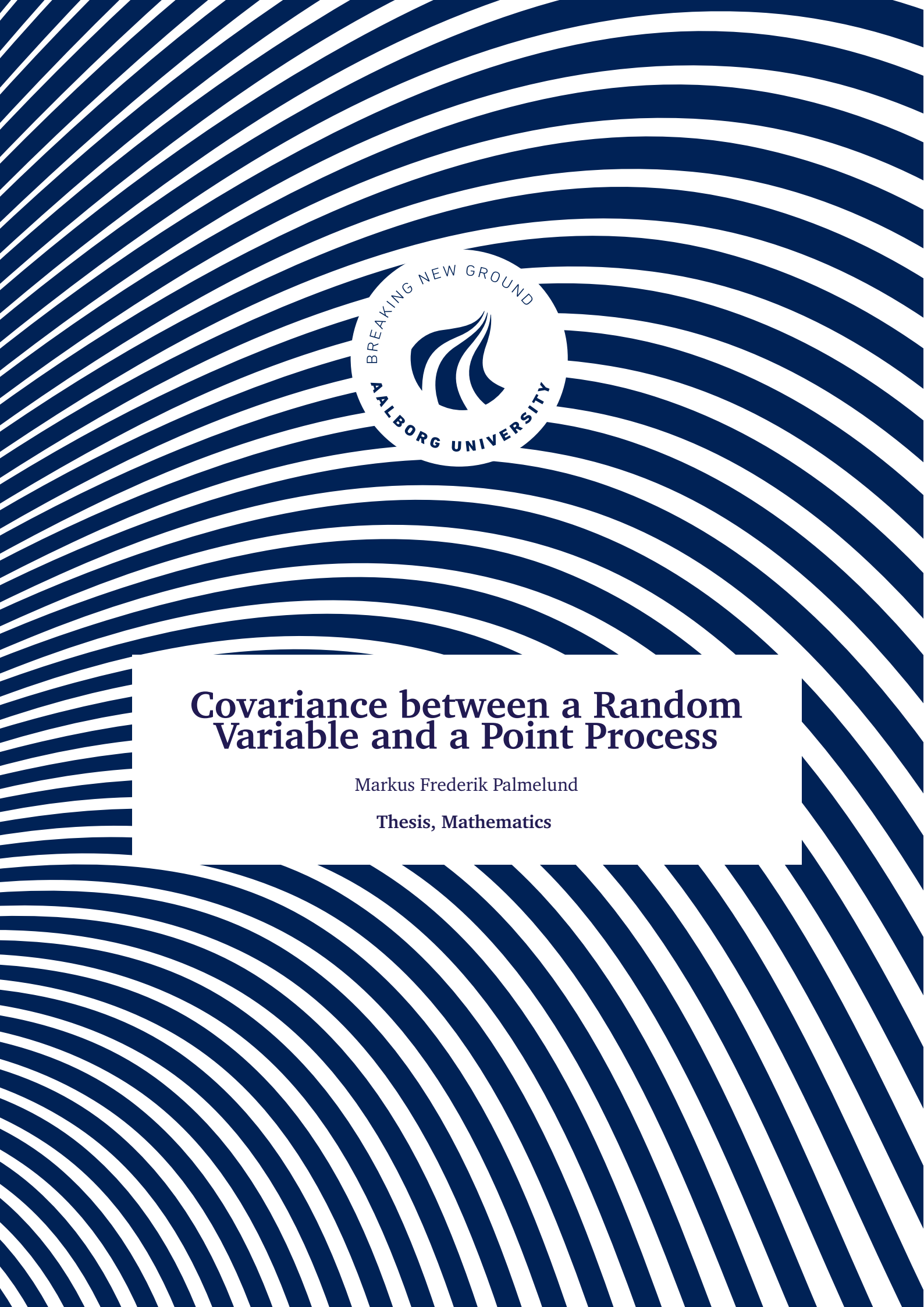

Covariance between a Random Variable and a Point Process

Markus Frederik Palmelund

Thesis, Mathematics

Dept. of Mathematical Sciences
Skjernvej 4A
DK-9220 Aalborg Ø
http://math.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title
Covariance between a Random Variable and a Point Process

Theme
Spatial Point Process

Project Period
Spring Semester 2025

Participant
Markus Frederik Palmelund

Supervisor
Rasmus Waagepetersen

Page Numbers
49

Date of Completion
May 28, 2025

Abstract

This thesis primarily examines the covariances between a point process and random variables, which are binary and continuous. From each covariance two summary functions are extracted, which are given non-parametric estimators, that both include and exclude an edge-corrected factor. A simulation study evaluates the estimators, which shows that edge-corrected estimators work best and is able to detect correlation and uncorrelation between a point process and random variables. Finally, the estimators are applied to a data set from a rainforest on Barro Colorado Island in Panama. Before introducing the summary function, the thesis shows some theory about point processes, including the pair correlation function, Poisson point process, Cox point process, and the covariance between two point processes.

Preface

This thesis was written by a master's student at the Department of Mathematical Sciences at Aalborg University.

Equations, definitions, propositions, theorems and lemmas are numbered according to the section, while figures and tables are numbered by order.

A thanks goes to Rasmus Waagepetersen for help with supervision during the writing of the thesis.

Aalborg University, May 28, 2025

Contents

1 Introduction

The main subject of this project is to explore the correlation between spatial point processes and random variables. The motivation is to be able to analyse more aspects of the correlations for spatial point processes. An example of a spatial point process and random variable could be the locations of trees and the pH value of the soil in a rainforest. I want to answer the question: Is the random variable, such as pH, at a location correlated with the spatial point process of trees?

First, I will introduce some general terms and two examples of spatial point processes. One of these point processes is the Poisson, which is independent of disjoint sets. One of the terms is the pair correlation function, which can confirm a point process is Poisson. I also compare the covariance of the Poisson and Cox point processes.

Hereafter, I will explore the covariance between a random variable and a point process. The random variables explored in the thesis consist of binary and continuous variables. From the covariances between the point process and both types of random variables, I derive two summary functions, which can determine the correlation between the random variable and the spatial point process. For the summary functions, I will give some non-parametric kernel estimates with and without edge-correction factor. I will examine the performance of the estimates in simulation studies of the uncorrelated and correlated cases. To simulate the correlated cases, I evaluate the summary function for the log Gaussian point processes in a special case. Finally, I intend to apply the estimates to an existing data set of trees and soil measurements from Barro Colorado Island in Panama.

2 Point Process

In this project, I will only work with spatial point processes defined within a state space $S \subseteq \mathbb{R}^2$. A point process $X \subset S$ is a random countable subset of points within the state space S. The point processes can in the plane be used to model the location patterns and describe the spatial aggregation within a given dataset. Fields where we have such interest could be e.g epidemiology, forestry and geography where we may have the locations of infected, trees and human settlements.

I will concentrate on $B \subseteq S$ bounded subsets of the state space. Also, in practice, a realisation of a point process is called a point pattern x, and when observing these, we are constrained to a smaller bounded part of the state space called the window and denoted W. Note the difference between a set and window is that a window is the set where we observe the points in practice.

The point process within either a set or window A is denoted as $X_A = X \cap A$ while the random number of points within A are denoted by $N(A)$ and the volume of A by $|A|$. A point pattern within A is denoted x_A.

If $N(x_A) < \infty$ then x is said to be locally finite. I will in the project only concentrate on locally finite point processes, which take their values from the space defined as

$$N_{lf} = \{x \subseteq S : N(x_B) < \infty \text{ for all bounded } B \subseteq S\}.$$

The elements in N_{lf} are called the locally finite point configurations.

Another notation used in the project is $W_u = \{v + u : v \in B\}$, which is the translation of W by $u \in \mathbb{R}^2$.

Throughout this section, I describe some basic theory about the Poisson point process, intensity function, the pair correlation function and Cox point process. This will in the end lead to analysis of the covariance between two spatial point process in subsection 2.5. This section is based on [1].

2.1 Poisson Point Process

In this subsection, I will introduce some basic aspects of the Poisson point process with special focus on the spatial randomness for Poisson point processes.

Before defining the Poisson point process, I will introduce the binomial point process.
Definition 2.1 (Binomial Point Process)**.** Let f_B be a density function on a set $B \subseteq S$. If a point process X consists of $N \in \mathbb{N}$ i.i.d points with density f_B, then X follows a binomial point process with n points in B and density f_B. This binomial point process is denoted $X \sim \text{binomial}(B, n, f_B)$.

In the definition of the Poisson point process, I will use a function $\lambda : S \to [0, \infty)$ called the intensity function. Furthermore, for any bounded set B, the integral $\mu(B) = \int_B \lambda(u)du < \infty$ is called the intensity measure. The intensity measure satisfies that $\mu(\{u\}) = 0$ for any $u \in S$. Heuristically, $\lambda(u)du$ is the probability that a point occurs within the circle with centre u and infinitesimally small radius.

In the following definition, $Y \sim \text{Pois}(\mu)$ denotes that Y follows a Poisson distribution with mean μ. Note that general the Poisson distribution is applied to find the probability that a number of events occur within an area, given the expected number of occurrences.

Definition 2.2 (Poisson Point Process)**.** The point process X is said to follow a Poisson point process on S with intensity function λ and denoted $X \sim \text{Poisson}(S, \lambda)$ if the following is satisfied:

- For any bounded set $B \subseteq S$ with $\mu(B) < \infty$, $N(B) \sim \text{Pois}(\mu(B))$.
- For any bounded set $B \subseteq S$ with $\mu(B) \in (0, \infty)$, $N(B) = n$ and $X_B \sim \text{binomial}\,(B, n, f_B)$ where $f_B(u) = \lambda(u)/\mu(B)$ for $u \in B$.

Furthermore, if λ is a constant over S, the point process is said to be homogeneous and else inhomogeneous.

In the project, the theory used will be for the inhomogeneous intensity. However, the theory can also be applied for the homogeneous case. Sometimes the methods are in fact more simple for the homogeneous intensities.

In the following, I will show the independence property of the Poisson point process. First, a proposition is introduced to prove the independent property.

Proposition 2.3. *(i) $X \sim$ Poisson(S, λ) if and only if for all bounded sets $B \subseteq S$ with $\mu(B) < \infty$ and all $F \subseteq N_{lf}$,*

$$P(X_B \in F) = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B \mathbf{1}[\{x_1, \ldots, x_n\} \in F] \prod_{i=1}^{n} \lambda(x_i) dx_1 \ldots dx_n.$$

(ii) If $X \sim$ Poisson(S, λ), then for a function $h : N_{lf} \to [0, \infty)$ and bounded set $B \subseteq S$ with $\mu(B) < \infty$

$$\mathbb{E}[h(X_B)] = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B h[\{x_1, \ldots, x_n\}] \prod_{i=1}^{n} \lambda(x_i) dx_1 \ldots dx_n.$$

Proof. Note that (ii) follows from (i) by the standard proof (See Lemma A.1).

To prove (i), let $p_N(n)$ be the density of the number of points, $f_{X|N}\left(\{x_1, \ldots x_n\} \mid n\right)$ be the density for the point process conditional on the number of points and $f_{X_i|N}\left(\{x_i\} \mid n\right)$ be the density for the point X_i in the point process conditional on the number of points. Then, by the law of total probability and definition 2.2,

$$\begin{aligned}
P(X_B \in F) &= \sum_{n=0}^{\infty} P\left(N(B) = n\right) P\left(X_B \in F \mid N(B) = n\right) \\
&= \sum_{n=0}^{\infty} p_N(n) \int_B \cdots \int_B \mathbf{1}\left[\{x_1, \ldots x_n\} \in F\right] f_{X|N}\left(\{x_1, \ldots x_n\} \mid n\right) dx_1 \ldots dx_n \\
&= \sum_{n=0}^{\infty} p_N(n) \int_B \cdots \int_B \mathbf{1}\left[\{x_1, \ldots x_n\} \in F\right] \prod_{i=1}^{n} f_{X_i|N}\left(\{x_i\} \mid n\right) dx_1 \ldots dx_n \\
&= \sum_{n=0}^{\infty} \frac{(\mu(B))^n \exp\left(-\mu(B)\right)}{n!} \int_B \cdots \int_B \mathbf{1}\left[\{x_1, \ldots x_n\} \in F\right] \frac{1}{\mu(B)^n} \prod_{i=1}^{n} \lambda(x_i) dx_1 \ldots dx_n \\
&= \sum_{n=0}^{\infty} \frac{\exp\left(-\mu(B)\right)}{n!} \int_B \cdots \int_B \mathbf{1}\left[\{x_1, \ldots x_n\} \in F\right] \prod_{i=1}^{n} \lambda(x_i) dx_1 \ldots dx_n.
\end{aligned}$$

□

Proposition 2.4. *If X is a Poisson point process on S and $B_1, B_2, \ldots \subseteq S$ are disjoint bounded sets then*

(i) $X_{B_1}, X_{B_2}, \ldots$ are independent Poisson point processes.

(ii) $N(B_1), N(B_2), \ldots$ are independent Poisson variables.

Proof. If $X_{B_1}, X_{B_2}, \ldots$ are independent Poisson point processes, then $N(B_1), N(B_2) \ldots$ are independent and follow a Poisson distribution, hence (ii) is a direct result of (i).

Note for disjoint sets $B_1, B_2, \ldots, B_n$, it holds that

$$\mu\left(\bigcup_{i=1}^{n} B_i\right) = \mathbb{E}\left[N\left(\bigcup_{i=1}^{n} B_i\right)\right] = \mathbb{E}\left[\sum_{i=1}^{n} N(B_i)\right] = \sum_{i=1}^{n} \mathbb{E}\left[N(B_i)\right] = \sum_{i=1}^{n} \mu(B_i).$$

(i) is proven by induction so in the base case consider B_1, B_2 and X_{B_1}, X_{B_2}. Let $F, G \subset N_{lf}$ denote sets of locally finite point configurations in S. The independence between X_{B_1} and X_{B_2} comes from showing that $P\left(X_{B_1} \in F, X_{B_2} \in G\right) = P\left(X_{B_1} \in F\right) P\left(X_{B_2} \in G\right)$. Let $\int_{A^n}$ denote the integral over n points in the set A. The probability is given as the following:

$$\begin{aligned}
&P(X_{B_1} \in F, X_{B_2} \in G) \\
&\overset{(a)}{=} \sum_{n=0}^{\infty} \frac{\exp\left(-\mu\left(B_1 \cup B_2\right)\right)}{n!} \int_{(B_1 \cup B_2)^n} \mathbf{1}\left[\{x_1, \ldots, x_n\} \cap B_1 \in F\right] \\
&\quad \mathbf{1}\left[\{x_1, \ldots, x_n\} \cap B_2 \in G\right] \prod_{i=1}^{n} \lambda(x_i) dx_n, \ldots, dx_1 \\
&\overset{(b)}{=} \sum_{n=0}^{\infty} \frac{\exp\left(-\mu\left(B_1 \cup B_2\right)\right)}{n!} \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} \\
&\quad \int_{B_1^m} \mathbf{1}\left[\{x_1, \ldots, x_m\} \cap B_1 \in F\right] \int_{B_2^{n-m}} \mathbf{1}\left[\{x_{m+1}, \ldots, x_n\} \cap B_2 \in G\right] \\
&\quad \prod_{i=1}^{n} \lambda(x_i) dx_n, \ldots, dx_1 \\
&\overset{(c)}{=} \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} \frac{\exp\left(-\mu\left(B_1 \cup B_2\right)\right)}{m!(n-m)!} \int_{B_1^m} \mathbf{1}\left[\{x_1, \ldots, x_m\} \cap B_1 \in F\right] \\
&\quad \int_{B_2^{n-m}} \mathbf{1}\left[\{x_{m+1}, \ldots, x_n\} \cap B_2 \in G\right] \prod_{i=1}^{n} \lambda(x_i) dx_n, \ldots, dx_1 \\
&\overset{(d)}{=} \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\exp\left(-\mu\left(B_1 \cup B_2\right)\right)}{m!k!} \int_{B_1^m} \mathbf{1}\left[\{x_1, \ldots, x_m\} \cap B_1 \in F\right] \\
&\quad \int_{B_2^k} \mathbf{1}\left[\{x_{m+1}, \ldots, x_{m+k}\} \cap B_2 \in G\right] \prod_{i=1}^{m+k} \lambda(x_i) dx_{m+k}, \ldots, dx_1 \\
&\overset{(e)}{=} \sum_{m=0}^{\infty} \frac{\exp\left(-\mu\left(B_1\right)\right)}{m!} \int_{B_1^m} \mathbf{1}\left[\{x_1, \ldots, x_m\} \cap B_1 \in F\right] \\
&\quad \prod_{i=1}^{m} \lambda(x_i) dx_m, \ldots, dx_1 P\left(X_{B_2} \in G\right) \\
&\overset{(f)}{=} P\left(X_{B_1} \in F\right) P\left(X_{B_2} \in G\right)
\end{aligned}$$

where (a), (e) and (f) derive from Proposition 2.3, in (c) the summations are interchanged and for (d), I let $k = n-m$. Furthermore, (b) derives by splitting the integral over $(B_1 \cup B_2)^n$ into integrals over B_1^m and B_2^{m-n} separately. Note that there exist $n!/(m!(n-m)!) = \binom{n}{m}$ ways of allocating m and $m-n$ points in B_1 and B_2. Hence, when ensuring each possible integral is considered, the sum $\sum_{m=0}^{n}$ and $n!/(m!(n-m)!)$ are included.

For the induction step, assume that $X_{B_1}, \ldots, X_{B_{n-1}}$ are independent for disjoint bounded sets $B_1, \ldots, B_{n-1}$. Then if $B_1, \ldots, B_n$ are disjoint then $\tilde{B} = \bigcup_{i=1}^{n-1} B_i$ and B_n are disjoint. Then from the basis case $X_{\tilde{B}}$ and X_{B_n} are independent such that $X_{B_1}, \ldots, X_{B_n}$ are independent. □

From this property, we have the terminology of no interaction and complete spatial randomness for Poisson point processes. Because of the independencies for disjoint bounded sets, then there are also an uncorrelation such the covariance is 0. The covariance for a point process is further studied in subsection 2.5.

Later, a correlation property will be shown for the Poisson point processes, and this property is proven by the Slivnyak-Mecke theorem.

Theorem 2.5 (Slivnyak-Mecke)**.** *If $X \sim Poisson(S, \lambda)$ then for any $n \in \mathbb{N}$ and function $h : S^n \times N_{lf} \to [0, \infty)$,*

$$\mathbb{E}\left[\sum_{\substack{u_1,\ldots u_n \in X \\ u_i \neq u_j \text{ for } i \neq j}} h\left(u_1, \ldots u_n, X \backslash \{u_1, \ldots, u_n\}\right)\right] = \int_S \cdots \int_S \mathbb{E}\left[h(u_1, \ldots, u_n, X)\right] \prod_{i=1}^{n} \lambda(u_i) du_1, \ldots, du_n$$

The proof is omitted from the project, but it can be found in [1].

