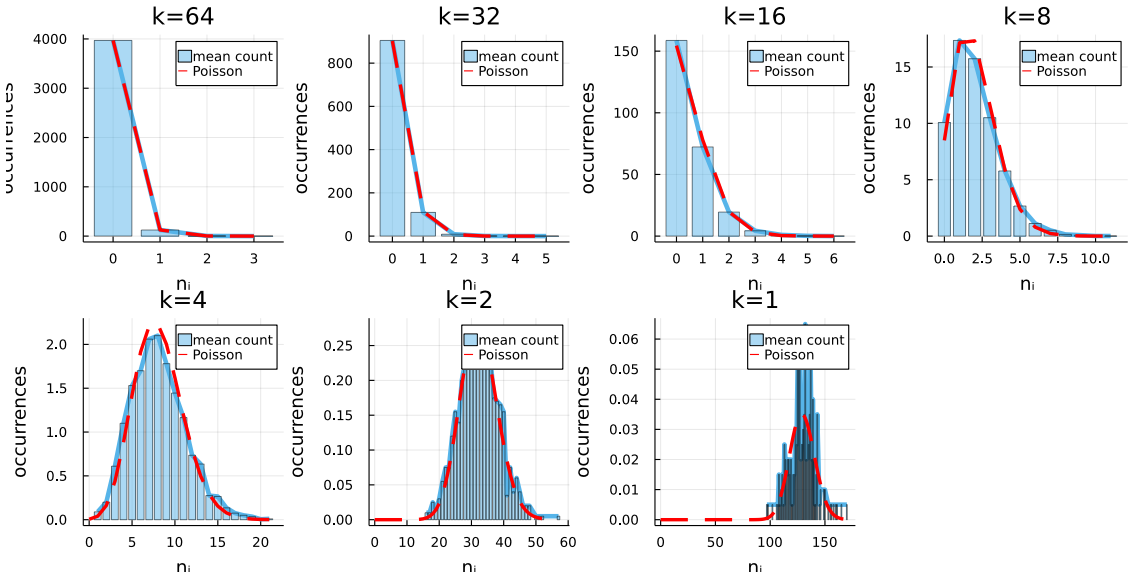


Figure 4.8: Histogram of quadrat counts, n_i, averaged over 200 samples of a Geyer process with $\theta_0 = \log 100$, $\theta_2 = \log 1.25$, and nuisance parameters $R = 0.05$ and $s^* = 4$. Also displayed are the corresponding theoretical distribution of counts from a Poissson point process (red dashed lines) for comparison. From the top left to the bottom right, the plots correspond to a quadrat schemes/partitions with increasingly lower resolution k (where k^2 is the number of quadrats), i.e. increasingly larger quadrats, with sides $a = 1/k$.

which is roughly the sum over the ratios of observed number of pairs of points in each quadrat, $n_i(n_i - 1)$, and the expected number of pairs of point in each quadrat $n_W(n_W - 1)/k^2$. Some (not so simple) algebra shows that this index is a linear transformation of D, depending on both the total number of points, n_W, and the number of quadrats k^2, $D = \frac{k^2}{k^2-1}\left((n_W - 1)M - (\bar{n} - 1)\right)$ so in essence it does not contain any new information compared to X^2 and D, but rather presents the information in a different way, and with a related but slightly different interpretation. The index, plotted (see figure 4.9) against quadrat side length $a = 1/k$, as well as fraction of contained area, shows a strong deviation from the expected value 1 for a Poisson point process for small quadrats, while approaching 1 for larger quadrats. In fact when plotted against the fraction of contained area, the relationship is almost linear in the intermediate regime.
We interpret the results presented in this section as further evidence of our hypothesis that the independence of the quadrat counts $N_1, \ldots, N_{k^2}$, as well as the Poisson-like marginal distribution and thus the validity of the Poisson approximation for large quadrat sizes, depends on the fraction of contained area, F_c, given in Eq. 4.5.

4.3.2 Area interaction process

While the argument for the Poisson approximation for large quadrats, and it's dependence on the fraction of contained area, seems reasonable for the local pairwise interactions of the Geyer model, it is not immediately clear whether this would also

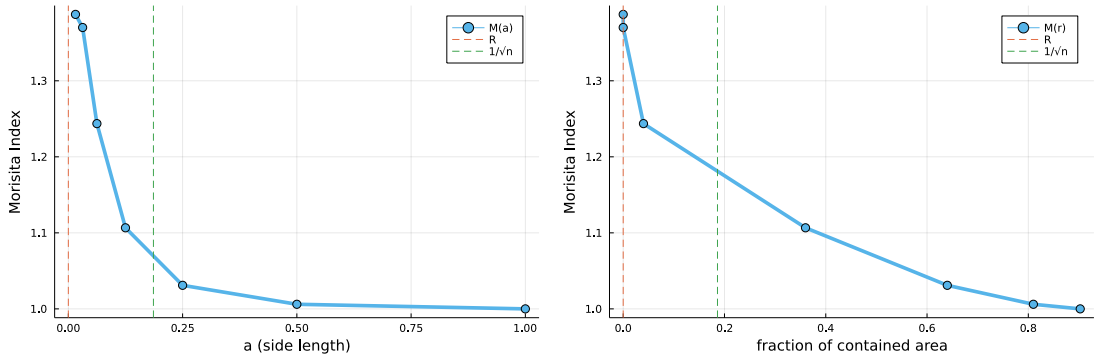


Figure 4.9: Morisita index as definded in Eq. 4.6, plotted against quadrat side length, $a = 1/k$ (left) and fraction of contained area, F_c (right).

hold for the area interaction process, which is not based on pairwise interactions, but rather on the area covered by the disks with centers at all $x \in \boldsymbol{x}$ (and is thus inherently nonlocal). We therefore spend an extra subsection on examining the same properties that we previously examined for the Geyer process, but now apply it to the area interaction process.
Fortunately, the results derived from the analysis of the Geyer process seem to readily apply also to the area interaction process (see figure 4.10), thus suggesting that they could potentially apply to all Gibbs processes.

4.4 Reconstructing point processes with spatially inhomogeneous covariates

In the previous sections we have found that for a homogeneous interacting point process, with nothing else but the counts in the different subregions, it is quite difficult to reconstruct a point configuration with the method outlined in this thesis, both in the direct (non-statistical) sense, as well as in the statistical sense. Typically, the quadrats need to be small enough so that either all points could potentially interact with points in other subregions, i.e. $a \lesssim 2R$, or small enough such that each subregion only contains one point $a \lesssim 1/\sqrt{n_W}$.
One might initially suspect that a more complicated point process, that also depends on some spatially inhomogeneous covariate, would be even harder to reconstruct with the method proposed so far in this thesis. On the other hand, a spatially inhomogeneous intensity of the point process could significantly reduce the volume of configuration space associated with higher probabilities. In other words, having areas with low- and high- intensity provides the point process more structure that can be picked up by the quadrat counts, and used in the reconstruction process.
Furthermore, if the information about the covariates is accessible to us, we could use it to improve the reconstruction process. In an epidemiological context for example, we might have access to data on population density, which would most likely act as a covariate for disease spread. It therefore would be reasonable to also use this data in the reconstruction process.
In the following section we first investigate how well the usual reconstruction pro-

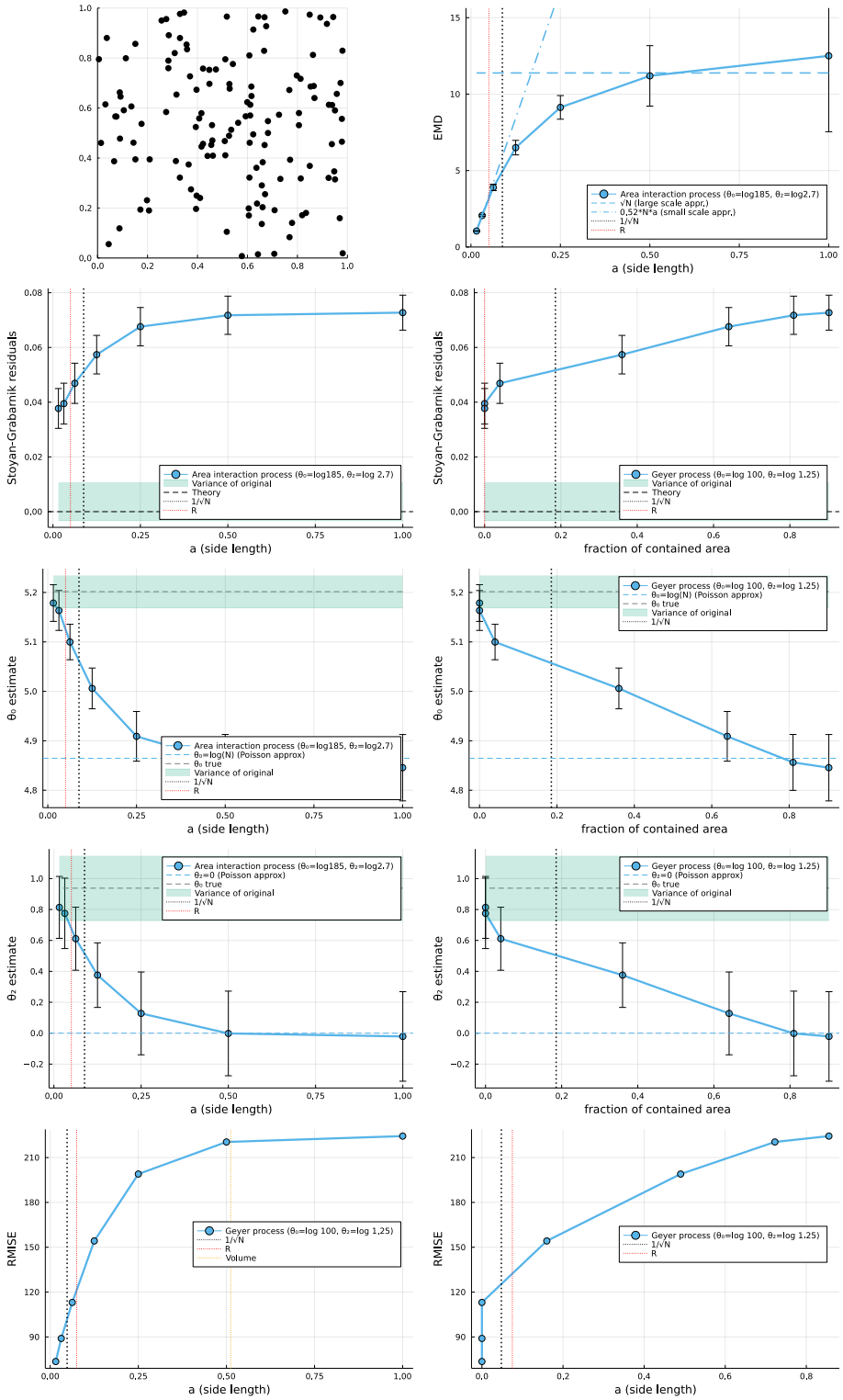


Figure 4.10: Same as in figures 4.6, but for an area interaction process with $\theta_0 = \log 185$, $\theta_2 = \log 2.7$, $R = 0.05$.

cess (using homogeneous Biniomial point process to reconstruct the points in the quadrats) works for an inhomogeneous point process. Then we attempt to reconstruct the point process using knowledge of the spatially inhomogeneous covariate

used to generate the point configuration.
For our spatial covariate, $S_1(u)$, we shall use a weighted sum of Gaussians:

$$S_1(u) = \sum_{i=1}^{m} w_i \exp\left(-\frac{(u_x - v_{x,i})^2}{2\sigma_{x,i}^2} - \frac{(u_y - v_{y,i})^2}{2\sigma_{y,i}^2}\right), \tag{4.7}$$

with weights w_i, centers $v_i = (v_{x_i}, v_{y,i})$ and widths $\sigma_{x,i}$ and $\sigma_{y,i}$, for $i = 1, \ldots, m$.

4.4.1 Inhomogeneous Poisson point process

As with the homogeneous point processes, we shall start out simple. In figure 4.11 we illustrate the importance of the scale of variation of the covariate density in the reconstruction process. The original point configuration is a sample from an inhomogeneous Poisson point process, with $\lambda(u) = \exp(\theta_0 + \theta_1 S_1(u))$, where $S_1(u)$ is a single Gaussian with a center and width chosen such that the peak is mostly contained in one of the partitions when $k = 4$, or equivalently, the side length is $a = 0.25$. It can be clearly seen that the reconstructed point configuration begins to qualitatively differ from the original configuration only once $a > 0.25$. Of course it should come as no surprise that the resolution needs to be higher than the typical scale of variation of the covariate density, in order for the reconstruction to capture the spatial variation of the intensity. Covariate densities encountered in real life situations will, however, usually not have a single scale, but several, or even contain a continuum of scales. For this reason, we turn next to somewhat more complicated covariate densities.