2.2 Intensity Function

In this subsection, I will introduce the Campbell Formula and how to non-parametric estimate the intensity function for a point process.

The general definition of the intensity measure for a point process within the set B is $\mu(B) = \mathbb{E}[N(B)]$, where the intensity function λ is the non-negative function given by the integrand in $\mu(B) = \int_B \lambda(u) du$. If $\lambda(u)$ is constant for $u \in B$, the intensity function is homogeneous and else inhomogeneous.

I will now introduce the Campbell formula, which will be used repeatedly for the point processes.

Proposition 2.6 (Campbell Formula)**.** *Let $h : \mathbb{R}^2 \to [0, \infty)$ be a function. If a point process X has intensity function λ then*

$$\mathbb{E}\left[\sum_{u \in X} h(u)\right] = \int_{\mathbb{R}^2} h(u)\lambda(u) du.$$

Proof. By the standard proof, it is sufficient to show the equality for the indicator function

$\mathbf{1}[u \in B]$ where $B \subseteq \mathbb{R}^2$. By the definition of the intensity function and measure we have

$$\mathbb{E}\left[\sum_{u \in X} \mathbf{1}[u \in B]\right] = \mathbb{E}\left[N(B)\right] = \mu(B) = \int_{\mathbb{R}^2} \mathbf{1}[u \in B]\lambda(u)du.$$

□

Different methods exist to estimate the intensity function, however, I will concentrate on a non-parametric kernel estimate for the inhomogeneous point processes and the homogeneous point processes. Let x_W be the point pattern within the window W. The estimations are made within a observed window W of the bigger state space $S \subseteq \mathbb{R}^2$, which gives some boundary problems from edge effects. Locations near the edges loses information from locations outside the window, which may have an correlation with the points within the window. Therefore treating locations near the edges as those in the centrum lead to a bias. To adjust for the problem an edge-correction factor is included in the estimation of the intensity function to less the bias [2]. The non-parametric kernel estimate for the inhomogeneous point processes is

$$\hat{\lambda}_b(u) = \sum_{v \in x_W} \frac{k_b(u-v)}{c_{W,b}(v)}, \quad u \in W,$$

where $k_b(u-v) = k\left((u-v)/b\right)/b^2$ is a kernel with bandwidth $b > 0$, and $c_{W,b}(v) = \int_W k_b(u-v)du$ is the edge-correction factor. The kernel k is a chosen density, where the choice of k is less important than the choice of b. Examples of kernels are the uniform and Epanechnikov, which are listed in appendix B. Observe that the estimated intensity measure given by $\hat{\mu}(W) = \int_W \hat{\lambda}_b(u)du$ is unbiased from the following:

$$\begin{aligned}
\mathbb{E}\left[\hat{\mu}(W)\right] = \mathbb{E}\left[\int_W \sum_{v \in x_W} \frac{k_b(u-v)}{c_{W,b}(v)}du\right] &= \int_W \mathbb{E}\left[\sum_{v \in x_W} \frac{k_b(u-v)}{c_{W,b}(v)}\right]du \\
&\overset{(a)}{=} \int_W \int_W \frac{k_b(u-v)}{c_{W,b}(v)}\lambda(v)dvdu \\
&= \int_W \lambda(v)dv = \mu(W),
\end{aligned}$$

where (a) derive from the Campbell Formula. Note that in my application, I will assume the intensity function is homogeneous. The unbiased estimate of this intensity function is given by

$$\begin{aligned}
\hat{\lambda} &= \frac{N(x_W)}{|W|} \\
\mathbb{E}\left[\hat{\lambda}\right] &= \frac{\mathbb{E}\left[N(W)\right]}{|W|} = \frac{\mu(W)}{|W|} = \frac{|W|\lambda}{|W|} = \lambda.
\end{aligned}$$

2.3 Joint Intensity and Pair Correlation Function

I will introduce in this subsection introduce a summary function called the pair correlation function. This function has the ability to indicate if a given point pattern deviates from a Poisson point process.

Before defining the pair correlation function, I will introduce the second-order product density.

Definition 2.7. Let $C \subseteq \mathbb{R}^2 \times \mathbb{R}^2$ and $\lambda^{(2)} : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty]$ be a function. The second-order factorial moment measure is

$$a^{(2)}(C) = \mathbb{E}\left[\sum_{\substack{u,v \in X \\ u \neq v}} \mathbf{1}\left[(u,v) \in C\right]\right].$$

If $a^{(2)}(C)$ can be rewritten into

$$a^{(2)}(C) = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \mathbf{1}\left[(u,v) \in C\right] \lambda^{(2)}(u,v) du dv,$$

then $\lambda^{(2)}$ is called the second-order product density.

Note that if $\lambda^{(2)}$ exists, then with the standard proof, we can also express the second-order Campbell formula

$$\mathbb{E}\left[\sum_{\substack{u,v \in X \\ u \neq v}} h(u,v)\right] = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} h(u,v)\lambda^{(2)}(u,v) du dv,$$

where $h : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$.

Definition 2.8. If both λ and $\lambda^{(2)}$ exist, the pair correlation function is defined as

$$g(u,v) := \frac{\lambda^{(2)}(u,v)}{\lambda(u)\lambda(v)},$$

where $g(u,v) = 0$ when $\lambda(u)\lambda(v) = 0$.

Proposition 2.9. *$X \sim$ Poisson(S, λ), and both λ and $\lambda^{(2)}$ exist then $\lambda^{(2)}(u,v) = \lambda(u)\lambda(v)$.*

Proof. Let $C = B_1 \times B_2$ where $B_1, B_2 \subseteq S$ then

$$\begin{aligned}
\int_S\int_S \mathbf{1}\left[(u,v) \in C\right] \lambda(u)\lambda(v) du dv &= \int_S\int_S \mathbb{E}\left[\mathbf{1}\left[(u,v) \in C\right]\right] \lambda(u)\lambda(v) du dv \\
&\overset{(a)}{=} \mathbb{E}\left[\sum_{\substack{u,v \in X \\ u \neq v}} \mathbf{1}\left[(u,v) \in C\right]\right] \\
&\overset{(b)}{=} \int_S\int_S \mathbf{1}\left[(u,v) \in C\right] \lambda^{(2)}(u,v) du dv,
\end{aligned}$$

where (a) derives from Slivnyak-Mecke Theorem 2.5 and (b) from second-order Campbell formula. Hence

$$\lambda(u)\lambda(v) = \lambda^{(2)}(u,v).$$

□

By the proposition, if we have a Poisson point process then $g(u,v) = 1$. On the contrary, $g(u,v) > 1$ is an indiction that a pair of points are more likely to occur at the same location than for a Poisson point process with the same intensity function and less likely

if $g(u,v) < 1$. A typical assumption throughout the project is that the pair correlation function satisfies $g(u,v) = g(\|u-v\|)$ (Isotropy). Hence, the correlation function doesn't depend on the points' specific locations but on the distances between the pairs. I will typically use the notation $r = \|u-v\|$ for this. For small values of r, $g(r) > 1$ indicates clustering of the points at a distance of r, while $g(r) < 1$ indicates repulsion. Also, under the isotropy assumption, we have an edge-corrected kernel estimate for the pair correlation function given as

$$\hat{g}(r) = \frac{1}{2\pi r} \sum_{\substack{u,v \in x_W \\ u \neq v}} \frac{k_b(r - \|v-u\|)}{\hat{\lambda}(u)\hat{\lambda}(v)\,|W \cap W_{v-u}|},$$

where $k_b(r - \|v-u\|) = k\left((r - \|v-u\|)/b\right)/b$ with kernel $k(\cdot)$ and bandwidth $b > 0$, and it is assumed that $\|v-u\| \leq r$ such that $|W \cap W_{v-u}| > 0$. Note that the choice of b is more impactful than that of $k_b(\cdot)$.

Proposition 2.10. *Let $\lambda(u)\lambda(v)$ be known. If b is sufficiently small, $\hat{g}(r)$ is close to $g(r)$, and if r is also sufficiently small, the estimator is upwards biased.*

Proof. The expected value of the estimator is

$$\begin{aligned}
\mathbb{E}\left[\hat{g}(r)\right] &= \mathbb{E}\left[\frac{1}{2\pi r} \sum_{\substack{u,v \in x_W \\ u \neq v}} \frac{k_b(r - \|v-u\|)}{\lambda(u)\lambda(v)\,|W \cap W_{v-u}|}\right] \\
&= \frac{1}{2\pi r}\mathbb{E}\left[\sum_{\substack{u,v \in x_W \\ u \neq v}} \frac{k_b(r - \|v-u\|)}{|W \cap W_{v-u}|\,\lambda^{(2)}(v,u)} g(u,v)\right] \\
&\overset{(a)}{=} \frac{1}{2\pi r}\int_W \int_W \frac{k_b(r - \|v-u\|)}{|W \cap W_{v-u}|\,\lambda^{(2)}(v,u)} \lambda^{(2)}(v,u) g(u,v) du dv \\
&= \frac{1}{2\pi r}\int_W \int_W \frac{k_b(r - \|v-u\|)}{|W \cap W_{v-u}|} g(\|v-u\|) du dv \\
&= \frac{1}{2\pi r}\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{k_b(r - \|v-u\|)}{|W \cap W_{v-u}|} g(\|v-u\|) \mathbf{1}[v \in W, u \in W] du dv,
\end{aligned}$$

where (a) derives from the second-order Campbell formula. Let $w = v - u$ where $dv = dw$ and substitute v with w in the integral such that

$$\begin{aligned}
\mathbb{E}\left[\hat{g}(r)\right] &= \frac{1}{2\pi r}\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{k_b(r - \|w\|)}{|W \cap W_w|} g(\|w\|) \mathbf{1}[w + u \in W, u \in W] du dw \\
&= \frac{1}{2\pi r}\int_{\mathbb{R}^2} \frac{k_b(r - \|w\|)}{|W \cap W_w|} g(\|w\|) \int_{\mathbb{R}^2} \mathbf{1}[u \in W_{-w}, u \in W] du dw \\
&= \frac{1}{2\pi r}\int_{\mathbb{R}^2} \frac{k_b(r - \|w\|)}{|W \cap W_w|} g(\|w\|)\,|W \cap W_{-w}|\, dw \\
&= \frac{1}{2\pi r}\int_{\mathbb{R}^2} \frac{k_b(r - \|w\|)}{|W \cap W_w|} g(\|w\|)\,|W \cap W_w|\, dw \\
&= \frac{1}{2\pi r}\int_{\mathbb{R}^2} k_b(r - \|w\|) g(\|w\|) dw.
\end{aligned}$$

Let $\tilde{r} = \|w\|$, such that $w = (w_1, w_2) = (\tilde{r}\cos(s), \tilde{r}\sin(s))$, and we have the following Jacobian

$$J_w = \begin{bmatrix} \frac{dw_1}{d\tilde{r}} & \frac{dw_1}{ds} \\ \frac{dw_2}{d\tilde{r}} & \frac{dw_2}{ds} \end{bmatrix} = \begin{bmatrix} \cos(s) & -\tilde{r}\sin(s) \\ \sin(s) & \tilde{r}\cos(s) \end{bmatrix}, \quad \det(J_w) = \tilde{r}\cos^2(s) + \tilde{r}\sin^2(s) = \tilde{r}.$$

Then

$$\mathbb{E}\left[\hat{g}(r)\right] = \frac{1}{2\pi r}\int_0^\infty \int_0^{2\pi} k_b(r-\tilde{r})g(\tilde{r})\tilde{r}dsd\tilde{r} = \int_0^\infty k_b(r-\tilde{r})g(\tilde{r})\frac{\tilde{r}}{r}d\tilde{r}.$$

Let $t = \frac{r-\tilde{r}}{b}$, such that $\tilde{r} = r - tb$ and $-\frac{1}{b}d\tilde{r} = dt$. Since the kernel functions have support in $[-1,1]$, $k(t)$ will take values other than 0 for $t \in \left[-1, \min(1, \frac{r}{b})\right]$. Hence

$$\begin{aligned}
\int_0^\infty g\left(\tilde{r}\right) k_b\left(r-\tilde{r}\right)\frac{\tilde{r}}{r}d\tilde{r} &= -\frac{1}{r}\int_{\min\left(1,\frac{r}{b}\right)}^{-1} g\left(r-tb\right)k(t)\left(r-tb\right)dt \\
&= \frac{1}{r}\int_{-1}^{\min\left(1,\frac{r}{b}\right)} g\left(r-tb\right)k(t)\left(r-tb\right)dt \\
&= \frac{b}{r}\int_{-1}^{\min\left(1,\frac{r}{b}\right)} g\left(r-tb\right)k(t)\left(\frac{r}{b}-t\right)dt.
\end{aligned}$$

For small b

$$\frac{b}{r}\int_{-1}^{\min\left(1,\frac{r}{b}\right)} g\left(r-tb\right)k(t)\left(\frac{r}{b}-t\right)dt \approx g\left(r\right)\frac{b}{r}\int_{-1}^{\min\left(1,\frac{r}{b}\right)} k(t)\left(\frac{r}{b}-t\right)dt.$$

For $b < r$

$$\lim_{b\to 0} g\left(r\right)\frac{1}{r}\int_{-1}^{1} k(t)\left(r-tb\right)dt = g\left(r\right)\frac{1}{r}\int_{-1}^{1} k(t)rdt = g\left(r\right).$$

Let $0 < r < b$. If r is so small that $\frac{b}{r}\int_0^1 k(-t)tdt \geq 1$, then

$$\begin{aligned}
\mathbb{E}\left[\hat{g}(r)\right] &= g\left(r\right)\frac{b}{r}\int_{-1}^{\frac{r}{b}} k(t)\left(\frac{r}{b}-t\right)dt \\
&= g\left(r\right)\frac{b}{r}\int_{-\frac{r}{b}}^{1} k(-t)\left(\frac{r}{b}+t\right)dt \\
&= g\left(r\right)\frac{b}{r}\left(\int_{-\frac{r}{b}}^{0} k(-t)\frac{r}{b}dt + \int_{-\frac{r}{b}}^{0} k(-t)tdt + \int_0^1 k(-t)\frac{r}{b}dt + \int_0^1 k(-t)tdt\right) \\
&\overset{(a)}{\geq} g\left(r\right)\frac{b}{r}\left(\int_{-\frac{r}{b}}^{0} k(-t)\frac{r}{b}dt - \int_{-\frac{r}{b}}^{0} k(-t)\frac{r}{b}dt + \int_0^1 k(-t)tdt\right) \\
&= g\left(r\right)\frac{b}{r}\int_0^1 k(-t)tdt \\
&\geq g\left(r\right),
\end{aligned}$$

where (a) derives from $\int_{-\frac{r}{b}}^0 k(-t)tdt \geq -\int_{-\frac{r}{b}}^0 k(-t)\frac{r}{b}dt$ and $\int_0^1 k(-t)\frac{r}{b}dt \geq 0$.

Hence, the estimate $\hat{g}(r)$ is upward biased for a sufficiently small r. Also, for a sufficiently small b, we have $\hat{g}(r) \approx g(r)$. □

In the case with a homogeneous intensity (more specific stationary point process), then it has been suggested to use $b = 0.15/\sqrt{\hat{\lambda}}$ for the Epanecnikov kernel and $b = 0.1/\sqrt{\hat{\lambda}}$ for the uniform kernel for $\hat{g}$ [3].

2.4 Cox Point Process

This subsection introduces an extension of the Poisson point process called the Cox point process and a special case called the log Gaussian Cox point process. This subsection is besides [1] also based on [4].

Before introducing a Cox point process, I will first give the definition of a random field in the Euclidean plane.
Definition 2.11. Let $(\Omega, \mathcal{C}, P)$ be a probability space and $S \subseteq \mathbb{R}^2$. Then, the set $R = \{R(\omega, u) : \omega \in \Omega, u \in S\}$ of random variables is called a random field.

Usually, $R(\cdot, u)$ is typically written as $R(u)$. If the intensity function in the Poisson point process is a realisation of a random field, the process is called a Cox point process.
Definition 2.12. Let $\Lambda = \{\Lambda(u) : u \in S\}$ be a non-negative random field with $P\left(\int_B \Lambda(u)du < \infty\right) = 1$ for each bounded $B \subseteq S$. If $X \mid \Lambda \sim \text{Poisson}(S, \Lambda)$, then X is said to be a Cox point process driven by Λ.

If $\mathbb{E}[\Lambda(u)] = \lambda(u)$ exists, then $\lambda(u)$ is the intensity function. Furthermore, the intensity measure for $X \mid \Lambda$ is a random measure given as

$$M(B) = \int_B \Lambda(u)du.$$

If $\text{Var}(\Lambda(u)) < \infty$ for all $u \in S$, the pair correlation function is

$$g(u, v) = \frac{\mathbb{E}\left[\Lambda(u)\Lambda(v)\right]}{\lambda(u)\lambda(v)},$$

where $\mathbb{E}\left[\Lambda(u)\Lambda(v)\right] = \lambda^{(2)}(u, v)$.

The random variables in a random field may follow a specific distribution e.g. a Gaussian distribution.
Definition 2.13. Let $R = \{R(u) : u \in S\}$ be a random field. If

$$\left[R\left(u_i\right)\right]_{1 \le i \le n} \sim \mathcal{N}\left(\left[\mathbb{E}\left[R\left(u_i\right)\right]\right]_{1 \le i \le n}, \left[\text{Cov}\left(R\left(u_i\right), R\left(u_j\right)\right)\right]_{1 \le i,j \le n}\right)$$

for any $n \in \mathbb{N}$ and subsets $\{u_1, \ldots, u_n\} \subseteq S$ then R is said to be a Gaussian random field (GRF).

With GRF a special case of the Cox point process can be defined called the log Gaussian Cox point process.
Definition 2.14. Let X be the Cox point process driven by $\Lambda = \exp(Y)$, where Y is a GRF. Then, X is said to be a log Gaussian Cox point process (LGCP).

Denote the mean and covariance function for the GRF Y by $m(u) = \mathbb{E}\left[Y(u)\right]$ and $C(u, v) = \text{Cov}\left(Y(u), Y(v)\right)$ for $u, v \in S$. When $m(u) = \mu$ is constant, we call the LGCP homogeneous.
Proposition 2.15. *The intensity and pair correlation functions for the LGCP are*

$$\lambda(u) = \exp\left(m(u) + \frac{C(u, u)}{2}\right), \quad g(u, v) = \exp\left(C(u, v)\right).$$

Proof. The Laplace transform of $Y(u) \sim \mathcal{N}\left(m(u), C(u, u)\right)$ is

$\mathbb{E}[\exp(Yt)] = \exp\left(m(u)t + \frac{C(u,u)t^2}{2}\right)$. Hence

$$\begin{aligned}
\lambda(u) = \mathbb{E}\left[\Lambda(u)\right] = \mathbb{E}\left[\exp\left(Y(u)\right)\right] &= \exp\left(m(u) + \frac{C(u,u)}{2}\right), \\
\mathbb{E}\left[\Lambda(u)\Lambda(v)\right] &= \mathbb{E}\left[\exp(Y(u) + Y(v))\right] \\
&= \exp\left(m(u) + m(v) + \frac{C(u,u) + C(v,v) + 2C(u,v)}{2}\right) \\
&= \lambda(u)\lambda(v)\exp\left(C(u,v)\right),
\end{aligned}$$

such that

$$g(u,v) = \frac{\mathbb{E}\left[\exp\left(\Lambda(u) + \Lambda(v)\right)\right]}{\lambda(u)\lambda(v)} = \exp\left(C(u,v)\right).$$

□

2.5 Covariance for Point Process.

In this subsection, I examine the covariance between two functions of the same point process. I will specific look on the cases with Poisson point process and LGCP.

Let $h_1, h_2 : S \to [0,\infty)$ be two non-negative functions on $S \subseteq \mathbb{R}^2$ and X a point process defined on S. Then

$$\begin{aligned}
\operatorname{Cov}\left(\sum_{v\in X} h_1(v), \sum_{v\in X} h_2(v)\right) &= \mathbb{E}\left(\sum_{v\in X} h_1(v)\sum_{v\in X} h_2(v)\right) - \mathbb{E}\left(\sum_{v\in X} h_1(v)\right)\mathbb{E}\left(\sum_{v\in X} h_2(v)\right) \\
&= \mathbb{E}\left(\sum_{\substack{u,v\in X \\ u\neq v}} h_1(u)h_2(v)\right) + \mathbb{E}\left(\sum_{v\in X} h_1(v)h_2(v)\right) \\
&\quad - \mathbb{E}\left(\sum_{v\in X} h_1(v)\right)\mathbb{E}\left(\sum_{v\in X} h_2(v)\right) \\
&\overset{(a)}{=} \int_S\int_S h_1(u)h_2(v)\lambda^{(2)}(u,v)dudv + \int_S h_1(v)h_2(v)\lambda(v)dv \\
&\quad - \int_S h_1(u)\lambda(u)du\int_S h_2(v)\lambda(v)dv \\
&= \int_S\int_S h_1(u)h_2(v)\left(\lambda^{(2)}(u,v) - \lambda(u)\lambda(v)\right)dudv \\
&\quad + \int_S h_1(v)h_2(v)\lambda(v)dv,
\end{aligned}$$

where equality (a) is derived from Campbell's two formulas. Observe, if X is a Poisson point process, then $\lambda^{(2)}(u,v) = \lambda(u)\lambda(v)$ such that $\operatorname{Cov}\left(\sum_{v\in X} h_1(v), \sum_{v\in X} h_2(v)\right) = \int_S h_1(v)h_2(v)\lambda(v)dv = \mathbb{E}\left[\sum_{v\in X} h_1(v)h_2(v)\right]$.

Let $B_1, B_2 \subseteq S$ be bounded sets. The covariance between the number of points in each set

is

$$\begin{aligned}\text{Cov}\left(N(B_1), N(B_2)\right) &= \text{Cov}\left(\sum_{v\in X}\mathbf{1}[v\in B_1], \sum_{v\in X}\mathbf{1}[v\in B_2]\right)\\ &= \int_S\int_S \mathbf{1}[v\in B_1, u\in B_2]\left(\lambda^{(2)}(u,v)-\lambda(u)\lambda(v)\right)dudv\\ &+ \int_S \mathbf{1}[v\in B_1, v\in B_2]\lambda(v)dv.\end{aligned}$$

If X follows a Poisson point process, then $\text{Cov}\left(N(B_1), N(B_2)\right) = \int_S \mathbf{1}[v\in B_1, v\in B_2]\lambda(v)dv$ and $\text{Var}\left(N(B)\right) = \int_B \lambda(v)dv = \mathbb{E}\left[N(B)\right]$ for any bounded set $B\subseteq S$. Furthermore, if B_1 and B_2 are disjoint, then $\mathbf{1}[v\in B_1, v\in B_2] = 0$ such that $\text{Cov}\left(N(B_1), N(B_2)\right) = 0$.

If X follows a Cox point process, then

$$\begin{aligned}\text{Cov}\left(N(B_1), N(B_2)\right) &= \int_S\int_S \mathbf{1}[v\in B_1, u\in B_2]\left(\lambda^{(2)}(u,v)-\lambda(u)\lambda(v)\right)dudv\\ &+ \int_S \mathbf{1}[v\in B_1, v\in B_2]\lambda(v)dv\\ &= \int_S\int_S \mathbf{1}[v\in B_1, u\in B_2]\left(\mathbb{E}\left[\Lambda(u)\Lambda(v)\right]-\mathbb{E}\left[\Lambda(u)\right]\mathbb{E}\left[\Lambda(v)\right]\right)dudv\\ &+ \mathbb{E}\left[\sum_{v\in X}\mathbf{1}[v\in B_1, v\in B_2]\right]\\ &= \int_S\int_S \mathbf{1}[v\in B_1, u\in B_2]\text{Cov}\left(\Lambda(u),\Lambda(v)\right)dudv + \mathbb{E}\left[N\left(B_1\cap B_2\right)\right].\end{aligned}$$

If B_1 and B_2 are disjoint, then

$$\text{Cov}\left(N(B_1), N(B_2)\right) = \int_S\int_S \mathbf{1}[v\in B_1, u\in B_2]\text{Cov}\left(\Lambda(u),\Lambda(v)\right)dudv.$$

If also $\Lambda(u)$ and $\Lambda(v)$ are uncorrelated for all $u\in B_1$ and $v\in B_2$ then $\text{Cov}\left(N(B_1), N(B_2)\right) = 0$. Furthermore, let $B\subseteq S$ be a bounded set, then the Cox point process $\text{Var}\left(N(B)\right) = \int_B\int_B \text{Cov}\left(\Lambda(u),\Lambda(v)\right)dudv + \mathbb{E}\left[N(B)\right]$, such that $\text{Var}\left(N(B)\right)\geq \mathbb{E}\left[N(B)\right]$.

3 Covariance

Given the tree locations within a window $W \subseteq \mathbb{R}^2$ of a rainforest, we can assess the number of trees and their locations as a spatial point process X_W. Besides the tree location, we may have more information about the area, which could arise as a binary or continuous random variable. Denote the binary as $I(v)$ and the continuous as $J(v)$ for location $v \in \mathbb{R}^2$. An example of a continuous random variable could be the pH value of the soil, and it can be transformed into a binary variable by e.g.

$$I(v) = \begin{cases} 0 & \text{If pH}(v) \geq 7 \text{ such that the soil at location } v \text{ is non-acidic.} \\ 1 & \text{If pH}(v) < 7 \text{ such that the soil at location } v \text{ is acidic.} \end{cases},$$

where $I(v)$ follows a Bernoulli distribution with $P(I(v) = 1) = p(v)$.

A question could arise whether the random variables are correlated to the point process of the trees. In subsections 3.1 and 3.2, I will deduce the covariances between a point process and the random variables, which are either binary or continuous. From the covariances, I derive two functions, c_{bin} and c_{con}, which are used as summary functions for the correlation between the point process and random variables. Both c_{bin} and c_{con} will hereafter be given two estimators and studied for a special case of LGCP.

3.1 Covariance Between Spatial Point Process and Binary Variable

This subsection is inspired by the earlier draft of [5] and will deduce the covariance between a spatial point process and a binary random variable.

Let X be a spatial point process within $S \subseteq \mathbb{R}^2$ and $I(v)$ be a binary random variable at location $v \in \mathbb{R}^2$ with $P(I(v) = 1) = p(v)$ such that $\mathbb{E}[I(v)] = p(v)$. Assume that both the intensities of X and the conditional intensities of X given $I(v)$ exist and denote them by $\lambda(\cdot)$ and $\lambda_{I(v)}(\cdot \mid v)$ respectively. Let $c_{\text{bin}}(u, v) = \lambda_1(u \mid v)/\lambda(u)$ be a normalised moment density of X and $I(v)$. Assume that $h(u)$ is a non-negative function, then

$$\begin{aligned}
\mathbb{E}\left[I(v)\sum_{u\in X} h(u)\right] &= \mathbb{E}\left[I(v)\mathbb{E}\left[\sum_{u\in X} h(u) \mid I(v)\right]\right] \\
&\overset{(a)}{=} \mathbb{E}\left[I(v)\int_S h(u)\lambda_{I(v)}(u \mid v)du\right] \\
&= 1p(v)\int_S h(u)\lambda_1(u \mid v)du + 0(1-p(v))\int_S h(u)\lambda_0(u \mid v)du, \\
&= p(v)\int_S h(u)\lambda_1(u \mid v)du,
\end{aligned}$$

where Campbell formula is used for (a). The covariance between the spatial point process

and the binary random variable is

$$\begin{aligned}\operatorname{Cov}\left[I(v), \sum_{u\in X} h(u)\right] &= \mathbb{E}\left[I(v)\sum_{u\in X} h(u)\right] - \mathbb{E}\left[I(v)\right]\mathbb{E}\left[\sum_{u\in X} h(u)\right] \\ &\overset{(a)}{=} p(v)\int_S h(u)\lambda_1(u \mid v)du - p(v)\int_S h(u)\lambda(u)du \\ &= p(v)\int_S h(u)\left(\lambda_1(u \mid v) - \lambda(u)\right)du \\ &= p(v)\int_S h(u)\lambda(u)\left(c_{\text{bin}}(u,v) - 1\right)du,\end{aligned}$$

where (a) derives from Campbell formula. If $c_{\text{bin}}(u,v) = 1$ for all $u \in S$, there is uncorrelation between $I(v)$ and $\sum_{u\in X} h(u)$. If $c_{\text{bin}}(u,v) < 1$ for all $u \in S$, then $\operatorname{Cov}\left[I(v), \sum_{u\in X} h(u)\right] < 0$ such that X and $I(v)$ are negatively correlated. If $c_{\text{bin}}(u,v) > 1$ for all $u \in S$, the opposite is the case. Hence, the $c_{\text{bin}}(u,v)$ works as a summary function for the correlation between the point process and binary random variable. The question is how to estimate $c_{\text{bin}}(u,v)$, which is explored in the following.

3.1.1 Estimates of c_{bin}

Let for a finite known set of locations $\mathcal{T} \subseteq \mathbb{R}^2$, the binary variable $I(v)$ have realisation $i(v)$ for each $v \in \mathcal{T}$, and let $X_W = x_W$ be a realised point pattern within a window W with intensity $\lambda(u)$. Assume that $c_{\text{bin}}(u,v) = c_{\text{bin},0}(\|u-v\|) = c_{\text{bin},0}(r)$. Then, an estimate of $c_{\text{bin},0}(r)$ could be given by

$$\hat{c}_{\text{bin},0}(r) = \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} i(v) \sum_{\substack{u\in x_W \\ u\neq v}} \frac{k_b(r - \|u-v\|)}{2\pi\|u-v\|\, p(v)\lambda(u)}.$$

In the following, the bias of this estimate is explored:

$$\begin{aligned}\mathbb{E}\left[\hat{c}_{\text{bin},0}(r)\right] &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} \mathbb{E}\left[I(v) \sum_{\substack{u\in X_W \\ u\neq v}} \frac{k_b\left(r - \|u-v\|\right)}{2\pi\|u-v\|\, p(v)\lambda(u)}\right] \\ &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} \mathbb{E}\left[I(v)\mathbb{E}\left[\sum_{\substack{u\in X_W \\ u\neq v}} \frac{k_b\left(r - \|u-v\|\right)}{2\pi\|u-v\|\, p(v)\lambda(u)} \mid I(v)\right]\right] \\ &\overset{(a)}{=} \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} \mathbb{E}\left[I(v)\int_W \frac{k_b\left(r - \|u-v\|\right)}{2\pi\|u-v\|\, p(v)\lambda(u)}\lambda_{I(v)}(u \mid v)du\right] \\ &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} p(v)\int_W \frac{k_b\left(r - \|u-v\|\right)}{2\pi\|u-v\|\, p(v)\lambda(u)}\lambda_1(u \mid v)du \\ &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} \int_W \frac{k_b\left(r - \|u-v\|\right)c_{\text{bin},0}\left(\|u-v\|\right)}{2\pi\|u-v\|}du \\ &\overset{(b)}{=} \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} \int_0^\infty \int_0^{2\pi} \frac{k_b\left(r-\tilde{r}\right)c_{\text{bin},0}\left(\tilde{r}\right)}{2\pi\tilde{r}}\mathbf{1}[v + \tilde{r}\left(\cos(s),\sin(s)\right) \in W]\tilde{r}dsd\tilde{r} \\ &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}} \int_0^\infty \int_0^{2\pi} \frac{k_b\left(r-\tilde{r}\right)c_{\text{bin},0}\left(\tilde{r}\right)}{2\pi}\mathbf{1}[v + \tilde{r}\left(\cos(s),\sin(s)\right) \in W]dsd\tilde{r},\end{aligned}$$

where Campbell formula was used in (a). The multiple variables integral substitution was used in (b) where the Jacobian of $u = (u_1, u_2) = v + \|v - u\| (\cos(s), \sin(s)) = v + \tilde{r} (\cos(s), \sin(s))$ is

$$J_u = \begin{bmatrix} \cos(s) & -\tilde{r}\sin(s) \\ \sin(s) & \tilde{r}\cos(s) \end{bmatrix}, \quad \det(J_u) = \tilde{r}.$$

Assume for all $v \in \mathcal{T}$ that $v + \tilde{r} (\cos(s), \sin(s)) \in W$, which could happen in the case of a sufficiently small $\tilde{r}$ and $\mathcal{T} \subseteq W$. Then

$$\begin{aligned}
&\frac{1}{N(\mathcal{T})} \sum_{v \in \mathcal{T}} \int_0^\infty \int_0^{2\pi} \frac{k_b (r - \tilde{r}) c_{\text{bin},0} (\tilde{r})}{2\pi} \mathbf{1}[v + \tilde{r} (\cos(s), \sin(s)) \in W] ds d\tilde{r} \\
&= \frac{1}{N(\mathcal{T})} \sum_{v \in \mathcal{T}} \int_0^\infty \int_0^{2\pi} \frac{k_b (r - \tilde{r}) c_{\text{bin},0} (\tilde{r})}{2\pi} ds d\tilde{r} \\
&= \frac{1}{N(\mathcal{T})} N(\mathcal{T}) \int_0^\infty k_b (r - \tilde{r}) c_{\text{bin},0} (\tilde{r}) d\tilde{r} \\
&= \int_0^\infty k_b (r - \tilde{r}) c_{\text{bin},0} (\tilde{r}) d\tilde{r} \\
&\approx c_{\text{bin},0} (r) \int_0^\infty k_b (r - \tilde{r}) d\tilde{r},
\end{aligned}$$

where in the approximation, I used that when b is sufficiently small, then $r \approx \tilde{r}$. Let $t = \frac{r - \tilde{r}}{b}$, such that $\tilde{r} = r - tb$ and $-\frac{1}{b} d\tilde{r} = dt$. Since the kernel functions are defined on $[-1, 1]$, $k(t)$ will take values from $t \in \left[-1, \min(1, \frac{r}{b})\right]$. Hence,

$$c_{\text{bin},0} (r) \int_0^\infty k_b (r - \tilde{r}) d\tilde{r} = -c_{\text{bin},0} (r) \int_{\min\left(1, \frac{r}{b}\right)}^{-1} k (t) dt = c_{\text{bin},0} (r) \int_{-1}^{\min\left(1, \frac{r}{b}\right)} k (t) dt.$$

Then for $r < b$

$$c_{\text{bin},0} (r) \int_{-1}^{\frac{r}{b}} k (t) dt \leq c_{\text{bin},0}(r),$$

while for $r \geq b$

$$c_{\text{bin},0} (r) \int_{-1}^{1} k (t) dt = c_{\text{bin},0}(r).$$

Hence, when b is sufficiently small, the estimate is downward biased if $r < b$, while it is unbiased for $r \geq b$. When r is sufficiently big, the assumption that $v + \tilde{r} (\cos(s), \sin(s)) \in W$ for all $v \in \mathcal{T}$ is violated even though $\mathcal{T} \subseteq \mathcal{W}$. For this reason an edge-correction factor could be included. Let

$$e (v, \|u - v\|) = \frac{1}{\int_0^{2\pi} \mathbf{1} [v + \|u - v\| (\cos(s), \sin(s)) \in W] ds}, \tag{3.1}$$

and the edge-corrected estimate given by

$$\tilde{c}_{\text{bin},0}(r) = \frac{1}{N(\mathcal{T})} \sum_{v \in \mathcal{T}} I(v) \sum_{\substack{u \in X_W \\ u \neq v}} \frac{k_b (r - \|u - v\|) e (v, \|u - v\|)}{\|u - v\| p(v) \lambda(u)}.$$

Beside including $e\left(v, \|u-v\|\right)$ in the estimate, 2π has also been removed from the denominator. Then

$$\begin{aligned}\mathbb{E}\left[\tilde{c}_{\text{bin},0}(r)\right] &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}}\int_0^\infty k_b\left(r-\tilde{r}\right)c_{\text{bin},0}\left(\tilde{r}\right)e\left(v,\tilde{r}\right)\int_0^{2\pi}\mathbf{1}[v+\tilde{r}\left(\cos(s),\sin(s)\right)\in W]dsd\tilde{r}\\ &= \frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}}\int_0^\infty k_b\left(r-\tilde{r}\right)c_{\text{bin},0}\left(\tilde{r}\right)d\tilde{r}.\end{aligned}$$

Note that $e\left(v, \|u-v\|\right)$ is the sum of arc angles within W and centrum v.

I will in my simulation study and application estimate p by the sample mean and λ by the homogenous estimate.

3.2 Covariance between Point Process and Continuous Variable

I will in this subsection consider the random variable as continuous.