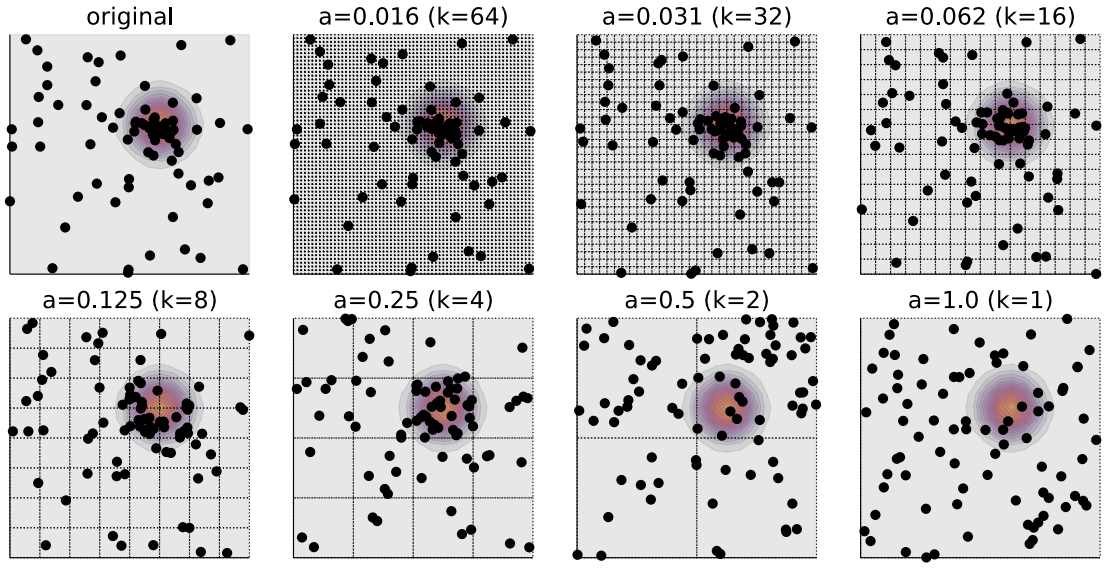


Figure 4.11: An original point configuration (top left) from an inhomogeneous Poisson point process $\lambda(u) = \exp(\theta_0 + \theta_1 S_1(u))$, with $S_1(u) = \exp(-(u - v)^2/2\sigma^2)$, $v = (0.625, 0.625)$ and $\sigma = 0.075$. The rest of the point configurations are reconstructions at resolutions indicated at the top.

In figure 4.12 we show sample point configurations, as well as EMD, Stoyan-Grabarnik residuals and parameter estimates, of samples from an inhomogeneous point process

with $\lambda(u) = \exp(\theta_0 + \theta_1 S_1(u))$, with $S_1(u)$ a weighted sum of three Gaussians. While $\theta_0 = 80$ is kept the same, the three columns represent the choices $\theta_1 = 0$, 2.5, 5 in increasing order. As can be seen, the dissimilarity measures are again small for side lengths $a < 1/\sqrt{n}$ (small quadrat limit), but approach the large quadrat limits in a more linear fashion than what we previously observed in the interaction cases, where the fraction of contained area determined the shape. In the inhomogeneous Poisson process interactions are absent, and thus the area of contained interactions is not relevant. The more linear shape is formed by the presence of multiple different scales of the three (overlapping) Gaussians, each with different widths in the x- and y-directions. This behaviour is observed for most randomly chosen configurations of weighted sums of Gaussians, and we expect it to be a more or less general feature when reconstructing inhomogeneous point processes.

4.4.1.1 Reconstructing using inhomogeneous binomial point process

It is interesting to investigate whether knowledge of the covariate density $S_1(u)$ can be used to improve the reconstruction process. Of course, if we had access to the full intensity $\lambda(u)$ of the inhomogeneous Poisson point process we wish to reconstruct, doing so by using a binomial point process with n_i points and density $j_i(u) = \lambda(u)/\int_{A_i} \lambda(u)\,\mathrm{d}u$ on corresponding quadrats A_i, would result again in a Poisson point process with density $\lambda(u)$ (see section 4.1). Thus, the statistical dissimilarity measures would be small, regardless of quadrat size, while the non-statistical dissimilarity measure, EMD, would behave similarly to the reconstruction of a homogeneous Poisson point process. Here however, the only knowledge of the covariate density $S_1(u)$ is known, while the parameters θ_0 and θ_1 are considered unknown. Thus the most non-informed reconstruction, apart from the homogeneous Binomial reconstruction, is one where we use an inhomogeneous binomial reconstruction, with density $j_i(u) \propto S_1(u)$, $u \in A_i$. Of course, this runs the risk of instead overestimating the spatial dependence of the intensity on the covariate density.
Indeed, as can be seen in figure 4.13, using the inhomogeneous binomial reconstruction causes a consistent overestimation of the covariate parameter θ_1, while underestimating θ_0. However, in contrast to the homogeneous reconstruction method, the inhomogeneous one becomes better the larger θ_1, i.e. the stronger the dependence of the original point process on the covariate density $S_1(u)$, the better the reconstruction. Thus, if you feel certain that the covariate density $S_1(u)$ should have a large influence, then using the inhomogeneous binomial reconstruction is probably the better choice.

4.4.1.2 Reconstructing inhomogeneous interacting point process

To conclude this section, we investigate the quality of reconstruction for an inhomogeneous interacting point process. For this we choose the inhomogeneous Geyer process, with Papangelou conditional intensity $\lambda_{\boldsymbol{\theta}}(u; \boldsymbol{x}) = \exp(\theta_0 + \theta_1 S_1(u) + \theta_2 S_2(u, \boldsymbol{x}))$, where $S_2(u; \boldsymbol{x})$ is the usual interaction function for the Geyer process, with nuisance parameters $R = 0.05$ and $s^* = 4$, while $S_1(u)$ is the covariate density as in the previous subsection. In figure 4.14 we show the dissimilarity measures for homogeneous binomial reconstructions, whereas in figure 4.15 we show them when using