Let $J(v)$ be a continuous variable at $v\in\mathbb{R}^2$ with density $f_v(j)$. Let X be the point process in $S\subseteq\mathbb{R}^2$ with intensity $\lambda(u)$ at $u\in X$ and $\lambda_J(u\mid v)$ the intensity at $u\in X$ given the continuous variable $J(v)$ at $v\in S$. Furthermore, let $c_J(u,v)=\frac{\lambda_J(u|v)}{\lambda(u)}$ and $c_{\text{con}}(u,v)=\int_{\mathbb{R}} jf_v(j)c_j(u,v)dj$. Then

$$\begin{aligned}\mathbb{E}\left[J(v)\sum_{u\in X}h(u)\right] &= \mathbb{E}\left[J(v)\mathbb{E}\left[\sum_{u\in X}h(u)\mid J(v)\right]\right]\\ &= \mathbb{E}\left[J(v)\int_S h(u)\lambda_J(u\mid v)du\right]\\ &= \mathbb{E}\left[J(v)\int_S h(u)\lambda(u)c_{J(v)}(u,v)du\right]\\ &= \int_{\mathbb{R}}\int_S h(u)\lambda(u)c_j(u,v)jf_v(j)dudj\\ &= \int_S h(u)\lambda(u)c_{\text{con}}(u,v)du,\end{aligned}\tag{3.2}$$

such that

$$\begin{aligned}\text{Cov}\left(J(v),\sum_{u\in X}h(u)\right) &= \mathbb{E}\left[J(v)\sum_{u\in X}h(u)\right]-\mathbb{E}[J(v)]\mathbb{E}\left[\sum_{u\in X}h(u)\right]\\ &= \int_S h(u)\lambda(u)c_{\text{con}}(u,v)du-\mathbb{E}[J(v)]\int_S h(u)\lambda(u)du\\ &= \int_S h(u)\lambda(u)\left(c_{\text{con}}(u,v)-\mathbb{E}\left[J(v)\right]\right)du.\end{aligned}$$

The type of correlation between $J(v)$ and $\sum_{u\in X}h(u)$ is deduced by $c_{\text{con}}(u,v)$ and $\mathbb{E}\left[J(v)\right]$. If $c_{\text{con}}(u,v)>\mathbb{E}\left[J(v)\right]$, they are positively correlated, if $c_{\text{con}}(u,v)<\mathbb{E}\left[J(v)\right]$, they are negatively correlated and if $c_{\text{con}}(u,v)=\mathbb{E}\left[J(v)\right]$, they are uncorrelated.

Assume that $f_v(j)=f(j)$ is independent of the location v. Then, the non-parametric method to estimate $\mathbb{E}\left[J(v)\right]=\int_{\mathbb{R}} jf(j)dj=\mathbb{E}\left[J\right]$ is by the sample mean given by $\hat{E}\left[J\right]=\frac{1}{N(\mathcal{T})}\sum_{v\in\mathcal{T}}j(v)$, where $\mathcal{T}\subseteq\mathbb{R}^2$ is a known set of locations and $J(v)$ has realisation $j(v)$ for $v\in\mathcal{T}$. In the following two estimates of $c_{\text{con}}(u,v)$ are deduced.

3.2.1 Estimates of c_{con}

Assume again that $f_v(j) = f(j)$ is independent of the location v. Assume also that $c_{\text{con}}(u,v) = c_{\text{con},0}(\|u-v\|) = c_{\text{con},0}(r)$. Then, an estimator of $c_{\text{con},0}(r)$ could be

$$\hat{c}_{\text{con},0}(r) = \frac{1}{N(\mathcal{T})} \sum_{v \in \mathcal{T}} j(v) \sum_{u \in x_W} \frac{k_b(r - \|u-v\|)}{2\pi r \lambda(u)} = \frac{1}{N(\mathcal{T})} \sum_{v \in \mathcal{T}} \hat{c}_{v,0}(r),$$

where x_W is the point pattern within the window W. The expected value of each entry in the estimator's first sum is

$$\begin{aligned}
\mathbb{E}\left[\hat{c}_{v,0}(r)\right] &= \mathbb{E}\left[J(v) \sum_{u \in X_W} \frac{k_b(r - \|u-v\|)}{2\pi r \lambda(u)} \right] \\
&\overset{(a)}{=} \int_W \frac{k_b(r - \|u-v\|)}{2\pi r \lambda(u)} \lambda(u) c_{\text{con}}(u,v) du \\
&= \frac{1}{2\pi r} \int_W k_b(r - \|u-v\|) c_{\text{con},0}(\|u-v\|) du \\
&\overset{(b)}{=} \frac{1}{2\pi r} \int_0^{\infty} \int_0^{2\pi} k_b(r-\tilde{r}) c_{\text{con},0}(\tilde{r}) \tilde{r} \, ds \, d\tilde{r} \\
&= \int_0^{\infty} k_b(r-\tilde{r}) c_{\text{con},0}(\tilde{r}) \frac{\tilde{r}}{r} d\tilde{r},
\end{aligned}$$

where equality (a) derives from (3.2) with $h(u) = k_b\left(r - \|u-v\|\right) / \left(2\pi r \lambda(u)\right)$ and for equality (b), I used the substitution $u = v + \tilde{r}\left(\cos(s), \sin(s)\right)$ and assumed for all $v \in \mathcal{T}$ that $v + \tilde{r}\left(\cos(s), \sin(s)\right) \in W$, as in subsection 3.1.

By equivalent arguments, as in the ending of the proof for proposition 2.10, when b is sufficiently small, $\hat{c}_{v,0}(r) \approx c_{v,0}(r)$, and if r is sufficiently small, the estimate is upwards biased. These biases are satisfied for $\hat{c}_{\text{con},0}(r)$ as well.

Again the edge-correction factor from (3.1) can be included to remove the assumption that $v + \tilde{r}\left(\cos(s), \sin(s)\right) \in W$ for all $v \in \mathcal{T}$ such the edge-corrected estimate is

$$\tilde{c}_{\text{con},0}(r) = \frac{1}{N(\mathcal{T})} \sum_{v \in \mathcal{T}} J(v) \sum_{u \in X_W} \frac{k_b\left(r - \|u-v\|\right) e\left(v, \|u-v\|\right)}{r \lambda(u)}.$$

3.3 c_{bin} and c_{con} for an LGCP

In subsection 4.3, I will simulate LGCP dependent on binary and continuous random variables. I do this to see how the estimates perform in the case with correlation. For this subsection, I will deduce the theoretical correlation used for these simulations.

Let X be a LGCP driven by $\Lambda = \exp(Y)$ with $\mathbb{E}[Y(u)] = m(u)$ for each location $u \in S \subseteq \mathbb{R}^2$. I will in the dependence simulations for binary variables consider the special case where the binary variable are defined $I(v) = \mathbf{1}[Y(v) < s]$ for $v \in S$. Furthermore, denote the joint density for $(Y(u), I(v))$ by $f_{Y,I}$, the conditional density for $(Y(u) \mid I(v))$ by $f_{Y|I}$, the joint density of $(Y(u), Y(v))$ by $f_{u,v}$, the conditional density for $(Y(u) \mid Y(v))$ by $f_{u|v}$, the marginal density of $Y(v)$ by f_v and $P(I(v) = 1) = p(v)$ such that

$$f_{Y,I}(y_u, 1) = \int_{-\infty}^{s} f_{u,v}(y_u, y_v) dy_v, \quad p(v) = \int_{-\infty}^{s} f_v(y_v) dy_v, \quad f_{Y|I}(y_u \mid 1) = \frac{f_{Y,I}(y_u, 1)}{p(v)}.$$

Note that for the conditional intensity:

$$\begin{aligned}\int_B \lambda_{I(v)}(u \mid v)du = \int_{\mathbb{R}^2} \mathbf{1}[u \in B]\lambda_{I(v)}(u \mid v)du &= \mathbb{E}\left[N(B) \mid I(v)\right] \\ &= \mathbb{E}\left[\int_B \exp\left(Y(u)\right) du \mid I(v)\right] \\ &= \int_B \mathbb{E}\left[\exp(Y(u) \mid I(v)\right] du\end{aligned}$$

Hence, the conditional intensity given $I(v) = 1$ is

$$\begin{aligned}\lambda_1(u \mid v) &= \mathbb{E}\left[\exp(Y(u)) \mid I(v) = 1\right] \\ &= \int_{\mathbb{R}} \exp(y_u) f_{Y|I}(y_u \mid 1) dy_u \\ &= \frac{\int_{\mathbb{R}} \exp(y_u) \int_{-\infty}^{s} f_{u,v}(y_u, y_v) dy_v dy_u}{p(v)} \\ &= \frac{\int_{-\infty}^{s} f_v(y_v) \int_{\mathbb{R}} \exp(y_u) f_{u|v}(y_u \mid y_v) dy_u dy_v}{p(v)} \\ &= \frac{\int_{-\infty}^{s} f_v(y_v) \mathbb{E}\left[\exp(Y_u) \mid Y_v = y_v\right] dy_v}{p(v)} \\ &\overset{(a)}{=} \frac{\int_{-\infty}^{s} f_v(y_v) \exp\left(\mathbb{E}\left[Y_u \mid Y_v = y_v\right] + \frac{\mathrm{Var}(Y_u \mid Y_v = y_v)}{2}\right) dy_v}{p(v)} \\ &\overset{(b)}{=} \frac{\int_{-\infty}^{s} f_v(y_v) \exp\left(m(u) + \frac{\mathrm{Cov}(Y_u, Y_v)}{\mathrm{Var}(Y_v)}\left(y_v - m(v)\right) + \frac{\mathrm{Var}(Y_u) - \frac{\mathrm{Cov}^2(Y_u, Y_v)}{\mathrm{Var}(Y_v)}}{2}\right) dy_v}{p(v)} \\ &= \exp\left(m(u) + \frac{C(u,u)}{2}\right) \frac{\int_{-\infty}^{s} f_v(y_v) \exp\left(\frac{C(u,v)}{C(v,v)}\left(y_v - m(v)\right) - \frac{C^2(u,v)}{2C(v,v)}\right) dy_v}{p(v)},\end{aligned}$$

where (a) derives from the Laplace transform for a normal distribution, and (b) from the conditional expectation and variance for a bivariate normal distribution. Remember from Proposition 2.15 that $\lambda(u) = \exp\left(m(u) + \frac{C(u,u)}{2}\right)$ for a LGCP. Hence the binary summary function from subsection 3.1 is given by

$$\begin{aligned}c_{\text{bin}}(u, v) &= \frac{\lambda_1(u \mid v)}{\lambda(u)} \\ &= \frac{\int_{-\infty}^{s} f_v(y_v) \exp\left(\frac{C(u,v)}{C(v,v)}\left(y_v - m(v)\right) - \frac{C^2(u,v)}{2C(v,v)}\right) dy_v}{p(v)} \\ &= \frac{\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi C(v,v)}} \exp\left(-\frac{1}{2C(v,v)}\left(y_v - \left(C(u,v) + m(v)\right)\right)^2\right) dy_v}{p(v)} \\ &\overset{(a)}{=} \frac{\frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{C(u,v) + m(v) - s}{\sqrt{2C(v,v)}}\right)\right)}{p(v)},\end{aligned}$$

where equality (a) derives from the cumulative distribution function for a normal distribution. If $m(v) = s$, then $p(v) = \frac{1}{2}$ such that $c_{\text{bin}}(u, v) = 1 - \mathrm{erf}\left(\frac{C(u,v)}{\sqrt{2C(v,v)}}\right)$, and if also $C(u, v) = 0$, then $c_{\text{bin}}(u, v) = 1$.

For the continuous case, I will consider the special case $J(v) = Y(v)$, such that the summary function from subsection 3.2 is rewritten into

$$c_{con}(u,v) = \int_{\mathbb{R}} y_v f_v(y_v) c_y(u,v) dy_v = \int_{\mathbb{R}} \frac{y_v f_v(y_v) \lambda_y(u \mid v)}{\lambda(u)} dy_v.$$

By Laplace transform and the conditional expectation and variance for a bivariate normal distribution, the conditional intensity is

$$\begin{aligned}\lambda_y(u \mid v) &= \mathbb{E}\left[\exp\left(Y(u)\right) \mid Y(v) = y_v\right] \\ &= \exp\left(m(u) + \frac{C(u,u)}{C(u,v)}\left(y_v - m(v)\right) + \frac{C(u,u) - \frac{C^2(u,u)}{C(u,v)}}{2}\right),\end{aligned}$$

where (a) derives from equivalent calculations for c_{bin}. Hence,

$$\begin{aligned}c_{con}(u,v) &= \int_{\mathbb{R}} \frac{y_v f_v(y_v) \exp\left(m(u) + \frac{C(u,v)}{C(v,v)}\left(y_v - m(v)\right) + \frac{C(u,u) - \frac{C^2(u,v)}{C(u,v)}}{2}\right)}{\exp\left(m(u) + \frac{C(u,u)}{2}\right)} dy_v \\ &= \int_{\mathbb{R}} y_v f_v(y_v) \exp\left(\frac{C(u,v)}{C(v,v)}\left(y_v - m(v)\right) - \frac{C^2(u,v)}{2C(u,v)}\right) dy_v \\ &\overset{(a)}{=} \int_{\mathbb{R}} y_v \frac{1}{\sqrt{2\pi C(v,v)}} \exp\left(-\frac{1}{2C(v,v)}\left(y_v - \left(C(u,v) + m(v)\right)\right)^2\right) dy_v \\ &\overset{(b)}{=} C(u,v) + m(v),\end{aligned}$$

where (a) derives from equivalent calculations for c_{bin}, and (b) uses the expected value for a normal distribution. Notice that $C(u,v) = 0$ if and only if $c_{\text{con}}(u,v) = m(v)$, and $m(v) = 0$, if and only if $c_{\text{con}}(u,v) = C(u,v)$.

For my simulation study, I will simulate a LGCP with the exponential covariance function given by

$$C(u,v) = C(\|u - v\|) = \sigma^2 \exp\left(-\frac{\|u - v\|}{\alpha}\right),$$

where σ^2 is the variance parameter and α is the scale parameter. This function is only 0 when $\sigma^2 = 0$, such that $C(v,v)$ is also 0, which isn't allowed. Hence, we expect correlation when generating X and the random variables from the same underlying GRF.

4 Simulation

This section displays estimates of c_{bin} and c_{con} from simulated data within a rectangular window $W = [0, w_1] \times [0, w_2]$. The binary and continuous variables are independently simulated from the homogeneous Poisson point process and LGCP in subsections 4.1 and 4.2. In subsection 4.3, the LGCP will be simulated dependently on the binary and continuous variables.

Before showing the simulation studies, I will introduce how I simulated the two point patterns and the realisations of the random variables. The methods used to simulate the point processes are based on [1].

When simulating a LGCP or an inhomogeneous Poisson point process, an independent thinning of a homogeneous Poisson point pattern is used.
Definition 4.1. Let X be a point process on S and $p : S \to [0, 1]$ be a function. If each point $u \in X$ are independently retained in the point process $X_{thin} \subseteq X$ with probability $p_r(u)$, then X_{thin} is called the independent thinning of X with retention probabilities $p_r(u)$ for $u \in S$.

A Poisson point pattern in W with homogeneous intensity λ is simulated by:

1. Generate a realisation of $N(W) \sim \text{po}(\lambda w_1 w_2)$ points.
2. Each point is independently allocated a location in W by the uniform distribution.

If an inhomogeneous Poison point process should be simulated, then let $\lambda^{\max}$ be the maximal intensity of the inhomogeneous Poison point process. Simulate a homogeneous Poison point pattern with intensity $\lambda^{\max}$. Hereafter, make an independent thinning of the current simulated homogeneous Poison point pattern with $p_r(u) = \lambda(u)/\lambda^{\max}$.

When simulating a LGCP, we first need to simulate a GRF Y_W. However, it is impossible to simulate Y_W completely smooth on a computer. Instead, we divide W into a finite number of disjoint regions W^u centred around each location u from a finite set $\mathcal{U}$ such that $W = \cup_{u \in \mathcal{U}} W^u$. Within each region W^u, we simulate Y_{W^u} as a constant. Different methods exist to simulate the realisations of each constant. However, the theory of these methods is omitted from this project. A LGCP in W is simulated by:

1. Simulate a realisation $y_W = \{y(u) : u \in W\}$ of the GRF Y_W.
2. Let $\lambda^{\max} = \max_{u \in W} (\exp(y(u)))$. Generate a realisation of $N(W) \sim \text{po}(\lambda^{\max} w_1 w_2)$ points.
3. Each point is independently allocated a location in W by the uniform distribution.
4. A thinning is performed with retention probabilities $p_r(u) = \exp(y(u))/\lambda^{\max}$ for $u \in W$.

The binary variables $I(\cdot)$, which were independent of the point patterns, were simulated by:

1. Create a $N_{\text{Grid}} \times N_{\text{Grid}}$ rectangular grid of points within $(0, w_1) \times (0, w_2)$.
2. Randomly allocated each location v to either set $\{v : I(v) = 1\}$ or $\{v : I(v) = 0\}$ with constant probability $P(I(v) = 1) = p$.

The independent continuous variables $J(\cdot)$ were simulated by:

1. Create a $N_{\text{Grid}} \times N_{\text{Grid}}$ rectangular grid of points within $(0, w_1) \times (0, w_2)$.

2. Simulate N^2_{Grid} normally distributed variables with both constant mean $\mathbb{E}[J]$ and variance $\text{Var}[J]$.

3. Allocate each normal distributed variable to a location v in the grid.

The simulations with dependence between the random variables and LGCP were based on the special case and theory from subsection 3.3. I first simulate a GRF where the random variables are given as $J(v) = Y_W(v)$ and $I(v) = \mathbf{1}[Y_W(v) < s]$ for some fixed s. Hereafter, the LGCP is simulated with the same GRF used for the random variables. In my studies, the simulations of the point patterns are all performed with the R package `spatstat`. The coding for the kernels and estimates are shown in appendices C.1, C.2 and C.3.

4.1 c_{bin} and c_{con} for Independent Simulated Poisson Point Processes

In this subsection, the Poisson point processes are independently simulated from the binary and continuous variables. Hence, the true $c_{\text{bin}}(u, v) = 1$ and $c_{\text{con}}(u, v) = \mathbb{E}[J(v)]$ for each $u, v \in W$. Appendices C.4 and C.5 show the code used for one simulation of the binary and continuous cases.

In the simulation studies, I will investigate the estimates of c_{bin} and $c_{\text{con}}/\mathbb{E}[J]$ when changing the parameters of the simulated Poisson point process and the random variables. One comparison comes by letting λ be both small and big. From this, it's possible to see what happens with the estimate when the average generated number of points in the point processes is small. I will change the boundaries of W to see how the estimate may change due to r extending out of the window. By not changing N_{Grid}, it's also possible to see if the estimates change when the distance increases between each location for the random variables. For the binary variables, I will also investigate what happens when p gets small or big since the size of $\{v : I(v) = 1\}$ gets small for a small p, while for a big p, the size gets big. It is expected that the smaller sample size leads to a worse estimate and a bigger bias. The sample size is also examined for the continuous variables by changing N_{grid}. Furthermore, I will also compare $\mathbb{E}[J]$ being a positive value and 0, since when $\mathbb{E}[J] = 0$, we cannot look at the ratio $c_{\text{con}}/\mathbb{E}[J]$. Instead, I will consider $c_{\text{con}} - \mathbb{E}[J]$, which has reference point 0. I consider $c_{\text{con}}/\mathbb{E}[J]$ in general, because it centres around 1 with smaller fluctuations compared to $c_{\text{con}} - \mathbb{E}[J]$ for big $\mathbb{E}[J]$. The variance of J is compared between a small and big value since a big value may increase the estimates' variance. For both the binary and continuous variables, I also investigate what happens when some of the locations v for the variables are on the edge of W since this will probably lead to a big difference between the edge-corrected and non-edge-corrected estimates. The choice of bandwidths are also explored from 4 different bandwidths and estimates of multiple simulations. The smoothness of the curves shows if the bandwidth should increase. Since the curve is oversmoothed if we have a big bias, which happens for a big b. While the variance is high, when b is small, the curve is undersmoothed. Tables 1 and 2 summarise the parameters used for each study in the binary and the continuous case. In each study of the binary variables, I let $N_{\text{Grid}} = 10$.