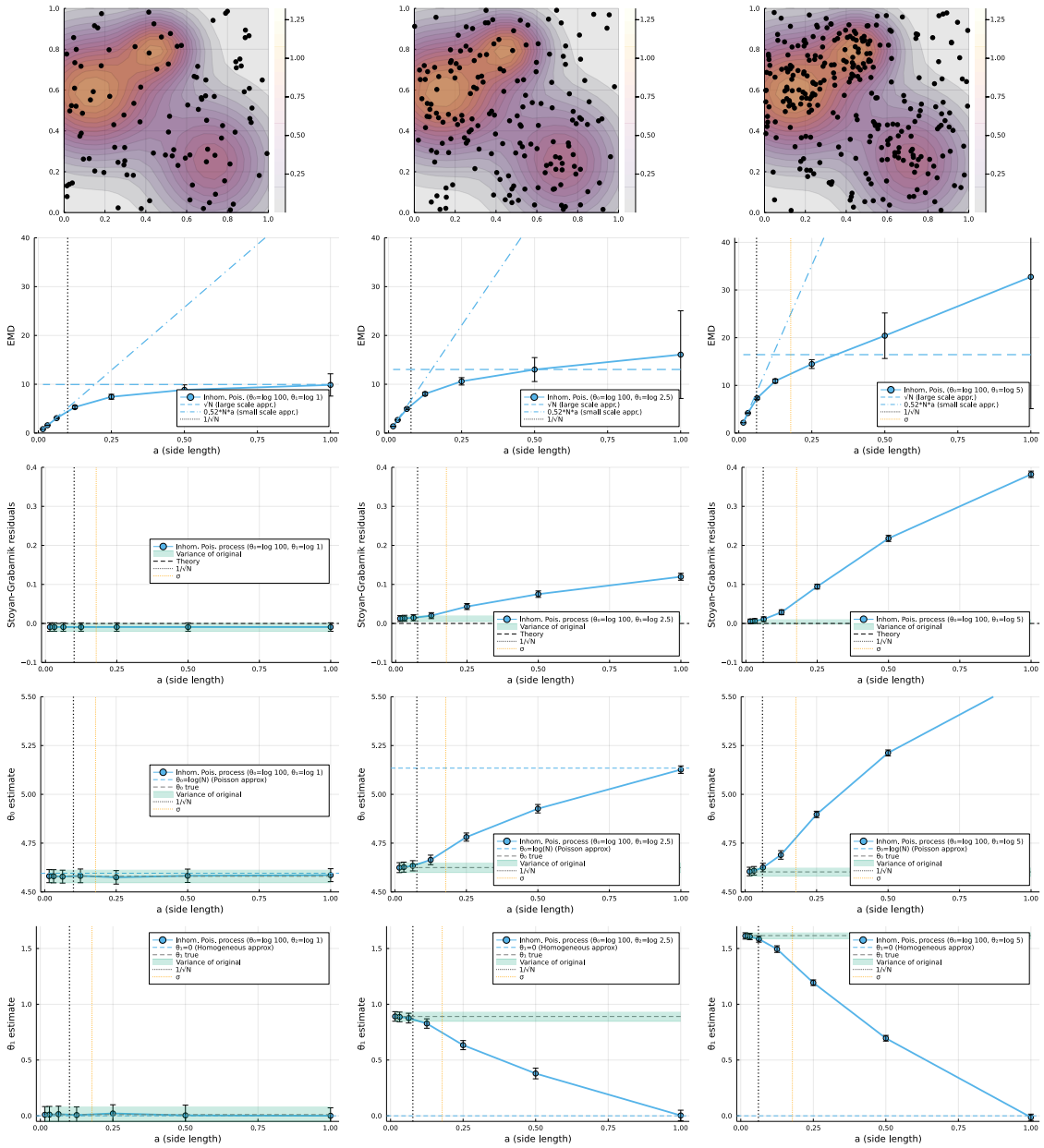


Figure 4.12: Samples and reconstruction dissimilarity measures for inhomogeneous Poisson point process, with $\lambda(u) = \exp(\theta_0 + \theta_1 S_1(u)$, where $S_1(u)$ is of the form 4.7. The parameters are chosen as $\theta_0 = 80$, and $\theta_1 = 0.0, 2.5, 5.0$ for left, center and right columns respectively. The top row shows the covariate density $S_1(u)$ and a sample from the inhomogeneous Poisson point process. Second to fourth rows show the dissimilariy measures EMD, Stoyan-Grabarnik, θ_0-estimate and θ_1-estimates as a function of the side length a of quadrats used in the reconstruction.

the inhomogeneous binomial reconstruction method.

Starting with the homogeneous reconstruction, figure 4.14, we can identify three regimes, roughly separated by the scales R or $1/\sqrt{n_W}$ of interactions, and the scale of spatial variation of $S_1(u; \boldsymbol{x})$, which here is represented by $2\bar{\sigma} = \frac{1}{3}\sum_{i=1}^{3}(\sigma_{x,i} + \sigma_{y,i})$. For large quadrats, $a > 2\bar{\sigma}$, the quality of reproduction is rather poor since the re-

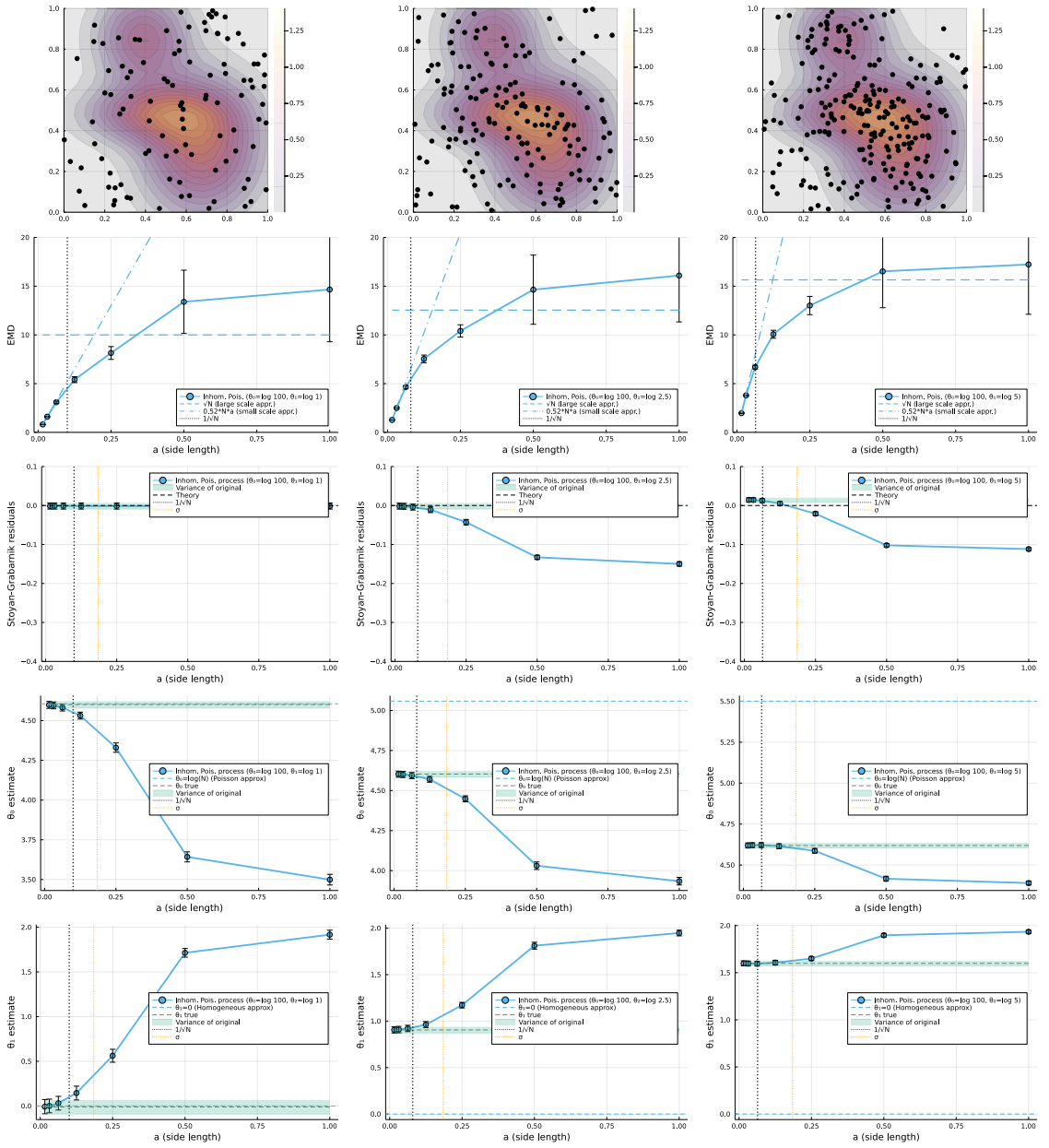


Figure 4.13: Same as in figure 4.12 but using an inhomogeneous binomial reconstruction process in the reconstruction, with $j_i(u) \propto S_1(u)$ as 1-point prob. density.