In the first study with binary variables, I let $W = [0, 1] \times [0, 1]$ and $\lambda = 50$, so I simulated an average of 50 points. I let the constant probability be $p = 0.25$ such that I generated on average 25 locations v with $I(v) = 1$. The Poisson point pattern and locations for $I(v) = 1$ are displayed in Figure 1 for one simulation.

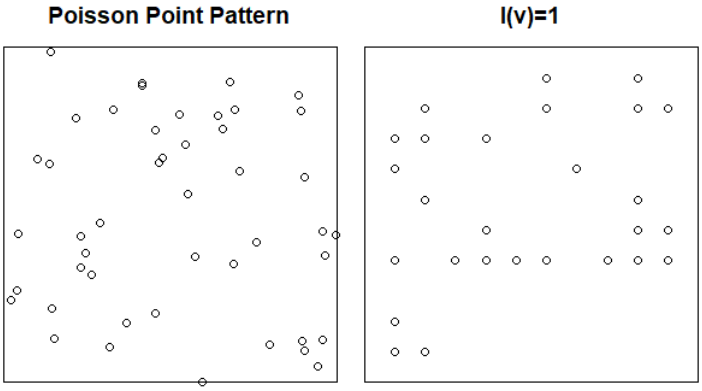

Figure 1: *One simulation of the first study with binary variables. (Left) The Poisson point pattern and (right) locations v where $I(v) = 1$*

After simulating the data, I estimated the p by the sample mean and intensity λ by $\hat{\lambda} = N(x_W)/|W|$. Hereafter, I tried to find a proper bandwidth from one simulation. In the first simulation, I tried using the suggested bandwidths for $\hat{g}$ with a uniform and Epanechnikov kernel given by $0.1/\sqrt{\hat{\lambda}}$ and $0.15/\sqrt{\hat{\lambda}}$. After this, c_{bin} was estimated by both $\hat{c}_{\text{bin},0}$ and $\tilde{c}_{\text{bin},0}$ with both kernels. Figure 2 displays the four estimates for small and large values of r.

The bandwidths seem to be too small since the curves are undersmoothed. Figure 3 displays how this first simulation performs with different bandwidths. The estimates with $b = 0.01$ were very undersmoothed, while $b = 0.2$ gave oversmoothed estimates. However, the bandwidths 0.05 and 0.1 seemed to give some more acceptable smoothing. The curves in both figures also suggest that it doesn't matter whether the non-edge-corrected or edge-corrected estimator is used for a small enough r. However, when r increases, the non-edge-corrected estimator gets smaller. Hence, I used the edge-corrected estimator for 39 simulations, in which I wanted to compare how the bandwidths 0.05 and 0.1 perform in general. The mean and boundaries of these simulations' estimates are shown in Figure 4. Observe from the figure, that $b = 0.05$ gives a wider range of values for the estimate compared to $b = 0.1$. We also see from the mean that in general the estimates are downwards biased for $r < b$. The choice of kernels doesn't matter for $r > b$.

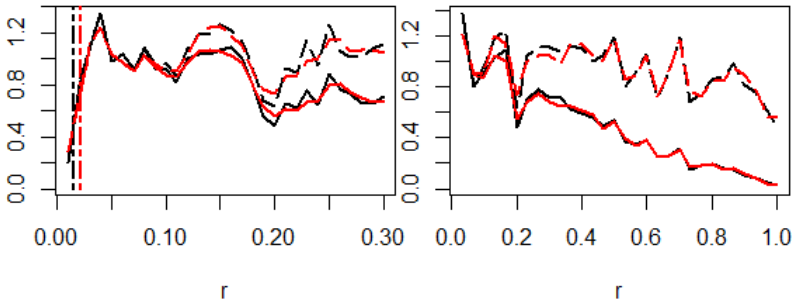

Figure 2: *The first simulation of c_{bin} for study 1. The solid lines are $\hat{c}_{bin,0}$, the dashed lines are $\tilde{c}_{bin,0}$, and the vertical lines are the used bandwidths. The black lines are for the uniform kernel, and the red lines are for the Epanechnikov kernel. (Left) The estimates with small r values and (right) for big r values.*

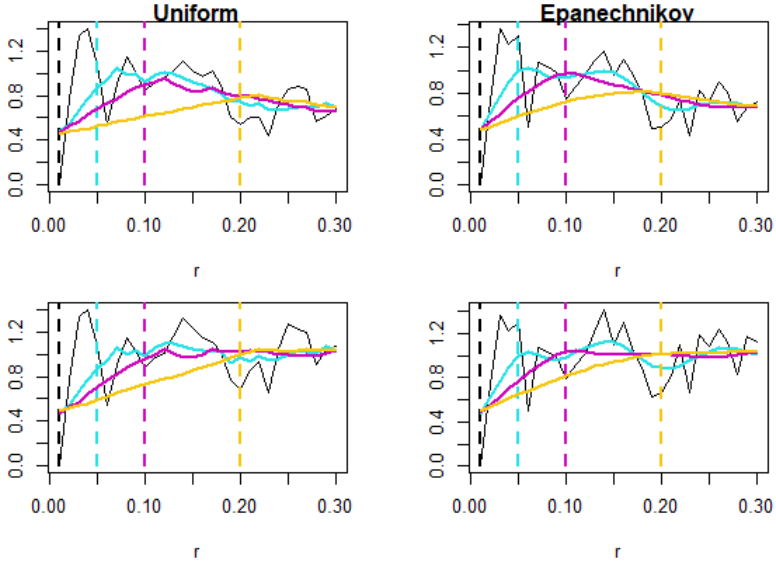

Figure 3: *The first simulation of c_{bin} for study 1. (Upper) The non-edge-corrected estimates and (lower) edge-corrected. (Left) Estimates with uniform kernel and (right) Epanechnikov kernel. The black estimates have bandwidth* 0.01*, the blue have bandwidth* 0.05*, the purple have bandwidth* 0.1*, and the yellow have bandwidth* 0.2*.*

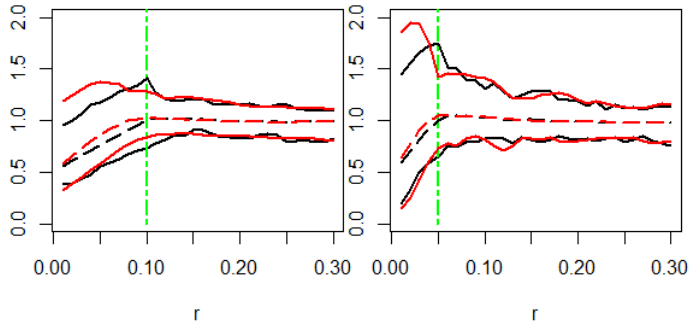

Figure 4: *The 39 simulation for study 1 of $\tilde{c}_{bin,0}$ with bandwidth (left)* 0.1 *and (right)* 0.05*. The solid lines are the minimum and maximum of $\tilde{c}_{bin,0}$, the dashed lines are the mean, and the vertical line are the used bandwidths. The black lines are for the uniform kernel, and the red are for the Epanechnikov kernel.*

Table 1 gives an overview of the parameters used for five other studies in the binary case. To make the estimations, I found a sufficient bandwidth from the smoothness of the estimates curves of a single simulation shown in Figure 5. The multiple simulations of the edge-corrected estimates are shown in Figure 6. Figure 5 shows again that as r increases, the non-edge-corrected estimates deviate from the edge-corrected estimate, especially for the study where some locations $v \in \{v : I(v) = 1\})$ are on the edges. Figure 6 again shows the mean of the estimates being 1 and a downward bias for $r < b$ as expected. However, study 2 has a much bigger maximum value for $r < b$ compared to the other estimates. The odd behaviour may be due to the small sample size, which on average is 10 for a point process with $\lambda = 10$ and $W = [0, 1] \times [0, 1]$. Its first simulation also has poor estimates in Figure 5. The third study had on average 250 points and looks fine for the edge-corrected estimate. From studies 4 and 5, where I simulated the binary variables with both a small and big value for p, it seemed that study 5 performed better since the smoothens of the estimate were the smallest for any of the studies, which resulted in the possibility of using a very small bandwidth. The sample sizes for the simulations of the two studies were also different since study 4 would, on average, have 5 locations v with $I(v) = 1$, while study 5 would have 95. Note that I switched around what was defined as $I(v) = 0$ and $I(v) = 1$ for the simulations. Hence, the estimate of these two studies would suggest that it is sometimes desirable to define $I(v) = 1$ for the event which occurs most frequently. From all of the simulations for the binary case, it also seems that the choice of kernels has only a small significance, since the estimates are nearly the same.

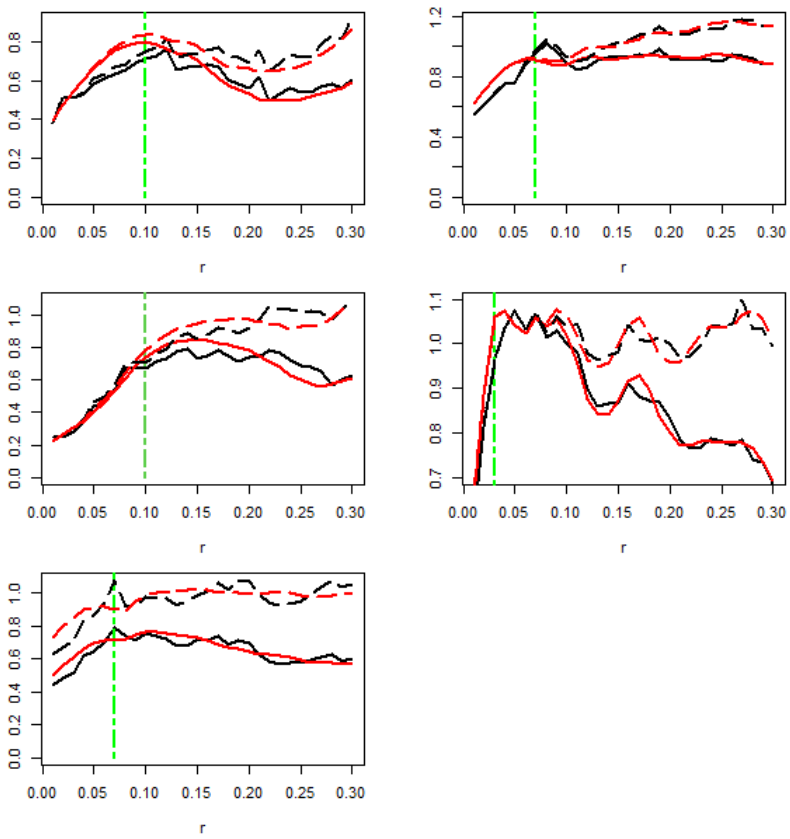

Figure 5: *The first simulation for studies 2-6 of c_{bin}. The solid lines are $\hat{c}_{bin,0}$, the dashed lines are $\tilde{c}_{bin,0}$, and the vertical lines are the used bandwidths. The black lines are for the uniform kernel, and the red lines are the Epanechnikov kernel. (Upper left) Study 2, (Upper right) study 3, (centrum left) study 4, (centrum right) study 5 and (lower) study 6.*

Simulation	λ	W	$p(v)$	$v \in \{v : I(v) = 1\}$ on edges of W
1	50	$[0,1] \times [0,1]$	0.25	No
2	10	$[0,1] \times [0,1]$	0.25	No
3	50	$[0,1] \times [0,5]$	0.25	No
4	50	$[0,1] \times [0,1]$	0.05	No
5	50	$[0,1] \times [0,1]$	0.95	No
6	50	$[0,1] \times [0,1]$	0.25	Yes

Table 1: *Overview of the studies for Poisson point processes in the binary case.*

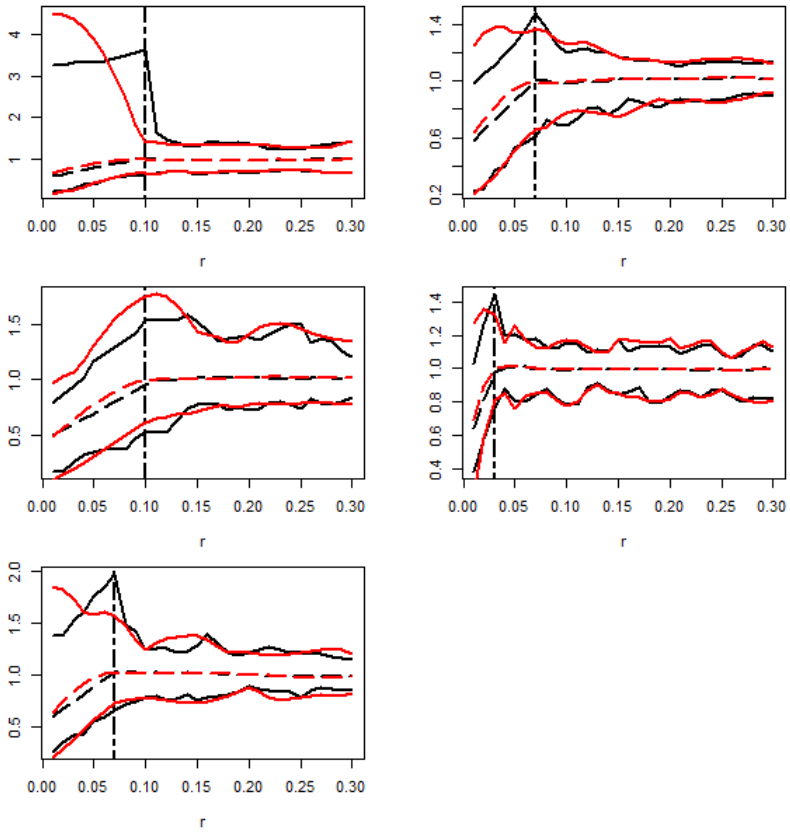

Figure 6: *The 39 simulation for study 2-6 of $\tilde{c}_{bin,0}$. The solid lines are the minimum and maximum of $\tilde{c}_{bin,0}$, the dashed lines are the mean, and the vertical line are the used bandwidths. The black lines are for the uniform kernel, and the red are for the Epanechnikov kernel. (Upper left) Study 2, (upper right) study 3, (centrum left) study 4, (centrum right) study 5 and (lower) study 6.*

For the continuous case, Table 2 gives an overview of the studies, Figure 7 displays the estimates for the first simulation of each study and Figures 9 and 10 show the estimates for 39 simulations. In the first simulation of study 1, the suggested bandwidths for $\hat{g}$ again resulted in an undersmoothed estimate. Figure 8 shows what happens with different choices of bandwidths, where the smoothness of the estimates with bandwidths 0.05 and 0.1 had proper smoothness. Hence, I used them again in the multiple simulations study in Figure 9.

From the multiple simulations study, the estimates with small b again have a wider range compared to the bigger b. The estimates for the first simulation of each study showed again that if the locations of the continuous variables aren't too close to the boundary of the window, the non-edge-corrected and edge-corrected estimates are close to being the same for small r. However, the edge-corrected estimates perform better than the non-edge-corrected for big r.
The multiple simulations in Figure 10 again suggested that the choice of kernels isn't crucial. Also, the mean values show an upwards bias for $r < b$ for the estimate, as expected from the theory. The mean of the multiple simulations all suggest that $\mathbb{E}[J] = c_{\text{con}}$. For the seventh study, I needed to use the estimate of $c_{\text{con}} - \mathbb{E}[J]$ since the true $\mathbb{E}[J]$ was 0. This resulted in a bigger spread compared to the others. However, the estimate still suggests $\mathbb{E}[J] = c_{\text{con}}$. The increase of variance in study 6 makes the estimates fluctuate a bit more compared to the others.

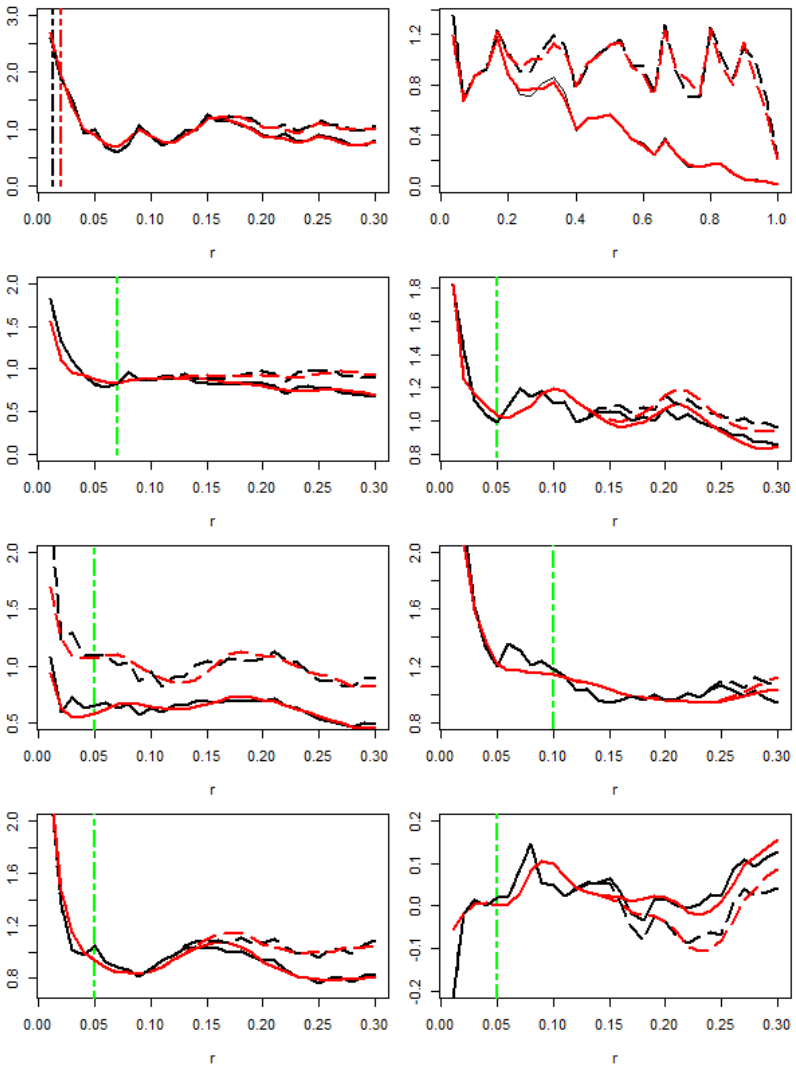

Figure 7: *First simulation of studies 1-7 of c_{con}. (Upper) Study 1 with both small and big values of r. (Second left) study 2, (second right) study 3, (third left) study 4, (third right) study 5, (lower left) study 6, and (lower right) study 7. The solid lines are $\hat{c}_{con,0}/\hat{E}[J]$, and the dashed lines are $\tilde{c}_{con,0}/\hat{E}[J]$, except for the lower right plot, which is $\hat{c}_{con,0} - \hat{E}[J]$ for the solid lines and $\tilde{c}_{con,0} - \hat{E}[J]$ for the dashed lines. The vertical lines are the used bandwidths. The black lines are for the uniform kernel, and the red are for the Epanechnikov kernel.*