production method does not capture neither the inhomogeneity nor the interaction (clustering). In the intermediate regime, $R, 1/\sqrt{n_W} < a < 2\bar{\sigma}$, the resolution is sufficiently fine so as to capture the spatial variation of the covariate density, but not fine enough to capture the interaction structure. This can be seen by the rapid improvement (as a becomes smaller) of the estimate for θ_1, while the estimate of θ_2 almost slows down. Indeed, we interpret the overestimation of θ_1 around $a \gtrsim R, 1/\sqrt{n_W}$ as the parameter estimation method missattributing some of the clustering to larger values of the covariate density, rather than to the interacting nature of the the point process. This is then (rapidly) corrected in the small quadrat regime, $a < R, 1/\sqrt{n_W}$, where the resolution is small enough to capture the clustering due to interactions

(as opposed to the covariate intensity).

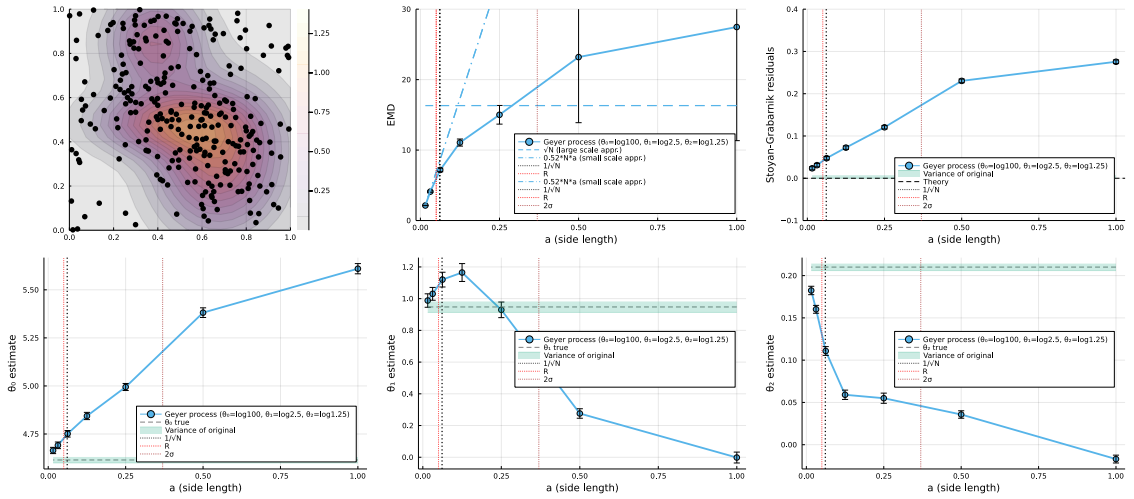


Figure 4.14: Reconstruction of an inhomogeneous Geyer process using a homogeneous binomial point process. On the top left we provide a sample of the original point process along with a density map of the covariate density. The graphs represent the different dissimilarity measures: EMD, Stoyan-Grabarnik residuals, as well as parameter estimates, as a function of quadrat side length a.

When using the inhomogeneous reconstruction method, we again observe the same three distinct regimes, but with somewhat different behaviour within them. In the large quadrat regime, $a > \bar{\sigma}$, the parameters estimated from the reconstructed point configurations show an overestimation of θ_1, accompanied with an underestimation of θ_0. This is due to the reconstruction process using a 1-point prob. density proportional to the covariate density to redistribute the points within the quadrats. In the intermediate regime, $R, 1/\sqrt{n_W} < a < 2\bar{\sigma}$, the resolution is fine enough to capture the spatial variation of the covariate density, making it easier to distinguish the correlation between covarariate density and the intensity of the point process, thus producing better estimates of the parameters θ_0 and θ_1. Meanwhile, the estimation of the interaction parameter θ_2, in this regime, worsens as a becomes smaller. We again attribute this effect as an inability to distinguish between clustering due to covariant density, and clustering due to point interaction. This is again corrected rapidly with the small quadrat regime $a < R, 1/\sqrt{n_W}$.
While most of the difference between the two methods of reconstruction can be attributed to the different emphasis placed on θ_0, θ_1 and θ_2, one might not necessarily be interested in the particular values of these parameters, but rather in whether it is possible to use the reconstructed point process to estimate the conditional intensity, $\lambda_\theta(u; \boldsymbol{x})$. Thus it is interesting to explore whether $\lambda_{\tilde{\boldsymbol{\theta}}}(u; \tilde{\boldsymbol{x}})$, constructed from the parameter estimates $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and the reconstructed point configuration $\tilde{\boldsymbol{x}}$. This is explored in figure 4.16. Interestingly, the wild variations in parameter estimates $\tilde{\boldsymbol{\theta}}$ as a function of quadrat side a is more or less absent in the RMISE, which has a much smoother and consistent behaviour. One can still somewhat distinguish the three different regimes, and at least in this case, the inhomogeneous reconstruction process produces smaller values of the RMISE, suggesting it provides a better estimate of

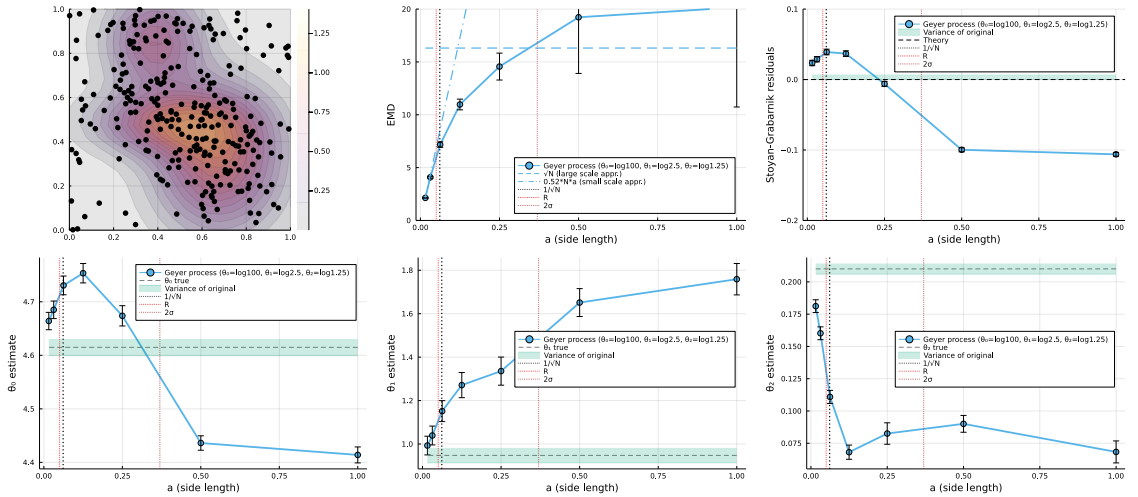


Figure 4.15: Same as in figure 4.14, but using an inhomogeneous binomial point process for reconstruction.

the Papangelou conditional intensity.

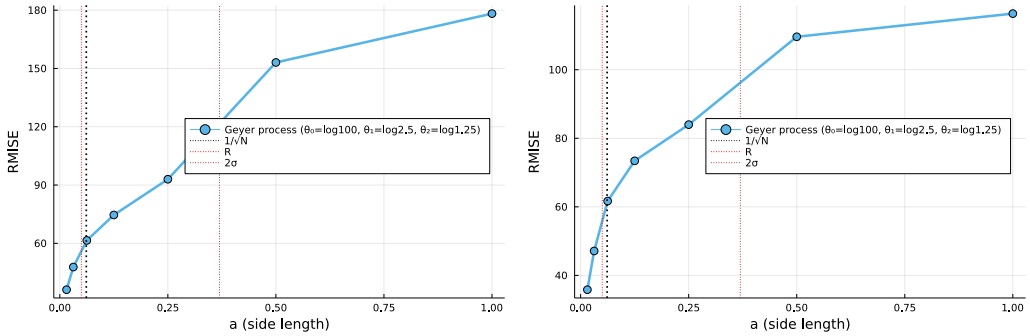


Figure 4.16: RMISE (root mean integrated square error) between $\lambda_{\boldsymbol{\theta}}(u;\boldsymbol{x})$ and $\lambda_{\tilde{\boldsymbol{\theta}}}(u;\tilde{\boldsymbol{x}})$, where $\lambda_{\tilde{\boldsymbol{\theta}}}(u;\tilde{\boldsymbol{x}})$, is constructed from the parameter estimates $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and the reconstructed point configuration $\tilde{\boldsymbol{x}}$. Using both a homogeneous (left) and inhomogeneous (right) point process in the reconstruction.

4.5 Reconstructing point processes on linear networks

To conclude the chapter we finish by exploring reconstruction of point processes on linear networks. To produce a somewhat random linear network, we choose randomly a set of pairs of points on the borders of the observation window W, and construct lines between them. This results in a set of random lines across the observation window. Then, for each line we identify its intersections with the other lines and split the line into linear segments such that no two segments intersect each other. The result is a linear network, i.e. a set of linear segments. The linear network can then be further processed by *thinning*, i.e. randomly removing a fraction

of the segments to deviate more from the random lines we started with. Figure 4.17 illustrates the process, as well as an example of a density map of a covariate density $S_1(u)$, formally defined over the entire observation window, but for a points process only the values on the segments of the linear network would be of importance.

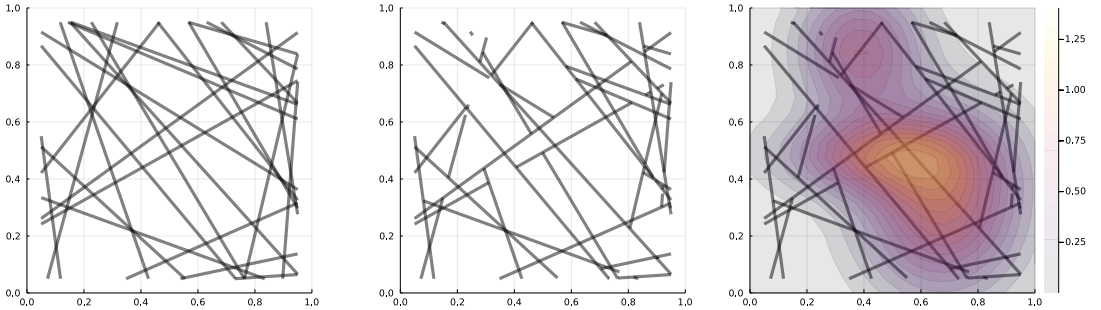


Figure 4.17: The formation of a linear network: The leftmost image shows a set of lines constructed randomly by drawing pairs of points from the borders of the observation window and drawing lines between them. Then one identifies the intersection points and splits the lines into linear segments, ℓ. The corresponding linear network, L, then consists of the set of these segments, $\ell \subset L \subset W$. The linear network can then be thinned, as in the center figure. On the right figure we illustrate a covariate density defined on the observation window W, and thus also on $\ell \subset W$.

Before presenting the results of the reconstructions, we shall address some necessary adjustments and restrictions necessary when dealing with linear networks, as opposed to the entire observation window. One issue relates to which distance metric to use, both in the context of interactions as well as in the dissimilarity measures (mainly EMD). As for interactions, it may, in many situations, be reasonable to use a shortest path distance along the network for example when it is reasonable to expect that the interactions can only be mediated through the network. However, since one of the motivations for this work is epidemic modelling, it seems unreasonable to assume that a point (infection) on one line segment would not affect the probability of infections in another close (in the 2D Euclidean sense) line segment, simply because the line segments are not directly connected. For this reason, we choose to use a 2D Euclidean distance metric defined on the observation window W in which L is embedded. On the other hand, when considering dissimilarity measures, one needs to consider the fact that the points can only be placed on the linear network, reducing the space of configurations significantly. Not only does this mean that our analysis in section 4.2 on the expected behaviour of EMD under reconstruction is invalid, but one would also need to rewrite the code used to compute the EMD. While not difficult, it is a task that would be a little too time consuming to be worthwhile given the time restrictions of this thesis. In a similar way, while the Stoyan-Grabarnik residuals can certainly be computed on linear networks, it introduces extra complexities and requires the code for the computation to be rewritten in a significant and time consuming way. While we have written such an algorithm, we found some issues with odd behaviour for point configurations reconstructed at very high resolutions. For this reason, and due to time constraints, we decided not

to include it in this thesis.
On the bright side, parameter estimation could be easily extended to be applied to point processes on linear networks. Furthermore, the theoretical analysis in section 4.1 can readily be extended to linear networks with minor modifications, and in particular the Poisson approximation at large quadrats should still apply to some extent.
To investigate the reproduction of point processes on linear networks, we decided to use both an inhomogeneous Poisson point process, as well as an interacting inhomogeneous Geyer process. In both cases, the covariate density is defined in the same way as in the previous 2D analysis. The results are presented in figure 4.18. As can be seen, the dissimilarity measures (here the parameter estimates) show a very similar behaviour as in the 2D case, compare 4.12 and 4.14. In particular, we can identify the three different regimes, separated by the interaction scale $R, 1/\sqrt{n_W}$ as well as the scale of spatial variation $2\bar{\sigma}$. We also observe the overestimation of θ_1 and related underestimation of θ_2 close to $a \gtrsim 1/\sqrt{n_W}$, when the resolution is fine enough to capture the spatial variation of the covariate density, but not fine enough to capture the interaction.