Study	λ	W	N^2_{grid}	$\mathbb{E}(J)$	$\text{Var}(J)$	v for $J(v)$ on edges of W
1	50	$[0,1] \times [0,1]$	25	10	1	No
2	10	$[0,1] \times [0,1]$	25	10	1	No
3	50	$[0,1] \times [0,5]$	25	10	1	No
4	50	$[0,1] \times [0,1]$	25	10	1	Yes
5	50	$[0,1] \times [0,1]$	4	10	1	No
6	50	$[0,1] \times [0,1]$	25	10	10	No
7	50	$[0,1] \times [0,1]$	25	0	1	No

Table 2: *Overview of the studies for the Poisson point processes in the continuous case.*

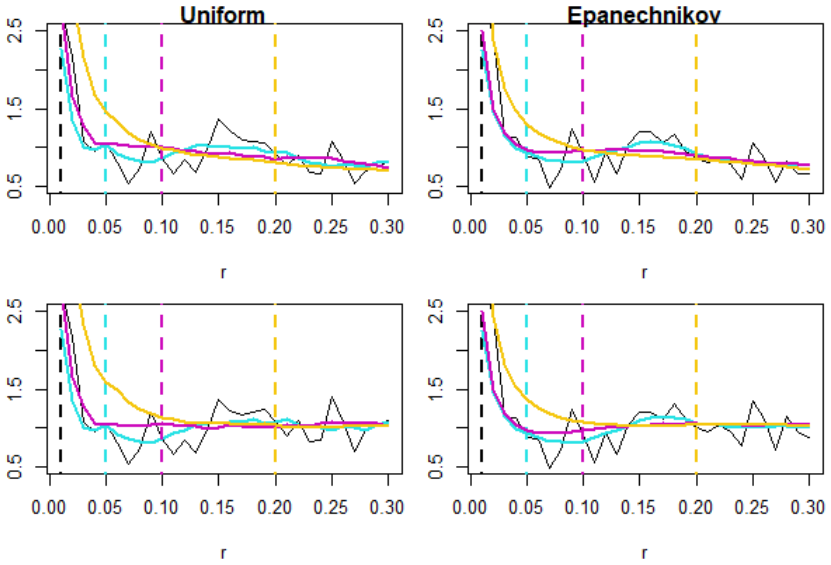

Figure 8: *First simulation for study 1 of $c_{con}/\mathbb{E}[J]$. (Upper) The non-edge-corrected estimates and (lower) edge-corrected. (Left) Estimates with uniform kernel and (right) Epanechnikov kernel. The black estimates have bandwidth* 0.01*, the blue have bandwidth* 0.05*, the purple have bandwidth* 0.1*, and the yellow have bandwidth* 0.2*.*

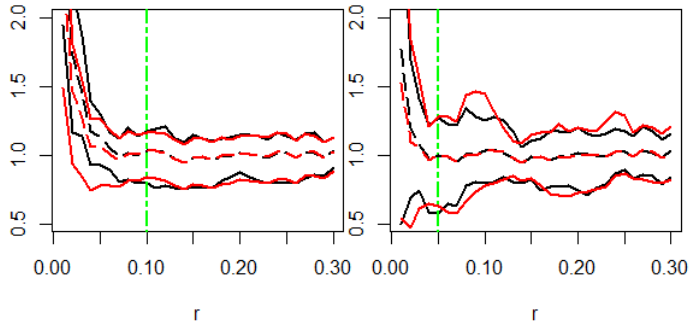

Figure 9: *The 39 simulations for study 1 of c_{con} with bandwidths (left)* 0.1 *and (right)* 0.05*. The solid lines are the minimum and maximum of $\tilde{c}_{bin,0}/\hat{\mathbb{E}}[J]$ and the dashed lines are the mean. The vertical lines are the used bandwidths. The black lines are for the uniform kernel, and the red are for the Epanechnikov kernel.*

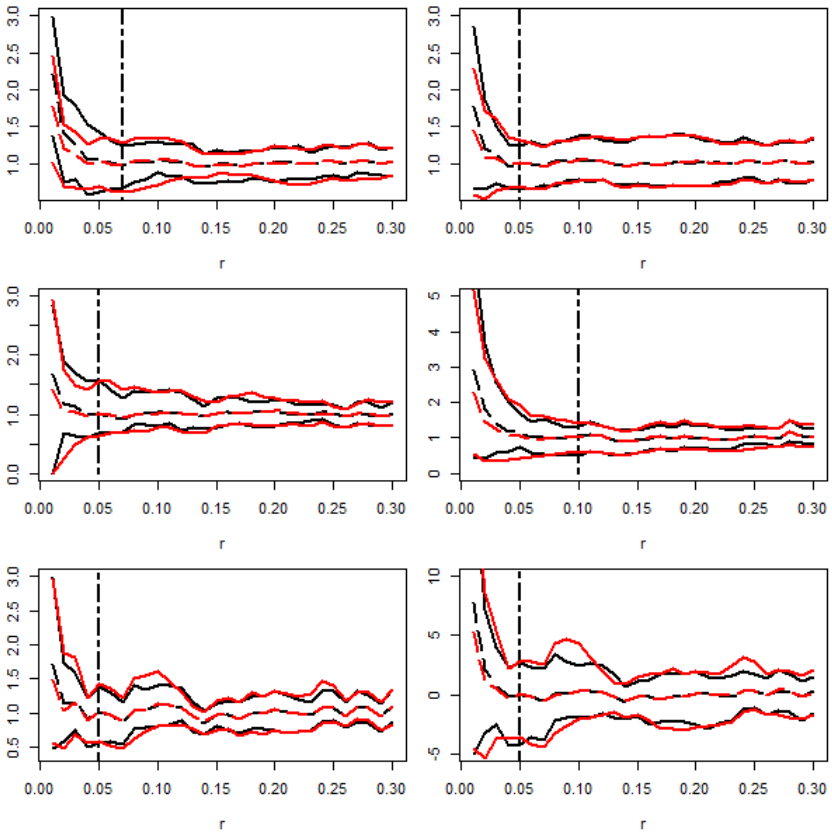

Figure 10: *The 39 simulations for studies 2-7 of c_{con}. (Upper left) Study 2, (upper right) study 3, (centrum left) study 4, (centrum right) study 5, (lower left) study 6 and (lower right) study 7. The solid lines are the minimum and maximum of $\tilde{c}_{bin,0}/\hat{\mathbb{E}}[J]$ and the dashed lines are the mean, expect for study 7, where $\tilde{c}_{bin,0} - \hat{\mathbb{E}}[J]$ is studied. The vertical line are the used bandwidths. The black lines are for the uniform kernel, and the red are for the Epanechnikov kernel.*

4.2 c_{bin} and c_{con} for Independent Simulated LGCPs

In this subsection, I work with LGCPs, which are independently simulated of the random variables. For each simulation, I use the window $W = [0,1] \times [0,1]$ and simulate the binary and continuous random variables with $p = 0.25$, $\mathbb{E}[J] = 10$ and $\text{Var}[J] = 1$. Since the LGCPs are simulated independently of the binary and continuous random variables, the true values of the estimates are $c_{\text{bin}} = 1$ and $c_{\text{con}} = \mathbb{E}[J]$ for each study. I will also only examine the edge-corrected estimate in this subsection since it has now been clarified that the non-edge-corrected estimate works worse.

I intend to investigate the estimates by changing the parameters for the LGCPs. The simulated LGCP shall be driven by a GRF with constant mean $m(u) = m_{\text{c}}$ and exponential covariance function with variance parameter σ^2 and scale α.

I used a big σ^2 and a small m_{c} for the first study, which would show what would happen if my GRF had a great correlation between points nearby and a big variance. The opposite could be observed from the second study, where σ^2 was small and m_{c} big. Note the first two studies don't have a small intensity λ. However, in the third study, I let both σ^2 and m_{c} be small so λ was, on average, also small. Since I work within $W = [0,1] \times [0,1]$, the expected number of points for each pattern is the expected λ, which is $\exp\left(m_{\text{c}} + \sigma^2/2\right)$ by Proposition 2.15. In the last study, I increased the scale so the GRF would have an increased correlation between each point.

The used parameters for each study are shown in Table 3. Figure 11 displays one simulation of the LGCP for each study, and the estimates are displayed in Figure 12 for the multiple simulations. How I simulated a LGCP is shown in appendix C.6. Instead of using a proper bandwidth, I used the same bandwidth, 0.05, for each estimate, so the variance and bias are more easily shown.

All the estimates are centred around 1 such that there is uncorrelation. Also, the binary case has a downward bias for $r < b$, while the continuous case has an upward bias.

The studies with the smallest range for the estimates are 1 and 4. Compared to study 2, the reason is probably due to the increased variance parameter for study 2. For study 3, we have a small sample sizes of the point patterns, so to get a proper estimate, we would need to increase the bandwidth and thereby increase the bias.

Study	σ^2	m_{c}	α	Expected λ
1	1	4	0.1	≈ 90.0
2	5	2	0.1	≈ 148
3	1	1	0.1	≈ 4.48
4	1	4	0.2	≈ 90.0

Table 3: *Overview of the studies for the LGCPs in both the binary and continuous case.*

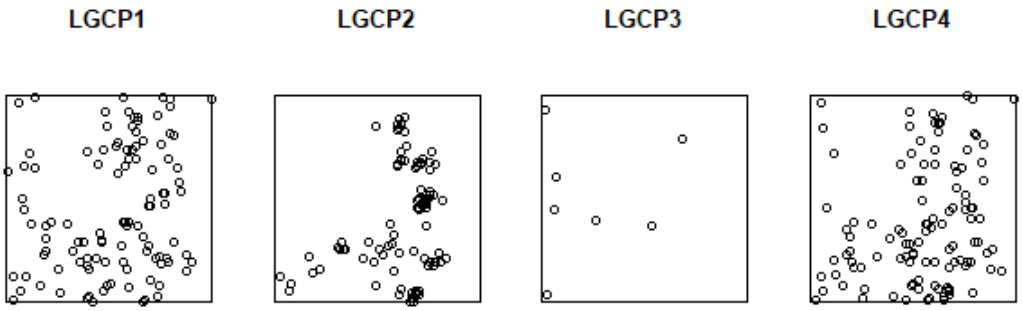

Figure 11: *One LGCP simulated from each study. (Upper left) Study 1, (upper right) study 2, (lower left) study 3 and (lower right) study 4.*

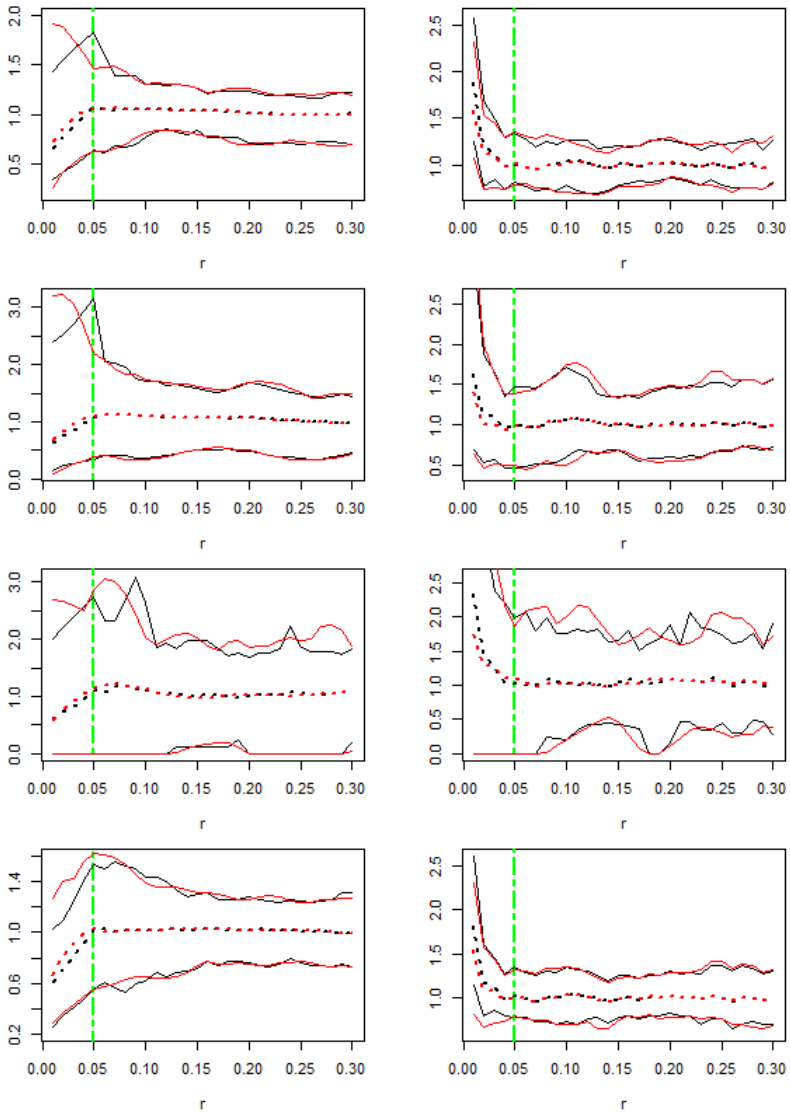

Figure 12: *Estimates for 39 simulations of the studies for independent LGCP. (Left) The edge-corrected estimates of c_{bin} and (right) $c_{con}/\mathbb{E}[J]$ from 39 simulations. (First row) Study 1, (second row) study 2, (third row) study 3 and (fourth row) study 4. The vertical line is the used bandwidth, the dashed lines are the mean of the estimate, the solid lines are the maximum and minium of the estimates. The black lines are for the uniform kernel, and the red lines are the Epanechnikov kernel.*

4.3 c_{bin} and c_{con} for Dependent Simulated LGCPs

In this subsection, I will simulate the LGCPs dependently on an underlying GRF, where my random variables originate. The theory of this is explained in subsection 3.3.

The GRFs will again be simulated with constant mean $m(u) = m_{\mathrm{c}}$ and the exponential covariance function with variance parameter σ^2 and scale α. Each simulation are performed in the window $W = [0,1] \times [0,1]$, such that the expected number of points for each point pattern is given by the expected λ. The GRFs are simulated from a grid where $N_{Grid} = 10$. To transform the continuous variables $J(v)$ to be binary, I used

$$I(v) = \begin{cases} 0 & \text{if } J(v) \geq m_{\mathrm{c}} \\ 1 & \text{if } J(v) < m_{\mathrm{c}} \end{cases}.$$

Remember,

$$c_{\mathrm{bin}}(u,v) = 1 - \operatorname{erf}\left(\frac{C(u,v)}{\sqrt{2C(v,v)}}\right) = 1 - \operatorname{erf}\left(\frac{\sigma \exp\left(-\frac{\|u-v\|}{\alpha}\right)}{\sqrt{2}}\right)$$
$$c_{\mathrm{con}}(u,v) = C(u,v) + m(v) = \sigma^2 \exp\left(-\frac{\|u-v\|}{\alpha}\right) + m_{\mathrm{c}},$$

in this special case by subsection 3.3. In this subsection, I will investigate the same parameters as in subsection 4.2. However, the parameters are expected to change the estimate more in this subsection. Since a big variance parameter for the GRF will give a larger correlation between the random variables and the point process from the true values of c_{bin} and c_{con}. A larger value of α should also increase the correlation compared to a smaller α. I again observe the estimate for a small intensity by letting both m_{c} and σ^2 be small such that the average number of points in the point processes is small, hence giving a small size and maybe a big biased estimate. Table 3 gives an overview of the studies for both the binary and continuous cases. Appendix C.6 shows the code for simulating a LGCP and the random variables from the given GRF.

Figure 13 displays the edge-corrected estimates for 39 simulations of each study. The graphs of the estimates show that estimates can identify correlations between a random variable and a point process.

The binary studies 1 and 3 have a mean nearly identical to the theoretical value of c_{bin}, while studies 2 and 4 are reactively near it. For the continuous studies, we have some lack of smoothness. However, the mean estimates are still centred around the theoretical value. For studies 2 and 3 there is a great range for the estimate of c_{con}, which is probably due to the big variance parameter and small sample size.

Note that the estimates seems neither to be upward or downward bias in general for $r < b$ to the theoretical value. For these studies, the choice of kernels again seems to be insignificant.

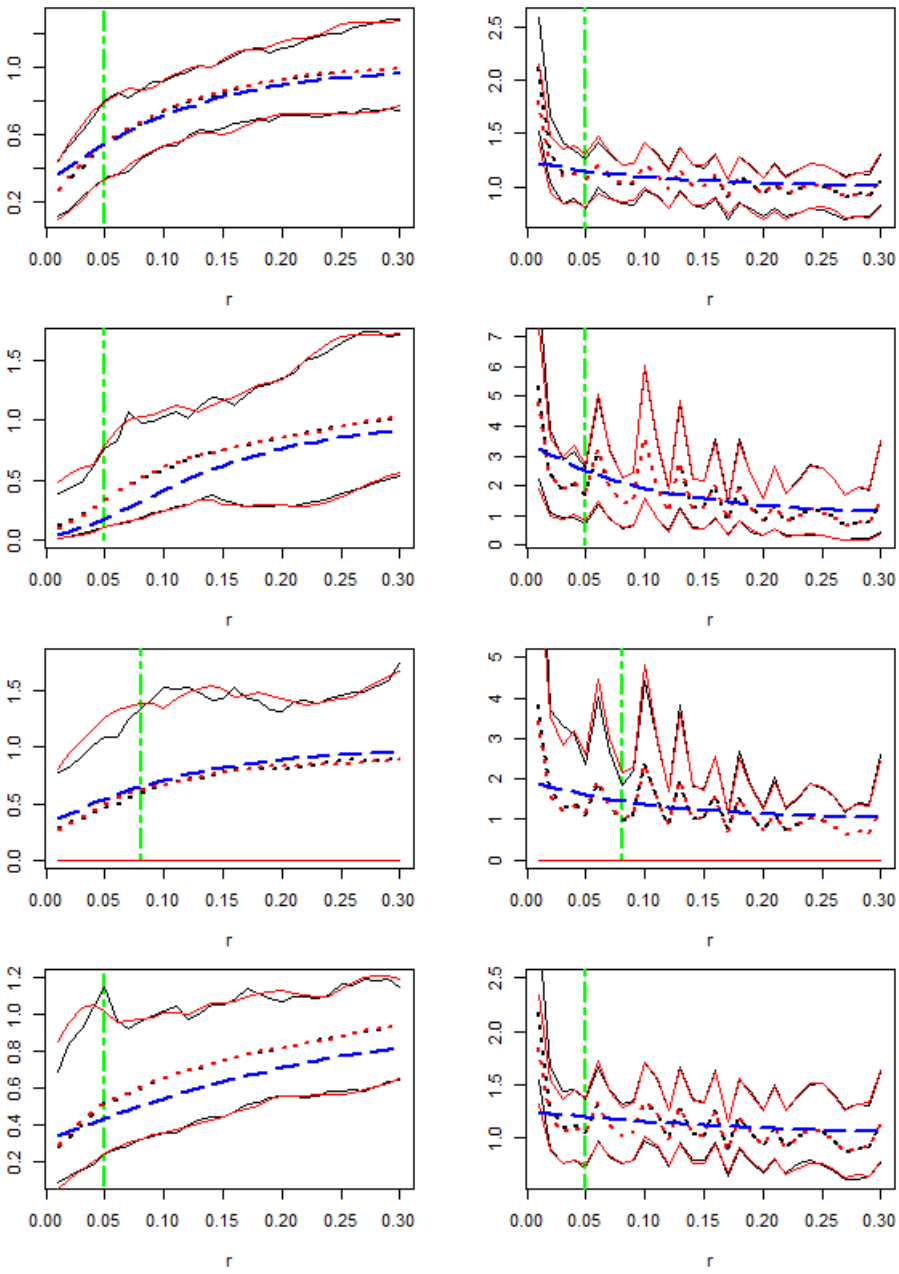

Figure 13: *Estimates for 39 simulations of the studies for dependent LGCP. (Left) The edge-corrected estimates of c_{bin} and (right) $c_{con}/\mathbb{E}[J]$ from 39 simulations. (First row) Study 1, (second row) study 2, (third row) study 3 and (fourth row) study 4. The vertical line is the used bandwidth, the dashed blue lines are the theoretical value, the dotted lines are the mean of the estimate, the solid lines are the maximum and minium of the estimates. The black lines are for the uniform kernel, and the red lines are the Epanechnikov kernel.*

5 Case study

In the case study, I will work with trees of the species Capparis frondosa within Barro Colorado Island in Panama. The trees are observed in a window $W = [0, 1000] \times [0, 500]$, where the distance is in meters. The state of the forest was observed over multiple years. However, I will concentrate on 2005 since 19 continuous variables were collected this year. I assume the intensity of the point process of trees is homogenous giving a estimated intensity $\hat{\lambda} = 0.00682$.

Each variable was observed at 300 sample locations in W, where 200 of these points were located in a regular grid in W and called the base points. An additional location was randomly selected near every other base point. A 1/3 of the additional locations were 2 meters away from the base point, another 1/3 were 8 meters away, and another 1/3 were 20 meters away. Figure 14 displays the tree and sample locations for 2005.

In my study, I work with the continuous variables iron (Fe), boron (B), zinc (Zn) and pH. The estimated mean values of the continuous variables are $\hat{\mathbb{E}}[\text{Fe}] \approx 181$, $\hat{\mathbb{E}}[\text{B}] \approx 0.966$, $\hat{\mathbb{E}}[\text{Zn}] \approx 5.53$ and $\hat{\mathbb{E}}[\text{pH}] \approx 5.67$. From the estimated mean values, it is fair to assume that $\mathbb{E}[J] \neq 0$, such that we do not have a problem assessing $c_{\text{con}}/\hat{\mathbb{E}}[J]$ in the continuous case. I let the transformations to the binary variables be

$$I(v) = \begin{cases} 0 & \text{If Fe}(v) \geq \hat{\mathbb{E}}[\text{Fe}] \text{ at location } v. \\ 1 & \text{If Fe}(v) < \hat{\mathbb{E}}[\text{Fe}] \text{ at location } v. \end{cases},$$

$$I(v) = \begin{cases} 0 & \text{If B}(v) \geq \hat{\mathbb{E}}[\text{B}] \text{ at location } v. \\ 1 & \text{If B}(v) < \hat{\mathbb{E}}[\text{B}] \text{ at location } v. \end{cases},$$

$$I(v) = \begin{cases} 0 & \text{If Zn}(v) \geq \hat{\mathbb{E}}[\text{Zn}] \text{ at location } v. \\ 1 & \text{If Zn}(v) < \hat{\mathbb{E}}[\text{Zn}] \text{ at location } v. \end{cases},$$

$$I(v) = \begin{cases} 0 & \text{If pH}(v) \geq \hat{\mathbb{E}}[\text{pH}] \text{ at location } v. \\ 1 & \text{If pH}(v) < \hat{\mathbb{E}}[\text{pH}] \text{ at location } v. \end{cases}.$$

The number of locations in the set $\{v : I(v) = 1\}$ for each variable is $N_{FE} = 167$, $N_B = 174$, $N_{Zn} = 190$ and $N_{pH} = 135$, so the estimates for p are $\hat{p}_{FE} \approx 0.557$, $\hat{p}_B = 0.58$, $\hat{p}_{Zn} \approx 0.633$ and $\hat{p}_{pH} = 0.45$. All the locations $v \in \{v : I(v) = 1\}$ are shown in Figure 15 and the smoothing of the continuous variables in Figure 16.

To validate the conclusion of the estimates, I permuted the original random variables 39 times. Each permuted random variable was created by randomly relocating each original random variable to another sample location. Note I only used the already existing sample locations. The permuted random variables and the point process of trees are independent since the new variables will not have contributed to the number or locations of trees. If the estimates of the permuted and original random variables are significantly different, then the conclusion is correlation between the point process of trees and the original random variables.

For each estimate, I generated the 30 r values. Denote for all the permuted random variables the estimated minimum and maximum for each r by

- $\tilde{c}_{\text{bin,min}}(r)$ and $\tilde{c}_{\text{bin,max}}(r)$ for the binary case .
- $\tilde{c}_{\text{con,min}}(r)$ and $\tilde{c}_{\text{con,max}}(r)$ for the continuous case .

Denote the original estimates by $\tilde{c}_{\text{bin},O}(r)$ and $\tilde{c}_{\text{con},O}(r)$ for each r. For 39 permutations, it holds that

$$P\left(\tilde{c}_{\text{bin},O}(r) < \tilde{c}_{\text{bin,min}}(r)\right) = P\left(\tilde{c}_{\text{bin},O}(r) > \tilde{c}_{\text{bin,min}}(r)\right) = 0.025.$$
$$P\left(\tilde{c}_{\text{con},O}(r) < \tilde{c}_{\text{con,min}}(r)\right) = P\left(\tilde{c}_{\text{con},O}(r) > \tilde{c}_{\text{con,min}}(r)\right) = 0.025.$$

A proper bandwidth for the estimates seemed to be $b = 1.5$. Figures 17 and 18 display the estimations of c_{bin} and $c_{\text{con}}/\mathbb{E}[J]$ with $r \in [0, 10]$.

The estimates suggest that iron is mostly likely to be uncorrelated with the point process of trees. The binary variables for boron and pH are also mostly uncorrelated, while zinc is negatively correlated. The continuous variables for boron, zinc and pH have a positive correlation for some r values. For both the binary and continuous case, we see some very small and big values for $r < b$, which comes for the downwards bias and upwards bias for each estimate.

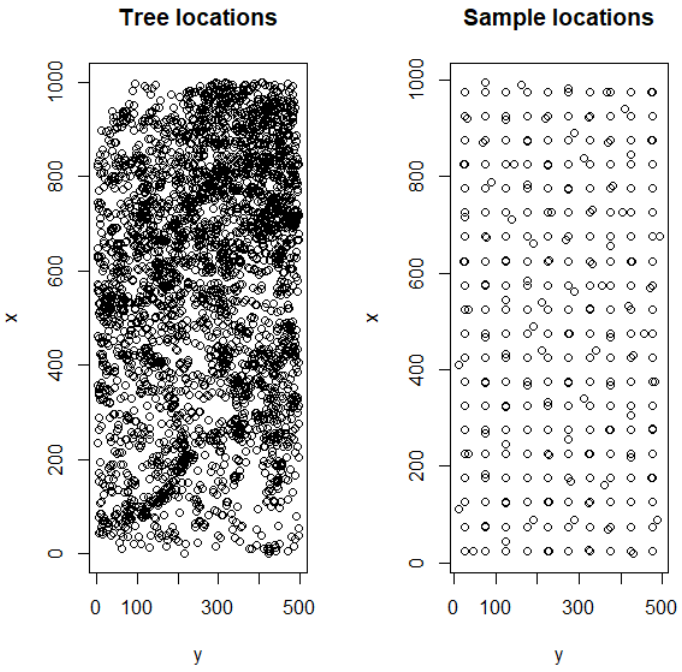

Figure 14: *(Left) Locations of the trees and (right) sample locations for the continuous variables in 2005.*

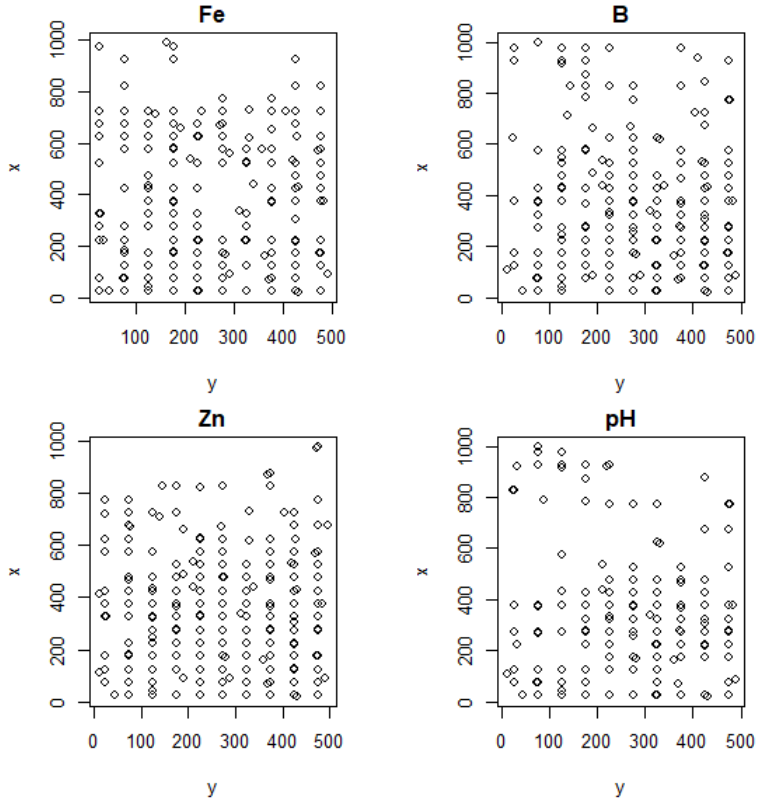

Figure 15: *All the locations $v \in W$ where $I(v) = 1$.*

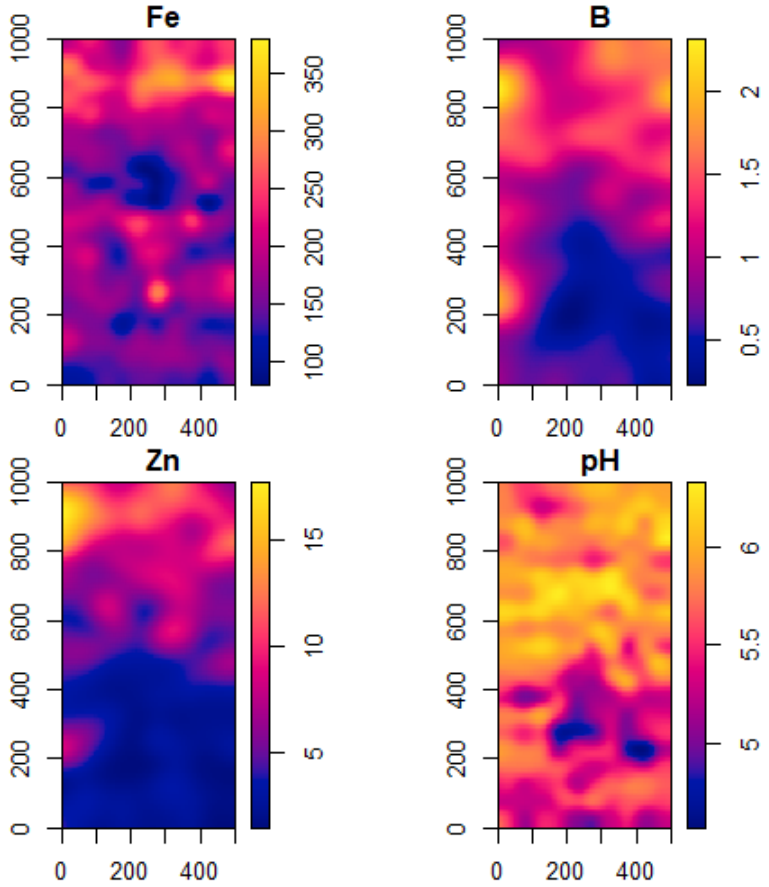

Figure 16: *Smoothing plot of the continuous variables.*

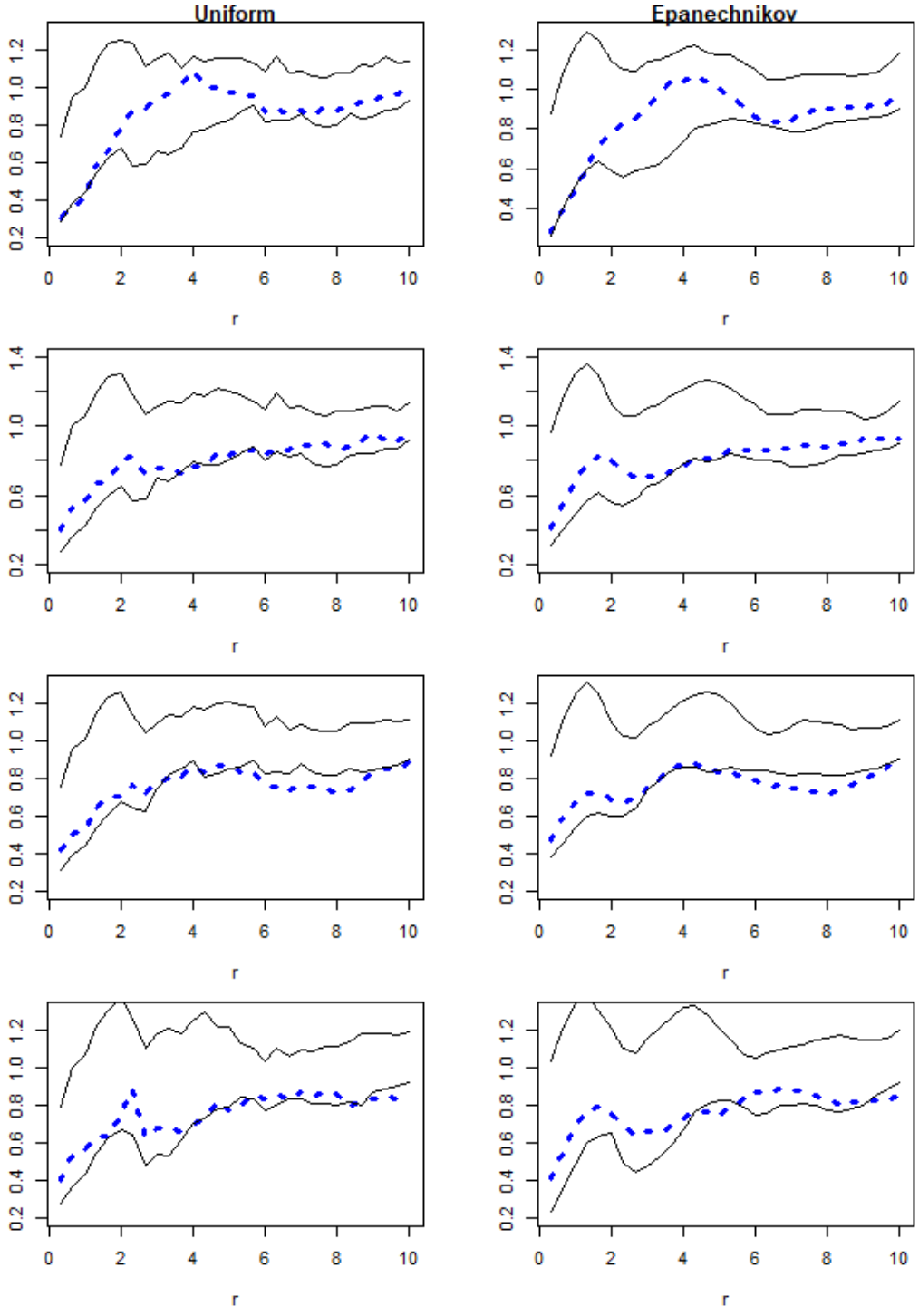

Figure 17: *The edge-corrected estimation of $c_{bin}/\mathbb{E}[J]$. The blue lines are the original estimate, the black lines are the maximum and minimum of the permuted estimates. (Left) Estimated by the uniform kernel and (right) the Epanechnikov. The used variables are (upper) iron, (second) boron, (third) zinc and (lower) pH.*

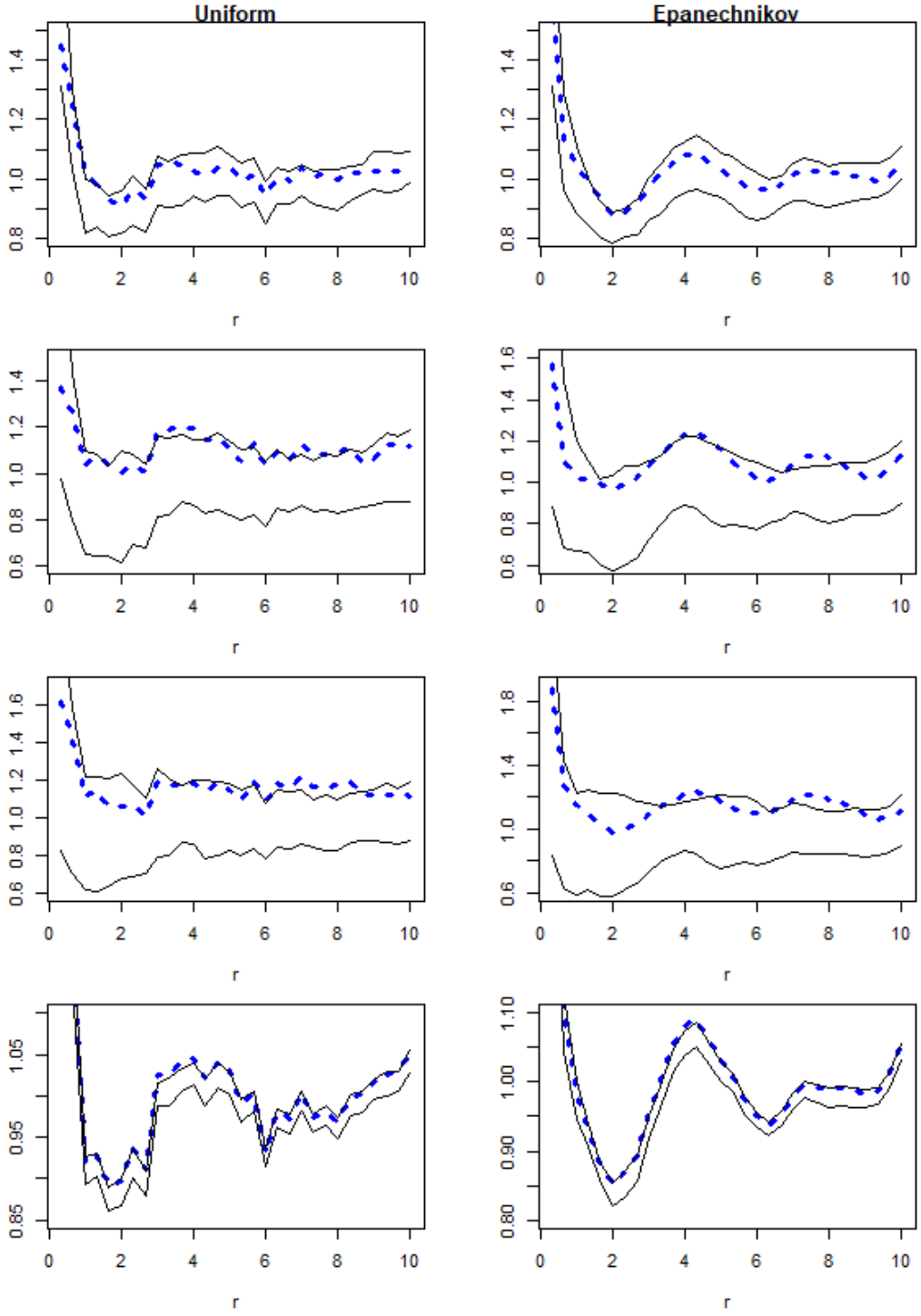

Figure 18: *The edge-corrected estimation of $c_{con}/\mathbb{E}[J]$. The blue line are the original estimate, the black lines are the maximum and minimum of the permuted estimates. (Left) Estimated by the uniform kernel and (right) the Epanechnikov. The used variables are (upper) iron, (second) boron, (third) zinc and (lower) pH.*

6 Discussion and Perspective

In this project, I have only worked with non-parametric kernel estimates, which led me to make some choices. I never found out what the most optimal bandwidths were. I chose the bandwidths, which seemed to give a good smoothness of curves for the estimates. In practice, it could be found by cross-validations. In [3], they chose a bandwidth for the estimate of g by minimising an estimate of the mean integrated squared error with least squares cross-validation.