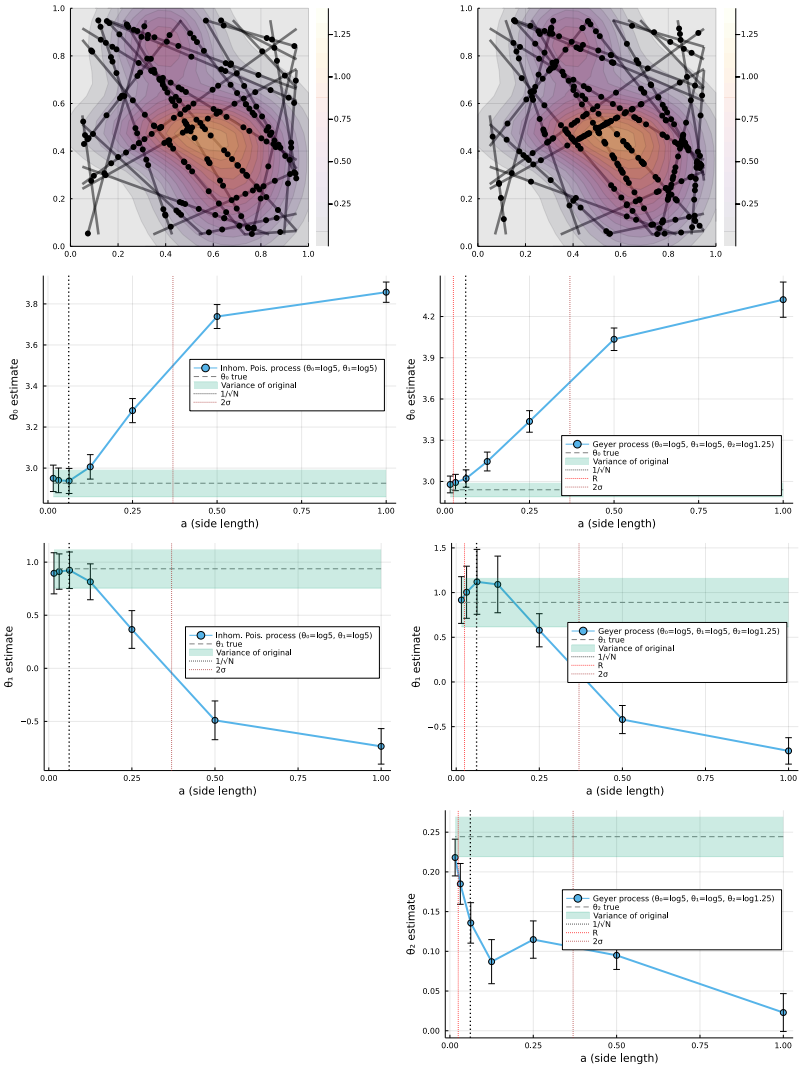


Figure 4.18: Reconstruction of point processes on a linear network. The left column of figures contains the results of an inhomogeneous Poisson point process, while the right column contains the results of an interacting inhomogeneous Geyer process. The top row shows the linear network, density maps of the covariate density, as well as samples of the point processes on the network. In rows 2-4 we show the parameter estimates for θ_0, the covariate parameter θ_1 and interaction parameter θ_2 (only for the Geyer process) of the corresponding point processes.

5

Conclusion

In this thesis we have briefly introduced the theory of finite point processes and the tools and methods used in our analysis, presented a theoretical analysis of the reconstruction process as well as results from both interacting and non-interacting as well as homogeneous and inhomogeneous point processes on a 2D planar geometry and on linear networks. The results were presented mainly in terms of a set of dissimilarity measures between the original- and reconstructed point configurations, as a function of the size (or resolution) of the partitions used in the reconstruction process.

The dissimilarity measures were separated into two types; direct (or non-statistical) and statistical dissimilarity. For the direct dissimilarity, we used the earth mover's distance (EMD), which showed a fairly universal behaviour. Even when reconstructing a homogeneous Poisson point process, using homogeneous binomial point process on each quadrat (in which case the reconstructed point configuration is again Poisson with the same intensity), the EMD showed a similar behaviour as when reconstructing other point processes. As such, it is deemed a rather useless measure if one wishes to measure how well the reconstructed point configurations preserve the statistical properties of the original process.

The statistical dissimilarity measures used were; Stoyan-Grabarnik residuals, parameter estimates, and root mean integrated square error (RMISE) of the original, and estimated/reconstructed Papangelou conditional intensities. In the case of homogeneous interacting point processes, exemplified by the Geyer process and area interaction processes, we identified two distinct regimes. The first, typically referred to as the "small quadrat regime", where the quadrats used in the reconstruction process had side lengths $a < R, 1/\sqrt{n_W}$, where R is the interaction range, and n_W is the number of points in the observation window W. In this regime the statistical properties (mainly interaction properties) were well preserved in the reconstruction process. In the second regime, $a^2 \approx W$, sometimes referred to as "large quadrat regime", or "Poisson approximation" the statistical properties of the reconstructed point configuration are more consistent with a Poisson point process than the interacting point process. We argued that the validity of the Poisson approximation is determined by the dependence-structure of the quadrat counts N_i, as well as the properties of the marginal distributions of these, i.e. whether they can be approximated by a Poisson distribution. Furthermore, we argued that this in turn depends on the available area for points capable/incapable of interacting with points in other quadrats. Thus we introduced the concept of "fraction of contained area", i.e. the fraction of quadrat area within which points are incapable of interacting with points from other quadrats. We argued that the validity of the Poisson approximation

should be proportional to this quantity, and plots of the statistical dissimilarity measures against this quantity showed a remarkably linear behaviour in the regime between small and large quadrats, thus providing an interpolation, and simple interpretation of the transition between the two. This provides a simple tool to estimate whether a reconstruction is expected to preserve the statistical properties of the original point process, only requiring the interaction range R and the size/geometry of the quadrats.

In the context of epidemic modelling: if only count data of subregions is available, a reconstruction method like the one presented in this thesis could produce a point configuration which preserves the statistical properties of the (unknown) original, if the interaction range and subregion size and geometry are such that the fraction of contained area can be considered large.

We also explored the ability to reconstruct inhomogeneous point processes in the case when the inhomogeneity can be attributed to some known covariate density (e.g. population density). Here we identified three regimes, small quadrat regime, $a < R, 1/\sqrt{n_W}$, intermediate regime, $R, 1/\sqrt{n_W} < a < 2\bar{\sigma}$ and large quadrat regime $2\bar{\sigma} < a \lesssim \sqrt{|W|}$. Here $2\bar{\sigma}$ is a measure of the average spatial scale of variation of the covariate density. Thus, the intermediate regime concerns cases when the resolution of the partition is fine enough to capture the spatial variation of the covariate density, but not fine enough to capture the interaction properties of the original point process. Similar results were obtained when applying the reconstruction method to linear networks. This result provides further guidance for what to expect when attempting to reconstruct a point process from quadrat counts: If the interest is mainly in estimating the dependence of the point process on some known covariate density, using a reconstruction method as outlined in this study would require the quadrats to be larger than the typical scale of variation of the covariate density. If one also wants to estimate the interactions, the reconstruction methods require the quadrats to be larger than R, and/or contain on average $\lesssim 1$ point, i.e. $n_W/a^2 \lesssim 1$.