In the project, I only used the uniform and Epanechnikov kernels. However, other kernels do exist and could have been used. However, the choice of the kernels seems to be insignificant compared to the choice of bandwidth, which is also the case for non-parametric estimates of the pair correlation function and inhomogeneous intensity function.

For the estimates, I could also have used alternative methods. For one, I assumed that the intensities for the point processes were homogenous in my simulations and case studies. However, the intensities could be inhomogeneous. Using the inhomogeneous intensity in estimates would increase the estimation time and add another bandwidth selection to the estimate. I used a factor to remove the edge effects in the estimates. However, alternative edge-correction factors or including buffer zones in the window boundary problem may have solved the boundary problem.

Instead of looking at the non-parametric kernel estimations, I could also have used estimations by regression models. In [5], they used the log-linear regression model

$$\lambda(u) = \exp\left(\beta Z(u)\right),$$

where $Z(u)$ is a covariate at location u with parameter β. To estimate β, we then need to know $Z(u)$ for each $u \in W$. However $Z(u)$ is only known for some locations. Interpolations of the Z's are used to handle the problem, which can lead to a bias. The estimations in this project don't have a problem with this interpolation and can be used as an alternative.

7 Conclusion

In conclusion of this project, I gained summary functions for the correlation between a point process and both binary and continuous variables. Furthermore, edge-corrected non-parametric estimates of the summary functions performed best. From the simulation study, the edge-corrected estimates seemed to perform well unless the sample size of the random variables or the number of points in the point pattern was small. From my application, I discovered that iron was likely uncorrelated with the tree locations in BCI, while boron, zinc and pH were positive correlated.

Appendices

A Standard Proof

This result is from [1] and is used to prove results throughout the project.
Lemma A.1. *Let h be a non-negative function defined on a measure space $(H, \mathcal{H})$. Suppose we want to prove an equality given by*

$$\mathbb{E}[h(X)] = H(h),$$

where H is some functional, such as the integral. Let A be a subset in $\mathcal{H}$. Then, it is enough to show, that for each A the expression is satisfied for $h(x) = \mathbf{1}[x \in A]$ to prove it.

B Kernels

This section will introduce kernels and two examples used in my application.

A kernel is a probability density function defined by $k : \mathbb{R} \to [0, \mathbb{R})$ or $k : \mathbb{R}^2 \to [0, \mathbb{R})$. Kernels have a bounded support, and I will work within $[-1, 1]$ and $[-1, 1] \times [-1, 1]$.

Uniform:

$$k(u) = \frac{1}{2}\mathbf{1}\left[|u| \leq 1\right] \qquad \text{for } u \in \mathbb{R}$$
$$k(u) = \frac{1}{4}\mathbf{1}\left[|u_1| \leq 1\right]\mathbf{1}\left[|u_2| \leq 1\right] \qquad \text{for } u = (u_1, u_2) \in \mathbb{R}^2.$$

Epanechnikov:

$$k(u) = \frac{3}{4}(1 - u^2)\mathbf{1}\left[|u| \leq 1\right] \qquad \text{for } u \in \mathbb{R}$$
$$k(u) = \frac{9}{16}(1 - u_1^2)(1 - u_2^2)\mathbf{1}\left[|u_1| \leq 1\right]\mathbf{1}\left[|u_2| \leq 1\right] \qquad \text{for } u = (u_1, u_2) \in \mathbb{R}^2$$

C Code

C.1 Kernels

```
ker_b_uni <- function(call,bandwidth){ #Uniform Kernel
  if (abs(call)/bandwidth <= 1 ){
    kb <- 1/2*1/bandwidth
  }
  else {
    kb <- 0
  }
  kb
}

ker_b_epi <- function(call,bandwidth){ #Epanechnikov Kernel
  if (abs(call)/bandwidth <= 1 ){
    k <- 3/4*(1-(call/bandwidth)^2)
    kb <- k/bandwidth
  }
  else {
    kb <- 0
  }
  kb
}
```

C.2 Estimation of $c_{bin,0}$.

In the code, the first function being defined is the non-edge-corrected estimate of $c_{\text{bin},0}$ and the second function is the edge-corrected estimate. Both estimates are with the uniform kernel. To get the estimates with the Epanechnikov kernel, change ker_b_uni to ker_b_epi.

```
library("spatstat")
hatc0homouni <- function(pat,binvar,r,bandwidth,p){ #The non-edge-corrected
    estimate
  X <- cbind(pat$x,pat$y)
  I <- cbind(binvar$x,binvar$y)
  N <- nrow(X)
  NI1 <- nrow(I)
  NItot <- NI1/p
  wind <- pat$window
  cpair <- crosspairs(pat,binvar,rmax = bandwidth+r,what = "indices") #Removes
      the worst locations pairs, that KB do not accept.
  lambda <- N/area(wind) #Estimated intensity
  s2 <- rep(NA,NI1)
  s1 <- matrix(NA,nrow = NI1,ncol = N)
  for(v in 1:NI1){
    if (v %in%  cpair$j){
      for (u in cpair$i[cpair$j==v ]){
        if (X[u,1] != I[v,1] | X[u,2] != I[v,2]){
          kB <- ker_b_uni(r-norm(t(X[u,]-I[v,]),type = "f"),bandwidth) #The
              kernel in the sum. Change to ker_b_epi for Epanechnikov.
          d <- 2*pi * norm(t(X[u,]-I[v,]),type = "f")*p*lambda #The denominator
               in the sum
          s1[v,u] <- kB/d #The inner sum
        }
      }
    }
    s2[v] <- sum(s1[v,],na.rm = TRUE) #The outer sum
  }
```

```
  sum(s2, na.rm = TRUE)/NItot #The estimate
}

tildec0homouni <- function(pat,binvar,r,bandwidth,p){ #The edge-corrected
    estimate
  X <- cbind(pat$x,pat$y)
  I <- cbind(binvar$x,binvar$y)
  N <- nrow(X)
  NI1 <- nrow(I)
  NItot <- NI1/p
  wind <- pat$window
  cpair <- crosspairs(pat,binvar,rmax = bandwidth+r,what = "indices") #Removes
      the worst locations pairs, that KB do not accept.
  lambda <- N/area(wind) #Estimated intensity.
  s2 <- rep(NA,NI1)
  s1 <- matrix(NA,nrow = NI1,ncol = N)
  for(v in 1:NI1){
    if (v %in%  cpair$j){
      for (u in cpair$i[cpair$j==v ]){
        if (X[u,1] != I[v,1] | X[u,2] != I[v,2]){
          rtilde <- norm(t(X[u,]-I[v,]),type = "f")
          kB <- ker_b_uni(r-rtilde,bandwidth) #The kernel in the sum. Change to
              ker_b_epi for Epanechnikov.
          d <- rtilde*p*lambda #The denominator in the sum.
          edgecor <- 1/(2*pi)*edge.Ripley(X = binvar[v],r=rtilde,W=wind,
              maxweight = Inf) #The edge-correction factor in the sum.
          s1[v,u] <- kB/d*edgecor #The inner sum.
        }
      }
    }
    s2[v] <- sum(s1[v,],na.rm = TRUE) #The outer sum.
  }
  sum(s2, na.rm = TRUE)/NItot #The estimate.
}
```

C.3 Estimation of $c_{con,0}$

In the code, the first function is the non-edge-corrected estimate of $c_{\mathrm{con},0}$, and the second is the edge-corrected estimate. Both estimates are with the uniform kernel. To get the estimates with the Epanechnikov kernel, change ker_b_uni to ker_b_epi.

```
library("spatstat")
hatc0conhomouni <- function(pat,contvar,r,bandwidth){ #The non-edge-corrected
    estimate
  X <- cbind(pat$x,pat$y)
  Cv <- cbind(contvar$x,contvar$y)
  Iv <- contvar$marks
  N <- nrow(X)
  NC <- nrow(Cv)
  wind <- pat$window
  cpair <- crosspairs(pat,contvar,rmax = bandwidth+r,what = "indices") #Removes
      the worst locations pairs, that KB do not accept.
  lambda <- N/area(wind) #Estimated intensity
  s2 <- rep(NA,NC)
  s1 <- matrix(NA,nrow = NC,ncol = N)
  for(v in 1:NC){
    if (v %in%  cpair$j){
      for (u in cpair$i[cpair$j==v ]){
        kB <- ker_b_uni(r-norm(t(X[u,]-Cv[v,]),type = "f"),bandwidth) #The
            kernel in the sum. Change to ker_b_epi for Epanechnikov.
```

```
        d <- 2*pi*r*lambda #The denominator in the sum
        s1[v,u] <- kB/d #The inner sum
      }
    }
    s2[v] <- Iv[v]*sum(s1[v,],na.rm = TRUE) #The outer sum
  }
  sum(s2, na.rm = TRUE)/NC #The estimate
}

tildec0conhomouni <- function(pat,contvar,r,bandwidth){ #The edge-corrected
    estimate
  X <- cbind(pat$x,pat$y)
  Cv <- cbind(contvar$x,contvar$y)
  Iv <- contvar$marks
  N <- nrow(X)
  NC <- nrow(Cv)
  wind <- pat$window
  cpair <- crosspairs(pat,contvar,rmax = bandwidth+r,what = "indices") #Removes
      the worst locations pairs, that KB do not accept.
  lambda <- N/area(wind) #Estimated intensity
  s2 <- rep(NA,NC)
  s1 <- matrix(NA,nrow = NC,ncol = N)
  for(v in 1:NC){
    if (v %in%  cpair$j){
      for (u in cpair$i[cpair$j==v ]){
        rtilde <- norm(t(X[u,]-Cv[v,]),type = "f")
        kB <- ker_b_uni(r-rtilde,bandwidth) #The kernel in the sum. Change to
            ker_b_epi for Epanechnikov.
        d <- r*lambda #The denominator in the sum.
        edgecor <- 1/(2*pi)*edge.Ripley(X=contvar[v],r=rtilde,W=wind,maxweight
            = Inf) #The edge-correction factor in the sum.
        s1[v,u] <- kB/d*edgecor #The inner sum.
      }
    }
    s2[v] <- Iv[v]*sum(s1[v,],na.rm = TRUE) #The outer sum.
  }
  sum(s2, na.rm = TRUE)/NC #The estimate.
}
```

C.4 Simulation of Poisson Point Process and Binary Variables

The code first shows one simulation from the Poisson Point Process. Hereafter, the binary variables are simulated. The code ends by showing how to insert the simulation in the estimate with an uniform kernel. The parameters used were $W = [0,1] \times [0,1]$, $\lambda = 50$, N_{10} and $p = 0.25$.

```
bx <- c(0,1) #Boundary for first coordinate in window.
by <- c(0,1) #Boundary for second coordinate in window.
lam <- 50 # Homogeneous intensity.
set.seed(123)
PP <- rpoispp(lambda = lam, win=owin(bx,by)) # Simulated Poisson point process.
grid_bin <- expand.grid((1:10)/11,(1:10)/11) # Grid of locations for the binary
    variables.
p_bin <- 0.25 # P(I(v)=1) for locations v.
set.seed(122)
rbin <- grid_bin[as.logical(rbinom(n = 10*10,size = 1,p=p_bin)), ] #Locations v
    allocated to the set with I(v)=1.
phat <- nrow(rbin)/nrow(grid_bin) #The sampled p value.
bin <- ppp(x=rbin[,1], y=rbin[,2],window = owin(bx,by))
```

```
b_uni_PP <- 0.1/sqrt(intensity(PP)) # Suggested bandwidth for the uniform
    estimate of g.

PP_t_cb_uni <-c() #Tilde estimates with uniform kernel.
PP_h_cb_uni <-c() #Hat estimates with uniform kernel.

for (i in 1:30){
  PP_t_cb_uni[i] <- tildec0homouni(pat = PP,binvar = bin,r = i/100, bandwidth =
       b_uni_PP,p = phat )
  PP_h_cb_uni[i] <- hatc0homouni(pat = PP,binvar = bin,r = i/100, bandwidth =
      b_uni_PP,p = phat )

}
```

C.5 Simulation of Poisson Point Process and Continuous Variables

The code first shows one simulation from the Poisson Point Process. Hereafter, the continuous variables are simulated. The code ends by showing how to insert the simulation in the estimate with an uniform kernel. The parameters used were $W = [0,1] \times [0,1]$, $\lambda = 50$, N_5, $\mathbb{E}[J] = 10$ and $\text{Var}[J] = 1$.

```
bx <- c(0,1) #Boundary for first coordinate in window.
by <- c(0,1) #Boundary for second coordinate in window.
lam <- 50 # Homogeneous intensity.
set.seed(122)
PP <- rpoispp(lambda = lam, win=owin(bx,by)) # Simulated Poisson point process.

grid_con <- expand.grid((1:5)/6,(1:5)/6) # Grid of locations for continuous
    variables.
N2_con <- 25 #Number of locations for continuous variables.

set.seed(12)
mark_con<- rnorm(mean = 10,sd =1,n = N_con) #Continuous variables.
mean_con <- mean(mark_con) #Sample mean of continuous variables used for
    relation.

con <- ppp(x=grid_con[,1], y=grid_con[,2],window = owin(bx,by),marks = mark_con
    ) #Continuous variables given a location.

b_uni_PP <- 0.1/sqrt(intensity(PP)) # Bandwidth suggested for uniform.

PP_t_cc_uni <- c() #Tilde estimates with uniform kernel.
PP_h_cc_uni <- c() #Hat estimates with uniform kernel.

for (i in 1:30){
  PP_t_cc_uni[i] <- tildec0conhomouni(pat = PP,contvar = con,r = i/100,
      bandwidth = b_uni_PP)
  PP_h_cc_uni[i] <- hatc0conhomouni(pat = PP,contvar = con,r = i/100, bandwidth
       = b_uni_PP)
}
```

C.6 Simulation of LGCP and dependent variables

The code shows one simulation from a LGCP with parameters $m_c = 4$, $\sigma^2 = 1$, $h = 0.1$ and $W = [0,1] \times [0,1]$. Hereafter, the code shows how to extract the simulated GRF from the LGCP. Finally, I create the sets of binary and continuous variables.

```
mc <- 4 # The used mean for the GRF.
var <- 1 # The used variance parameter for the GRF.
```

```
scale <- 0.1 # The used scale for GRF.
bx <- c(0,1) #The boundaries for the first coordinate of the window.
by <- c(0,1) #The boundaries for the second coordinate of the window.
set.seed(123)
Depcox <- rLGCP(mu = mc,scale=scale,var=var,model = "exp",win=owin(bx,by),eps =
     0.1) #Generates the LGCP.

marksC <- c()
for (i in 1:(10*10)){
  marksC[i] <-  log(attr(Depcox, "Lambda")$v[i])# Pulls out the estimated
      values for the GRF.
}
y <- c()
for (j in 1:10){
  y[j] <- 1/11*j
}
y <- rep(y,10)
x <- y[order(y)]

sample_mean <- mean(marksC) # Sample mean of the GRF.
marksI <- as.integer(marksC < sample_mean) # Sample of the binary variables.
phat <- sum(marksI)/length(marksI) # Estimate of P(I(v)=1)=p(v).

CV <- ppp(x = x, y = y, marks = marksC) #Converting the continuous variables to
     a point process.
IV <- ppp(x = x[marksI==1], y = y[marksI==1]) #Converting the binary variables
    to a point process.
```

References

[1] Jesper. Møller and Rasmus Plenge. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Chapman & Hall/CRC, Boca Raton, Fla, 2003. ISBN 1-58488-265-4.

[2] Daniel A. Griffith. The boundary value problem in spatial statistical analysis. *Journal of regional science*, 23(3):377–387, 1983. ISSN 0022-4146.

[3] Abdollah Jalilian and Rasmus Waagepetersen and. Fast bandwidth selection for estimation of the pair correlation function. *Journal of Statistical Computation and Simulation*, 88(10):2001–2011, 2018. doi: 10.1080/00949655.2018.1428606. URL https://doi.org/10.1080/00949655.2018.1428606.

[4] Daniel Hug and Evgeny Spodarev. Random polytopes. In *Stochastic Geometry, Spatial Statistics and Random Fields*, volume 2068 of *Lecture Notes in Mathematics*, pages 205–238. Springer Berlin / Heidelberg, Germany, 2013. ISBN 9783642333040.

[5] Abdollah Jalilian, Francisco Cuevas-Pacheco, Ganggang Xu, and Rasmus Waagepetersen. Composite likelihood inference for space-time point processes, 2024. URL https://arxiv.org/abs/2402.12548.