In light of these sombering results, we attempted to improve the reconstruction process by making use of the known covariate density. This was achieved by using instead an inhomogeneous binomial point process for each quadrat, with a 1-point probability density proportional to the covariate density. This did indeed result in a better reconstruction, provided the original point process had a strong dependence on the covariate density. On the other hand, if the dependence on the covariate density was weak, then this method of reconstruction resulted in a worse reconstruction. In other words, in order to justify this method of reconstruction, one would have to be certain that the underlying (original) point process is strongly dependent on the covariate density.

In the context of epidemic modelling, these results are somewhat disheartening. Although it is difficult to estimate, or even clearly define, an interaction radius when dealing with the spread of disease, it is unlikely that this interaction range would be close to the typical size of a reporting subregion wherein cases are counted. Thus, if for example, we are interested in the relative importance of interactions (i.e. clustering propensity) versus some covariate variable such as population density, we would require sizes of the subregions to be small enough that infections across subregion borders are not only common, but the norm.

5.1 Discussion and future work

Limitations on time and resources have certainly led to many stones left unturned. In this section I would like to present some possible avenues that the work could have explored further.
The main idea behind the reconstruction scheme was to be able to estimate the underlying point process model, when the only accessible data is point counts in subregions. For concreteness, suppose we are interested in the parameters of a given model. As can be seen in this thesis, in order to accurately estimate the parameter associated with a covariate density, one needs subregions that are smaller than the typical scale of variation of such a covariate. In order to accurately estimate the interaction parameter, one needs subregions on the scale of the interaction distance. On the other hand, in the case of interactions, we presented here a simple model to predict how much the interaction parameter would be underestimated (in terms of the fraction of contained area). This opens up the possibility of adjusting the estimate of the interaction parameter accordingly. Presumably, a similar model, and corresponding adjustment, could be obtained for the parameter associated with the covariate density. In future work, this could be explored and the validity of such adjustments could be tested in simulation studies.
Furthermore, we concluded in section 4.1 that the main information that provides the reconstruction method with any properties from the original point process, is encoded in the joint probability density of the quadrat counts, N_i. This begs the question if one could not infer the parameters of a given point process model directly from these counts. One such approach is Bayesian likelihood free inference, or ABC (approximate Bayesian computation). The idea here is to define a parameter space Θ, and possibly a prior $p(\theta)$, $\theta \in \Theta$, and a model $p(\boldsymbol{x}|\theta)$ from which we may generate samples $\boldsymbol{x}$. The samples may be used to compute the counts $n_i = |\boldsymbol{x} \cap A_i|$ in each of the quadrats A_i. Then, a rejection scheme can be implemented where one draws samples $\theta^\star \sim p(\theta)$ and then generates a point configuration $\boldsymbol{x}^\star \tilde{p}(\boldsymbol{x}|\theta^\star)$. The parameter value $\theta^\star$ is then accepted, if the corresponding counts $n_i^\star = |\boldsymbol{x}^\star \cap A_i|$ are not sufficiently different from the observed counts n_i^{obs}, as measured by some distance $d(\boldsymbol{n}^\star, \boldsymbol{n}^{\text{obs}}) < \varepsilon$ with some threshold ε. Repeating this procedure produces a set of parameter values which is an approximate sample of the posterior distribution, $p(\theta|\boldsymbol{x}^{\text{obs}})$. While we did experiment with this approach, including some improvements like regression adjustment, synthetic likelihood and ABC-SMC[27], the approach was ultimately too time consuming and warrants it's own study.

Bibliography

[1] D. Giuliani, MM. Dickson, and F. Santi. Modelling and predicting the spatio-temporal spread of covid-19 in italy. *BMC Infect Dis*, 20:1–10, 2020.

[2] AB. Lawson and J. Kim. Space-time covid-19 bayesian sir modeling in south carolina. *PLoS ONE*, 16:e0242777, 2021.

[3] A. Iftimi, M.N.M. van Lieshout, and F. Montes. A multi-scale area-interaction model for spatio-temporal point patterns. *Spatial Statistics*, 26:38–55, 2018.

[4] M.N.M. van Lieshout. *Markov Processes and Their Applications*. Imperial College Press, London, 2000.

[5] A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns - Methodology and Applications with R*. Chapman and Hall/CRC, Boca Raton, FL, 2016.

[6] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. Imperial College Press, London, 2008.

[7] D. J. Daley and D. Vere-Jones. *Introduction to the Theory of Point Processes: Elementary Theory and Methods*. Springer, New York, second edition, 2006.

[8] Strauss, D.J. A model for clustering. *Biometrica*, 62(2):467–475, 08 1975.

[9] Geyer, C. *Stochastic Geometry: Likelihood inference for spatial point processes*, chapter 3, pages 79–140. Chapman and Hall–CRC, Boca Raton, FL, 1999.

[10] A. Baddeley and M.N.M. van Lieshout. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, 47:601–619, 1995.

[11] F. Papangelou. The conditional intensity of general point processes and an application to line processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 28(3):207–226, 1974.

[12] H.-O. Georgii. Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics*, 48:31–51, 1976.

[13] X. Nguyen and H. Zessin. Integral and differential characterizations of gibbs processes. *Mathematische Nachrichten*, 88:105–115, 1979.

[14] M. N. M. van Lieshout. Nearest-neighbour markov point processes on graphs with euclidean edges. *Advances in Applied Probability*, 50(4):1275–1293, 2018.

[15] T.J. Leininger and A.E. Gelfand. Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Analysis*, 12(1):1–30, 2017.

[16] J.E. Besag. Statistical analysis of non-lattice data. *The statistician*, 24:758–763, 1975.

[17] J.E. Besag. Some mehtods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77–92, 1978.

[18] J. Moller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton, FL, 2004.

[19] M. Berman and R. Turner. Approximating point process likelihoods with GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):31–38, 1992.

[20] Baddeley A. and R. Turner. Practical Maximum Pseudolikelihood for Spatial Point Patterns. *Australian and New Zealand journal of statistics*, 42(3):283–322, 2000.

[21] A. Baddeley, J.-F. Coeurjolly, E. Rubak, and R. Waagepetersen. Logistic regression for spatial Gibbs point processes. *Biometrica*, 101(2):377–392, 2014.

[22] H.W. Kuhn. The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955.

[23] J. Munkres. Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[24] D. Stoyan and P. Grabarnik. Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachrichten*, 151:95–100, 1991.

[25] A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual Analysis for Spatial Point Processes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(5):617–666, 2005.

[26] J.F.C. Kingman. *Poisson processes*, volume 3. Oxford university press, 1992.

[27] S.A. Sisson, Y. Fan, and M.A. Beaumont. *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, Boca Raton, FL., 2019.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY