RUMLIGE PUNKTPROCESSER

- Modellering af hussalg ved anvendelse af log Gaussiske Cox processer -

AALBORG UNIVERSITET

KANDIDATSPECIALE

Camilla Louise MADSEN
Joachim REIMER

AALBORG UNIVERSITET
STUDENTERRAPPORT

School of Engineering and Science
Institut for matematiske fag
Skjernvej 4a
9220 Aalborg Øst
http://www.math.aau.dk/

Titel:

Rumlige punktprocesser: Modellering af hussalg ved anvendelse af log Gaussiske Cox processer

Tema:

Statistik

Projekt:

Kandidatspeciale

Projektperiode:

Efterår 2019

Projektgruppe:

5.230a

Gruppemedlemmer:

Camilla Louise Madsen
Joachim Reimer

Vejleder:

Jakob Gulddahl Rasmussen

Sidetal: 115

Afsluttet den 09/01-2020

Synopsis:

I dette speciale anvendes teori indenfor rumlige punktprocesser herunder Poisson- og Cox processer med henblik på, at tilpasse log Gaussiske Cox processer til et datasæt. Dette datasæt består af lokationer for hussalg i Aalborg Kommune, som repræsenteres gennem et todimensionelt punktmønster. Ved hjælp af `spatstat`-pakken tilpasses forskellige log Gaussiske Cox processer til punktmønsteret via composite likelihood estimation. I alt tilpasses ni modeller bestående af log Gaussiske Cox processer med tre forskellige intensitetsfunktioner, hver med tre forskellige korrelationsfunktioner for det tilhørende Gaussisk stokastiske felt. Ved anvendelse af punktvise envelopes undersøges det, hvor godt disse modeller tilpasser data.

Forord

Dette speciale er udarbejdet i efteråret 2019 af to studerende på kandidaten på matematikuddannelsen ved Aalborg Universitet. Læserne forventes at have kendskab til basal mål- og integrationsteori samt statistisk teori og herunder stokastiske felter.
Til litteraturreferencer bruges Vancouver-modellen. I starten af hvert kapitel klargøres det, hvilken kilde der er brugt til at skrive kapitlet. Hvis der er brugt andre kilder, vil det refereres umiddelbart efter den givne paragraf.
Definitioner, sætninger, propositioner, lemmaer og korollarer er nummereret efter hvert delafsnit. Ligninger og figurer er nummereret efter kapitel. Når der refereres til disse, gøres det ved at skrive "ifølge Sætning x"eller "ifølge ligning (x)". Et bevis afsluttes med ■ i nedre højre hjørne.
Alle figurer er udarbejdet via platformen RStudio af gruppens medlemmer, og der er derfor ikke kildehenvisninger til disse. I specialet anvendes R som programmeringssprog, og den tilhørende kode er versionsstyret via GitHub.
Specialet indeholder fem kapitler eksklusiv Indledning og Konklussion. I Kapitel 2 introduceres rumlige punktprocesser og Poisson processen. Kapitel 3 beskæftiger sig med summary statistics for punktprocesser og ikke-parametriske estimater af disse. I Kapitel 4 defineres Cox processer, og gennem disse introduceres log Gaussiske Cox processer. Som redskab til modelfitting af punktprocesser introduceres composite likelihood estimation i Kapitel 5. Kapitel 6 beskæftiger sig med at modellere hussalg i Aalborg Kommune ud fra log Gaussiske Cox processer. I Appendiks A-C findes teori som til dels ligger udenfor specialets teoretiske ramme, som kan give en dybere forståelse af nogle af begreberne introduceret i Kapitel 2-5.
Vi vil gerne takke vejleder Jakob Gulddahl Rasmussen for god og konstruktiv vejledning samt at stille data til rådighed. Desuden vil vi gerne takke Ege Rubak for rådgivning omkring anvendelse af `spatstat`-pakken.

Aalborg Universitet, 9. januar 2020

Camilla Louise Madsen
clma13@student.aau.dk

Joachim Reimer
jreim14@student.aau.dk

Abstract

In this thesis, spatial point processes are used to model locations of house sales in Aalborg Kommune. More specifically, the thesis examines how to use composite likelihood estimation in order to fit different log Gaussian Cox processes to a point pattern representing these locations.

In Chapter 2, the theory regarding spatial point processes, and Poisson processes are treated in order to later introduce the more central model class: The log Gaussian Cox process. In Chapter 3, summary statistics are described in order to provide tools for performing an exploratory analysis of spatial point patterns. Herein non parametric estimates for these summary stastistics are given. Pointwise envelopes are also described in this chapter. These are used for hypothesis testing for whether a point process model fits a given point pattern.

In Chapter 4, the log Gaussian Cox process is defined as a Poisson process with intensity function given as the exponential function taken on a Gaussian random field. In this chapter we prove that the intensity, and paircorrelation of the log Gaussian Cox process is expressed exclusively in terms of the mean, and covariance of the Gaussian random field. Furthermore, it is proven that the log Gaussian Cox process is homogeneous if, and only if the underlying Gaussian random field is stationary.
In order to fit different log Gaussian Cox processes to the described spatial point pattern composite likelihood estimation is applied. Composite likelihood methods and estimates are introduced in Chapter 5, where it is shown that the composite likelihood estimator is consistent in the case of homogeneous log Gaussian Cox processes.

In Chapter 6, the composite likelihood method is used to fit nine different types of log Gaussian Cox processes to the spatial point pattern which represents house sales in Aalborg Kommune. These models consist of three different intensity functions, and three different correlation function types for the underlying Gaussian random field. The intensity functions follow a constant model, a linear model, and a third order polynomial model. The correlation types are exponential, Gaussian, and stable. Using pointwise envelopes, we get an indication of which models seem to be performing the best with respect to fitting the dataset.

In conclusion, the models with exponential correlation seem to be performing well for all three intensity functions. The log Gaussian Cox processes with stable correlation seem to be performing well for most interpoint distances. The models with Gaussian correlation does not seem to fit the point pattern well. The inhomogeneous log Gaussian Cox process with third order intensity however, seem to be a slightly better fit than the two other models with Gaussian correlation.

Indholdsfortegnelse

Indledning 1

For at forstå konceptet bag en rumlig punktproces, skal begrebet omkring et rumligt punktmønster konkretiseres. Et rumligt punktmønster er et datasæt bestående af observerede lokationer for rumlige hændelser. En rumlig punktproces kan da ses som en tilfældig størrelse, hvis udfald er et punktmønster. Punktprocesser kan anvendes til at modellere adskillige fænomener ud fra fænomenernes rumlige position. Disse fænomener kan for eksempel være jordskælv, træer eller stjerner i galakser, hvis rumlige placering kan observeres, og dermed repræsenteres via et punktmønster. Disse punktmønstre vil da ses som udfald af hver deres tilhørende punktproces.
En simpel model for en rumlig punktproces er Poisson processen. Poisson processen modellerer rumlig uafhængighed, hvilket vil sige, at punkterne i processen forekommer uafhængigt af hinanden. I 1955 indførte David Cox hvad han da kaldte dobbelt stokastiske processer, som senere er blevet kaldt Cox processer [14]. Cox processen er løst beskrevet en Poisson proces, som er drevet af et underliggende stokastisk felt. Denne modelklasse kan modificeres ved at lade Poisson processen være drevet af en transformation af et stokastisk felt. Et specialtilfælde af dette er de log Gaussiske Cox processer, hvor processen, er drevet af eksponentialfunktionen taget på et Gaussisk stokastisk felt. Denne modeltype blev introduceret indenfor statistik af Møller et al. i 1998 [8], og er omdrejningspunktet for dette speciale.

I specialet tages der udgangspunkt i et datasæt for hussalg i Danmark i perioden 2004 til 2016. Dette kan omdannes til et punktmønster, som repræsenterer hussalgenes geografiske placering. På grund af størrelsen af datasættet, ses der kun på en delmængde af data givet som Aalborg Kommune. I lyset af den teoretiske rammesætning for specialet kan det være interessant at undersøge, hvorvidt dette punktmønster kan modelleres ud fra en log Gaussisk Cox proces.
Problemstillingen for dette speciale kan da formuleres på følgende måde:

Hvordan kan punktmønsteret for hussalg i Aalborg Kommune modelleres med log Gaussiske Cox processer, og hvordan vurderes hvor godt disse modeller passer på datasættet?

Punktprocesser 2

I dette kapitel introduceres rumlige punktprocesser, og disse eksemplificeres ved Poisson processen. Poisson processen anvendes, i de efterfølgende kapitler, som referenceramme for hvordan punkterne i et punktmønster afhænger af hinanden. Desuden har Poisson processen en særlig rolle, idet den anvendes til at definere Cox processen i Kapitel 4. Væsentlige resultater og egenskaber såsom Slivyak-Meckes Sætning bevises i dette kapitel.
Dette kapitel er baseret på [14].

2.1 Rumlige punktprocesser

I dette afsnit gives en formel definition af rumlig punktprocesser. Herudover introduceres void sandsynligheder, som viser sig at være nok til at karakterisere fordelingen af en rumlig punktproces.

Lad S definere en delmængde af $\mathbb{R}^d$.
Kort beskrevet er en *rumlig punktproces* X en stokastisk tællelig delmængde af rummet $S \subseteq \mathbb{R}^d$. For en klarere definition af en rumlig punktproces introduceres følgende nødvendige begreber.

Definition 2.1.1 (Lokalt endelig punktkonfiguration)
For enhver delmængde $x \subseteq S$ kaldet en *punktkonfiguration* lades $n(x)$ være kardinaliteten af x.
For en begrænset delmængde $B \subseteq S$ siges x at være *lokalt endelig* hvis $n(x_B) < \infty$, hvor $x_B = x \cap B$ er restriktionen af x til B.

På en lignende måde er X_B restriktion af punktprocessen X til $B \subseteq S$.

Definition 2.1.2 (Rummet for lokalt endelige punktkonfigurationer)
Rummet bestående af lokalt endelige punktkonfigurationer defineres som

$$N_{le} = \{x \subseteq B : n(x_B) < \infty \text{ for alle begrænsede } B \subseteq S\}.$$

Definitionen af rumlige punktprocesser på S indebærer en målelig afbildning, som kan sikres ved at tildele S en Borel-algebra $\mathcal{B}$. På lignende facon tildeles N_{le} en sigma algebra

$$\mathcal{N}_{le} = \sigma\big(\{x \in N_{le} : n(x_B) = m\} : B \in \mathcal{B}_0, m \in \mathbb{N}_0\big),$$

hvor $\mathcal{B}_0$ betegner mængden af begrænsede Borel-mængder. Herefter følger definitionen på rumlige punktprocesser.

Definition 2.1.3 (Rumlige punktprocesser)
En punktproces X på $S \subseteq \mathbb{R}^d$ er en målelig afbildning $X : (\Omega, \mathcal{F}, \mathcal{P}) \rightarrow (N_{le}, \mathcal{N}_{le})$ mellem et sandsynlighedsrum $(\Omega, \mathcal{F}, \mathcal{P})$ og det målelige rum $(N_{le}, \mathcal{N}_{le})$. Fordelingen P_X af X er givet $P_X(F) = P(\{\omega \in \Omega : X(\omega) \in F\})$ for $F \in \mathcal{N}_{le}$.

Rumlige punktprocesser kan forstås som målelige afbildninger mellem et sandsynlighedsrum og et måleligt rum, som er en delmængde af $\mathbb{R}^d$. Da X antager værdier i N_{le}, er en punktkonfiguration x et udfald for en punktproces X, som observeres på en delmængde $B \subseteq S$. Punktkonfigurationer kaldes også *punktmønstre*.

2.1.1 Karakterisering af punktprocesser ud fra void hændelser

Lad X være en punktproces på $S \subseteq \mathbb{R}^d$, og lad som før $\mathcal{B}_0 = \{B \in \mathcal{B} : B \text{ er begrænset}\}$. Da defineres *tællefunktionen* som

$$N(B) = n(X_B),$$

hvilket er det antal af punkter tilhørende X der optræder i $B \subseteq S$.
Delmængder $F_B \subseteq N_{le}$ som er på formen $F_B = \{x \in N_{le} : n(x_B) = 0\}$, hvor $B \in \mathcal{B}_0$, kaldes *void hændelser*. Void hændelser forstås som mængden af punktkonfigurationer i N_{le} uden punkter i B. Herom gælder at $X \in F_B$ hvis og kun hvis $N(B) = 0$.
Sandsynligheden for at hændelsen $N(B) = 0$ indtræffer, kaldes *void sandsynligheden*. Det viser sig, at void sandsynligheden er nok til at bestemme fordelingen af punktprocessen X. *Void sandsynligheder* er givet ved

$$v(B) = P(N(B) = 0), \quad B \in \mathcal{B}_0.$$

Her er $v(B)$ sandsynligheden for, at der ikke forekommer nogle punkter i B.
Det følgende resultat viser, at void sandsynligheder er tilstrækkelige for at karakterisere punktprocesser.

Sætning 2.1.4
En punktproces X på $S \subseteq \mathbb{R}^d$ er entydig bestemt af dens void sandsynligheder.

Bevis
Antag at borelmængden $B \subseteq S$ som punktprocessen X observeres på, kan inddeles i delmængder B_i for $i = 1, 2, \dots$ sådan at hver B_i indeholder højest et punkt. Da vil fordelingen

$P(X_{B_i} \in F)$ for $F \in \mathcal{N}_{le}$ kunne beskrives ud fra sandsynligheden for om B_i indeholder et punkt eller ikke. Da bliver $P(X_{B_i} \in F) = P(n(x_{B_i}) = 0)$ på N_{le}^0 defineret i ligning (A.1). Dermed giver Lemma A.1.5 og A.1.3, at fordelingen er givet entydigt ud fra void sandsynligheden. ■

2.1.2 Stardard beviset

Et redskab som dette speciale vil benytte sig af i flere tilfælde, er det såkaldte standard bevis. Bevisteknikken anvender følgende resultat fra integrationsteori.

Proposition 2.1.5
Lad X være en mængde med tilhørende σ-algebra $\mathbb{F}$. Så gælder for alle positive og målelige funktioner $f : X \to [0, \infty]$, at der findes en voksende følge $s_1 \leq s_2 \leq \ldots$ af simple $\mathbb{F}$-målelige funktioner $s_n : X \to [0, \infty[$ sådan at $f = \lim_{n\to\infty} s_n$ [12][s. 79].

Bevis
For bevis se [12][s. 79-80]. ■

Antag at det skal bevises, at følgende lighed er opfyldt

$$\mathbb{E} \sum_{\xi \in X} h(\xi) = \int h(\xi)\rho(\xi)d\xi, \tag{2.1}$$

hvor $h : S \to [0, \infty)$ og $\rho : S \to [0, \infty)$ er ikke negative funktioner. Senere introduceres ρ som intensitetsfunktionen i Definition 2.2.2. Antag at både venstreside og højreside af lighedstegnet i ligning (2.1) bliver til mål, når h er en indikatorfunktion. Herefter defineres henholdsvis venstresiden og højresiden af ligning (2.1) som

$$\nu_{vs}(A) = \mathbb{E} \sum_{\xi \in X} \mathbb{1}[\xi \in A]$$
$$\nu_{hs}(A) = \int \mathbb{1}[\xi \in A]\rho(\xi)d\xi = \int_A \rho(\xi)d\xi$$

for $A \subseteq S$. For at undersøge om ligheden i ligning (2.1) er opfyldt, er det på baggrund af Proposition 2.1.5 nok at undersøge, om den er opfyldt når h er en indikatorfunktion på formen $\mathbb{1}[\xi \in X]$. Det skyldes, at simple funktioner er summer af indikatorfunktioner. Ovenstående illustrerer fremgangsmåden i standard beviset. Altså for at vise at en lighed er opfyldt, kan begge sider af lighedstegnet erstattes med simple funktioner, og herefter eftertjekkes.

2.2 Poisson punktprocesser

Poisson punktprocesser eller *Poisson processer* er en modelklasse for fuldstændig rumlig uafhængighed. Derfor spiller Poisson punktprocesser en afgørende rolle, når der senere indføres mere avancerede punktproces modeller.

Først defineres *den binomiale punktproces*, da det herefter er muligt at indføre Poisson processen.

Definition 2.2.1 (Binomial punktproces)
Lad f være en tæthedsfunktion på $B \subseteq S$, hvor $S \subseteq \mathbb{R}^d$, og lad $n \in \mathbb{N}$. En punktproces X bestående af n uafhængige identisk fordelte punkter med tæthed f, kaldes en binomial punktproces af n punkter i B med tæthed f. Da skrives $X \sim \text{Binomial}(B, n, f)$.

På rummet $S \subseteq \mathbb{R}^d$ hvor Poisson punktprocessen er defineret, defineres også *intensitetsfunktionen*, som er med til at specificere Poisson punktprocessen.

Definition 2.2.2 (Intensitetsfunktion)
Hvis funktionen $\rho : S \to [0, \infty)$ er lokal integrabel, det vil sige at $\int_B \rho(\xi) d\xi < \infty$ for alle begrænsede $B \subseteq S$, da kaldes ρ for intensitetsfunktionen for punktprocessen X på S.

Bemærk at Definition 2.2.2 er generel for punktprocesser X på S, og ikke kun for Poisson processer.
Hvis intensitetsfunktionen ρ er lokalt integrabel, defineres integralet af ρ over $B \subseteq S$ som *intensitetsmålet* μ.

Definition 2.2.3 (Intensitetsmål)
Intensitetsmålet μ er givet som $\mu(B) = \int_B \rho(\xi) d\xi$ hvor $B \subseteq S$. Intensitetsmålet er lokalt endeligt, det vil sige $\mu(B) < \infty$ for begrænsede $B \subseteq S$.

Det ses, at der for intensitetsmålet gælder at $\mu(\{\xi\}) = 0$ for $\xi \in S$.

Nu er det muligt at definere en Poisson proces.

Definition 2.2.4 (Poisson punktproces)
En punktproces X på $S \subseteq \mathbb{R}^d$ er en Poisson punktproces med intensitetsfunktion ρ, hvis følgende egenskaber er gældende:

(i) For enhver delmængde $B \subseteq S$ med $\mu(B) < \infty$, er $N(B) \sim \text{po}(\mu(B))$, altså er antallet af punkter Poisson fordelt med middelværdi $\mu(B)$ (hvis $\mu(B) = 0$ så er $N(B) = 0$).

(ii) For ethvert $n \in \mathbb{N}$ og $B \subseteq S$ med $0 < \mu(B) < \infty$, betinget på $N(B) = n$, så er $X_B \sim \text{binomial}(B, n, f)$ med tæthed $f(\xi) = \frac{\rho(\xi)}{\mu(B)}$.

Da skrives $X \sim \text{Poisson}(S, \rho)$.

Der gælder altså for intensitetsmålet μ, at for ethvert begrænset $B \subseteq S$ bestemmer μ det forventede antal punkter i B. Det kan skrives

$$\mu(B) = \mathbb{E}[N(B)]. \tag{2.2}$$

Ud fra Definition 2.2.3 kan $\rho(\xi)d\xi$ ses som sandsynligheden for, at et punkt optræder på en uendelig lille kugle med centrum ξ og volumen $d\xi$. Dette betyder, at hvis intensiteten er konstant, vil sandsynligheden for om et punkt forekommer være ens overalt på S.

Definition 2.2.5 (Homogen og inhomogen Poisson proces)
Hvis ρ er konstant, så kaldes Poisson punktprocessen en *homogen Poisson proces* på S med intensitet ρ. Hvis ρ ikke er konstant siges processen at være en *inhomogen Poisson proces* på S.

Der gælder for en homogen Poisson proces, at den både er translations invariant og isotropisk.

Definition 2.2.6 (Translations invariant og isotropisk punktproces)
En punktproces X på $\mathbb{R}^d$ er translationsinvariant, hvis fordelingen af $X + s = \{\xi + s : \xi \in X\}$ er den samme fordeling som den for X givet ethvert $s \in \mathbb{R}^d$.
Punktprocessen X er isotropisk, hvis fordelingen af $\mathcal{O}X = \{\mathcal{O}\xi : \xi \in X\}$ er den samme fordeling som den for X for enhver rotation $\mathcal{O}$ om Origo.

Følgende proposition giver en måde til at identificere om X er en Poisson proces. Desuden giver propositionen et udtryk for fordelingen af X_B for en Poisson proces X.

Proposition 2.2.7 (i) Der gælder, at $X \sim \text{Poisson}(S, \rho)$ hvis og kun hvis der for alle $B \subseteq S$ med $\mu(B) = \int_B \rho(\xi)d\xi < \infty$ og alle $F \subseteq N_{le}$ kan skrives:

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

hvor integralet for $n = 0$ skal ses som $\mathbb{1}[\emptyset \in F]$.

(ii) Hvis der gælder at $X \sim \text{Poisson}(S, \rho)$, så gælder der for funktioner $h : N_{le} \to [0, \infty)$ og $B \subseteq S$ med $\mu(B) < \infty$ at

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

Bevis

(i) Det antages $X \sim \text{Poisson}(S, \rho)$, hvoraf der gælder at $N(B) \sim \text{po}(\mu(B))$ per Definition 2.2.4. Poissonfordelingens tæthed er

$$P(N(B)) = \exp(-\mu(B)) \frac{\mu(B)^k}{k!}.$$

For at udregne $P(X_B \in F)$ anvendes loven om total sandsynlighed

$$P(X_B \in F) = \sum_{n=0}^{\infty} P\big(X_B \in F | N(B) = n\big) P\big(N(B) = n\big).$$

For at udregne $P\big(X_B \in F | N(B) = n\big)$ anvendes at udfaldene er uafhængige, samt at sandsynligheden kan findes ved at integrere tætheden. Deraf fås

$$P(X_B \in F) = \sum_{n=0}^{\infty} \Big(P\big(N(B) = n\big)\Big) \int \cdots \int_B \mathbb{1}\big[\{x_1, \dots, x_n\} \in F\big] \prod_{i=1}^{n} f(x_i) dx_1, \dots, x_n,$$

hvor indikatorfunktionen indføres, for at sikre at punkterne er i F. Til sidst indsættes Poissontætheden for $P(N(B) = n)$ samt at $f(x_i) = \frac{\rho(x_i)}{\mu(B)}$

$$\begin{aligned} P(X_B \in F) &= \sum_{n=0}^{\infty} \exp(-\mu(B)) \frac{\mu(B)^n}{n!} \int \cdots \int_B \mathbb{1}\big[\{x_1, \dots, x_n\} \in F\big] \prod_{i=1}^{n} \frac{\rho(x_i)}{\mu(B)} dx_1, \dots, x_n \\ &= \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int \cdots \int_B \mathbb{1}\big[\{x_1, \dots, x_n\} \in F\big] \prod_{i=1}^{n} \rho(x_i) dx_1, \dots, x_n \end{aligned}$$

(ii) Det skal nu vises at følgende lighed holder

$$\mathbb{E}[h(X_B)] = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B h(\{x_1, \dots, x_n\}) \prod_{i=1}^{n} \rho(x_i) dx_1 \cdots dx_n. \tag{2.3}$$

Teknikken bag standard beviset anvendes, så det antages at h kan skrives som grænseværdien af en følge af simple funktioner: $s_n : N_{le} \to [0, \infty)$. Dermed kan h erstattes med en indikatorfunktion for at undersøge, om højresiden og venstresiden i ligning (2.3) er ens. Dette giver

$$\mathbb{E}\big[\mathbb{1}[X_B \in F]\big] = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B \mathbb{1}[\{x_1, \ldots, x_n\} \in F] \prod_{i=1}^{n} \rho(x_i) dx_1 \cdots dx_n. \tag{2.4}$$

Men idet indikatorfunktionen har netop to udfald 1 og 0 fås

$$\mathbb{E}\big[\mathbb{1}[X_B \in F]\big] = 1 \cdot P(X_B \in F) + 0 \cdot P(X_B \notin F) = P(X_B \in F)$$

hvilket per (i) i Proposition 2.2.7 opfylder ligheden i ligning (2.4). Heraf følger det af standard beviset, at ligning (2.3) er opfyldt. ■

Følgende sætning viser eksistensen og entydigheden for en Poisson proces givet dens void sandsynlighed. Bemærk, at hvis X er en Poisson proces er $N(B) \sim \text{po}(\mu(B))$, hvormed void sandsynligheden er givet

$$v(B) = P(N(B) = 0) = \exp(-\mu(B)) \frac{\mu(B)^0}{0!} = \exp(-\mu(B)). \tag{2.5}$$

Sætning 2.2.8

Poisson punktprocessen, $X \sim \text{Poisson}(S, \rho)$, eksisterer og er entydig bestemt ved dens void sandsynlighed

$$v(B) = \exp(-\mu(B)), \quad \text{for alle begrænsede } \; B \subseteq S.$$

Bevis

Eksistensen vises ved at konstruere Poisson processen. Lad $\xi_0 \in S$ være et vilkårligt punkt og sæt $B_i = \{\eta \in S : i - 1 \leq ||\eta - \xi_0|| < i\}$ for $i \in \mathbb{N}$. Det ses at S er en disjunkt forening af de begrænsede B_i mængder, da B_i for hvert i vil indeholde forskellige elementer. Betragt en følge af Poisson fordelte stokastiske variable $N_i \sim \text{po}(\mu(B_i))$. Betinget på $N_i = n$ lades X_i være en binomial proces på B_i med n punkter, mål μ_i og tæthed $f(\xi) = \frac{\rho_i(\xi)}{\mu(B_i)}$, hvor μ_i og ρ_i er restriktionen af henholdsvis μ og ρ til B_i. Hvis processerne X_i konstrueres uafhængigt af hinanden, vises at $X = \bigcup_{i=1}^{\infty} X_i$ er en Poisson proces på S. For begrænsede $B \subseteq S$ fås af loven om total sandsynlighed at,

$$P(n(X_B) = 0) = \prod_{i=1}^{\infty} P(n(X_i \cap B) = 0) = \prod_{i=1}^{\infty} \sum_{n=0}^{\infty} P(n(X_i \cap B) = 0 | N_i = n) P(N_i = n). \tag{2.6}$$

Da N_i er Poisson fordelt er $P(N_i = n) = \exp(-\mu(B_i))\frac{\mu(B_i)^n}{n!}$.
Bemærk at den betingede void sandsynlighed $P(n(X_i \cap B) = 0 | N_i = n)$ kan skrives som

$$P(n(X_i \cap B) = 0 | N_i = n) = (1 - P(\xi_1 \in X_i \cap B)) \cdots (1 - P(\xi_n \in X_i \cap B)).$$

Tætheden integreres for at bestemme sandsynligheden så

$$P(n(X_i \cap B) = 0 | N_i = n) = \left(1 - \int_{B \cap B_i} \frac{\rho_i(\xi)}{\mu(B_i)} d\xi\right)^n = \left(1 - \frac{\mu(B \cap B_i)}{\mu(B_i)}\right)^n.$$

Dette indsættes i ligning (2.6). Ved at anvende at eksponentialfunktionen på potensrækkeformen er givet

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

fås af ligning (2.6)

$$\begin{aligned}
P(n(X_i \cap B) &= \prod_{i=1}^{\infty} \sum_{n=0}^{\infty} \left(1 - \frac{\mu(B \cap B_i)}{\mu(B_i)}\right)^n \exp(-\mu(B_i)) \frac{\mu(B_i)^n}{n!} \\
&= \prod_{i=1}^{\infty} \sum_{n=0}^{\infty} \frac{\big(\mu(B_i) - \mu(B \cap B_i)\big)^n}{n!} \exp(-\mu(B_i)) \\
&= \prod_{i=1}^{\infty} \exp(\mu(B_i) - \mu(B \cap B_i)) \exp(-\mu(B_i)) \\
&= \prod_{i=1}^{\infty} \exp(-\mu(B \cap B_i)) = \exp\left(-\sum_{i=1}^{\infty} \mu(B \cap B_i)\right) \\
&= \exp(-\mu(B)). \qquad (2.7)
\end{aligned}$$

Udtrykket i ligning (2.7) er void sandsynligheden for en Poisson proces med intensitetsmål μ. Dermed fås også entydigheden af Poisson processen fra Sætning 2.1.4 [9][s. 5]. ■

En udvidelse af Sætning 2.2.8 til uafhængige Poisson processer er, at hvis X_1 og X_2 begge er uafhængige Poisson processer på S med intensitetsmål μ_1 og μ_2, så er fordelingen af (X_1, X_2) entydigt bestemt af deres void sandsynligheder. Disse skrives

$$P(X_1 \cap A = \emptyset, X_2 \cap B = \emptyset) = \exp(-\mu_1(A) - \mu_2(B)) \qquad (2.8)$$

for begrænsede delmængder $A, B \subseteq S$.
Nedenstående proposition viser, at der for disjunkte mængder er fuldstændig rumlig uafhængighed i en Poisson proces.

Proposition 2.2.9
Hvis X er en Poisson proces på S, så er $X_{B_1}, X_{B_2}, \dots,$ uafhængige for disjunkte mængder $B_1, B_2, \dots \subseteq S$.

Bevis
Det skal vises, at $X_{B_1}, X_{B_2}, \dots, X_{B_n}$ er uafhængige for disjunkte mængder $B_1, \dots, B_n$ for $n \geq 2$. For $n = 2$ haves, at $B_1, B_2 \subseteq S$ er disjunkte og begrænsede, da sættes $B = B_1 \cup B_2$. Det der vises er

$$P(X_{B_1} \in F_1, X_{B_2} \in F_2) = P(X_{B_1} \in F_1)P(X_{B_2} \in F_2).$$

Det anvendes at X er en Poisson proces på $B \subseteq S$, hvormed Proposition 2.2.7 giver

$$\begin{aligned}
&P(X_{B_1} \in F_1, X_{B_2} \in F_2) = \\
&\quad \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_B \cdots \int_B \\
&\quad \mathbb{1}[\{x_1, \dots x_n\} \cap B_1 \in F_1, \{x_1, \dots x_n\} \cap B_2 \in F_2] \prod_{i=1}^{n} \rho(x_i) dx_1, \dots, dx_n \\
&= \sum_{n=0}^{\infty} \frac{\exp(-\mu(B_1 \cup B_2))}{n!} \sum_{k=0}^{n} \binom{n}{k} \int_{B_1} \cdots \int_{B_1} \int_{B_2} \cdots \int_{B_2} \\
&\quad \mathbb{1}[\{x_1, \dots x_k\} \in F_1] \mathbb{1}[\{x_{k+1}, \dots x_n\} \in F_2] \prod_{i=1}^{n} \rho(x_i) dx_1, \dots, dx_n.
\end{aligned} \tag{2.9}$$

Anden lighed i ligning (2.9) følger af, at der er $\binom{n}{k}$ måder at trække k ud af n punkter, samt at de individuelle sandsynligheder summeres.
Herefter flyttes den indre sum udenfor, og sumtegnene omarrangeres, hvilket giver

$$\begin{aligned}
&P(X_{B_1} \in F_1, X_{B_2} \in F_2) = \\
&\quad \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} \frac{\exp(-\mu(B_1) - \mu(B_2))}{n!} \frac{n!}{k!(n-k)!} \\
&\quad \int_{B_1^k} \int_{B_2^{n-k}} \mathbb{1}[\{x_1, \dots x_k\} \in F_1] \mathbb{1}[\{x_{k+1}, \dots x_n\} \in F_2] \prod_{i=1}^{n} \rho(x_i) dx_1, \dots, dx_n \\
&= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} \frac{\exp(-\mu(B_1) - \mu(B_2))}{k!(n-k)!} \int_{B_1^k} \mathbb{1}[\{x_1, \dots x_k\} \in F_1] \prod_{i=1}^{k} \rho(x_i) dx_1, \dots, dx_k \\
&\quad \int_{B_2^{n-k}} \mathbb{1}[\{x_{k+1}, \dots x_n\} \in F_2] \prod_{i=k+1}^{n} \rho(x_i) dx_{k+1}, \dots, dx_n.
\end{aligned}$$

Anden lighed i ovenstående følger af at flytte de funktioner, som ikke afhænger af $x_1, \dots, x_k$ udenfor integralet der afhænger af $x_{k+1}, \dots, x_n$. Ved at reparametrisere $m = n - k$ fås da det følgende

$$
\begin{aligned}
&P(X_{B_1} \in F_1, X_{B_2} \in F_2) = \\
&\quad \sum_{k=0}^{\infty} \frac{\exp(-\mu(B_1))}{k!} \int_{B_1^k} \mathbb{1}[\{x_1, \ldots x_k\} \in F_1] \prod_{i=1}^{k} \rho(x_i) dx_1, \ldots, dx_k \cdot \\
&\quad \sum_{m=0}^{\infty} \frac{\exp(-\mu(B_2))}{m!} \int_{B_2^m} \mathbb{1}[\{x_1, \ldots x_m\} \in F_2] \prod_{i=1}^{m} \rho(x_i) dx_1, \ldots, dx_m \\
&= P(X_{B_1} \in F_1) P(X_{B_2} \in F_2).
\end{aligned}
$$

Den sidste lighed følger af genanvendelse af Proposition 2.2.7(i).
I induktionstrinnet antages, at

$$P(X_{B_1} \in F_1, \ldots, X_{B_j} \in F_j) = P(X_{B_1} \in F_1) \cdots P(X_{B_j} \in F_j)$$

er opfyldt for disjunkte $B_1, \ldots, B_j \subseteq S$. For at vise induktionstrinnet skal det vises at $X_{B_1}, \ldots, X_{B_{j+1}}$ er uafhængige. Dette følger direkte af anvendelse af basistrinnet for $n = 2$ og induktionsantagelsen så

$$P(X_{B_1} \in F_1, \ldots, X_{B_{j+1}} \in F_{j+1}) = P(X_{B_1} \in F_1) \cdots P(X_{B_{j+1}} \in F_{j+1}).$$

■

Sætning 2.2.9 viser, at Poisson processen er uafhængig, hvis S indeles i disjunkte mængder. Poisson processen udviser, hvad der betegnes fuldstændig rumlig uafhængighed, ved at alle punkter, som er forskellige fra hinanden, er uafhængige. I Kapitel 3 indføres kriterier, som kan anvendes til at afgøre, hvorvidt en punktproces udviser fuldstændig rumlig uafhængighed, hvilket er ækvivalent med, om en punktproces er en Poisson proces.
Følgende proposition angiver en metode til at konstruere homogene Poisson processer, ud fra uafhængige delmængder af $\mathbb{R}^d$ med passende fordelinger.
Lad ω_d være volumen af den d-dimensionelle enhedskugle givet som

$$\omega_d = \pi^{d/2} / \Gamma(1 + d/2).$$

Proposition 2.2.10
Lad $S_1, U_1, S_2, U_2, \ldots$ være gensidigt uafhængige, hvor hvert U_i er uniformt fordelt på $\{u \in \mathbb{R}^d : ||u|| = 1\}$, og hvert $S_i \sim \text{Exp}(\rho\omega_d)$ er eksponentielt fordelt for $\rho > 0$. Lad $R_0 = 0$ og $R_i^d = R_{i-1}^d + S_i$ for $i = 1, 2, \ldots$. Da er $X = \{R_1U_1, R_2U_2, \ldots\} \sim \text{Poisson}(\mathbb{R}^d, \rho)$.

Bevis
For bevis se [14][s. 18-19]. ■

2.2.1 Superposition og udtynding

To grundlæggende operationer for punktprocesser er *superpositionering* og *udtynding*.

Definition 2.2.11 (Superposition)
En disjunkt forening $\bigcup_{i=1}^{\infty} X_i$ af punktprocesser $X_1, X_2, \ldots,$ kaldes en superposition.

Definition 2.2.12 (Udtynding)
Lad $p : S \to [0,1]$ være en funktion og lad X være en punktproces på S. Punktprocessen $X_{thin} \subseteq X$ som er opnået ved at inkludere $\xi \in X$ i X_{thin} med sandsynlighed $p(\xi)$, hvor punkterne er inkluderet/ekskluderet uafhængigt af hinanden, kaldes en uafhængig udtynding af X med *retention sandsynligheder* $p(\xi), \ \ \xi \in S$. Dette kan også skrives som

$$X_{thin} = \{\xi \in X : R(\xi) \leq p(\xi)\},$$

hvor $R(\xi) \sim \text{Uniform}[0,1]$ med $\xi \in S$ er gensidigt uafhængige og uafhængige af X.

Det viser sig, at Poisson processer er lukkede under superposition. Følgende proposition anvender terminologien: *Med sandsynlighed et*, som forklares i Appendiks A.1.

Proposition 2.2.13
Hvis $X_i \sim \text{Poisson}(S, \rho_i), \ \ i = 1, 2, \ldots,$ er gensidigt uafhængige og $\rho = \sum \rho_i$ er lokalt integrabel, så er $X = \bigcup_{i=1}^{\infty} X_i$, med sandsynlighed et, en disjunkt forening og $X \sim \text{Poisson}(S, \rho)$.

Bevis
Første del vises ikke, men kan findes i [14][s. 23].

At $X = \bigcup_{i=1}^{\infty} X_i$ er en Poisson proces følger af Sætning 2.2.8, altså at den eksisterer og er entydig bestemt ved dens void sandsynlighed. For begrænset $B \subseteq S$ kan der på baggrund af Sætning 2.2.8 skrives

$$P(X_B = \emptyset) = \prod_{i=1}^{\infty} P(X_i \cap B = \emptyset) = \prod_{i=1}^{\infty} \exp(-\mu_i(B)) = \exp(-\mu(B))$$

hvor $\mu_i(B) = \int_B \rho_i(\xi) d\xi$. ■

Næste proposition viser, at en Poisson punktproces også er lukket under uafhængig udtynding.

Proposition 2.2.14

Antag at $X \sim$ Poisson(S, ρ) er underlagt uafhængig udtynding med retention sandsynligheder $p(\xi)$ hvor $\xi \in S$, og lad

$$\rho_{thin} = p(\xi)\rho(\xi), \;\; \xi \in S.$$

Så er X_{thin} og $X \backslash X_{thin}$ uafhængige Poisson processer med intensitets funktioner henholdsvis ρ_{thin} og $\rho - \rho_{thin}$.

Bevis

Lad μ_{thin} være givet ved $\mu_{thin} = \int \rho_{thin}(\xi) d\xi$. Givet udvidelsen af Sætning 2.2.8 giver ligning (2.8), at det kun er nødvendigt at vise

$$P(X_{thin} \cap A = \emptyset, X \backslash X_{thin} \cap B = \emptyset) = \exp(-\mu_{thin}(A) - \mu(B) + \mu_{thin}(B)) \tag{2.10}$$

for begrænsede $A, B \subseteq S$. Herfra vises det, at X_{thin} samt $X \setminus X_{thin}$ er Poisson processer.
Lad $C \subseteq S$ være begrænset. Da $X \sim$ Poisson(S, ρ) er $N(C) \sim$ po$(\mu(C))$.
Loven om total sandsynlighed giver for udtrykket $P(X_{thin} \cap C = \emptyset)$ at

$$\begin{aligned} P(X_{thin} \cap C = \emptyset) &= \sum_{n=0}^{\infty} P\big(X_{thin} \cap C = \emptyset | N(C) = n\big) P\big(N(C) = n\big) \\ &= \sum_{n=0}^{\infty} \frac{\mu(C)^n}{n!} e^{-\mu(C)} P\big(X_{thin} \cap C = \emptyset | N(C) = n\big). \end{aligned}$$

For at udregne $P\big(X_{thin} \cap C = \emptyset | N(C) = n\big)$ anvendes Definition 2.2.12, hvor $X_{thin} \cap C | N(C)$ har tæthed $p(\xi) f(\xi)$, og f er tætheden for $X \cap C | N(C)$. Dermed må $X_{thin} \cap C = \emptyset | N(C) = n$ have tæthed $(1 - p(\xi)) f(\xi)$. Dette benyttes til at bestemme sandsynligheden ved at integrere tætheden

$$\begin{aligned} P(X_{thin} \cap C = \emptyset) &= \sum_{n=0}^{\infty} \frac{\mu(C)^n}{n!} e^{-\mu(C)} \Big(\int_C \big(1 - p(\xi)\big) f(\xi) d\xi \Big)^n \\ &= \sum_{n=0}^{\infty} \frac{\mu(C)^n}{n!} e^{-\mu(C)} \Big(\int_C \big(1 - p(\xi)\big) \frac{\rho(\xi)}{\mu(C)} d\xi \Big)^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} e^{-\mu(C)} \left(\int_C (1 - p(\xi)) \rho(\xi) d\xi \right)^n \\ &= \exp\left(\int_C (1 - p(\xi)) \rho(\xi) d\xi \right) \exp(-\mu(C)) \\ &= \exp(\mu(C) - \mu_{thin}(C)) \exp(-\mu(C)) = \exp(-\mu_{thin}(C)). \end{aligned}$$

Her er potensrækkeformen for eksponentialfunktionen anvendt, samt at udtyndingen er uafhængig. Per Sætning 2.2.8 er X_{thin} er en Poisson proces.

Ved symmetri har $X \setminus X_{thin} \cap C = \emptyset | N(C) = n$ tæthed $p(\xi)f(\xi)$, og det fås at

$$P((X \backslash X_{thin}) \cap C = \emptyset) = \exp(-(\mu - \mu_{thin})(C)).$$

Da fås, at $X \backslash X_{thin}$ også er en Poisson proces, som er karakteriseret ved dens void sandsynlighed.
Det ses, at $A \cap B, A \backslash B, B \setminus A$ er disjunkte og at $\{X_{thin} \cap A, X \backslash X_{thin} \cap B\} = \{X \cap A \cap B, X_{thin} \cap (A \backslash B), (X \backslash X_{thin}) \cap (B \backslash A)\}$ for begrænsede $A, B \subseteq S$.
Resultatet følger nu af Proposition 2.2.9 og ligning (2.8).

$$\begin{aligned}
&P(X_{thin} \cap A = \emptyset, (X \backslash X_{thin}) \cap B = \emptyset) \\
&= P(X \cap A \cap B = \emptyset) P(X_{thin} \cap (A \backslash B) = \emptyset) P((X \backslash X_{thin}) \cap (B \backslash A) = \emptyset) \\
&= \exp(-\mu(A \cap B) - \mu_{thin}(A \backslash B) - (\mu - \mu_{thin})(B \backslash A))
\end{aligned}$$

Dermed er ligning (2.10) opfyldt. ■

Inhomogene Poisson processer kan ofte simuleres ved udtynding af en homogen Poisson proces. Dette viser følgende korollar.

Korollar 2.2.15
Antag at $X \sim$ Poisson$(\mathbb{R}^d, \rho)$, hvor intensitetsfunktionen ρ er begrænset af en endelig konstant c. Så er X fordelt som en uafhængig udtynding af Poisson$(\mathbb{R}^d, c)$ med retention sandsynligheder $p(\xi) = \frac{\rho(\xi)}{c}$.

Bevis
Per antagelse er void sandsynligheden for X givet ved

$$v(B) = P(X \cap B = \emptyset) = \exp\Big(-\int_B \rho(\xi) d\xi\Big) = \exp(-\mu(B)) \tag{2.11}$$

jævnfør Sætning 2.2.8.
Givet en homogen Poisson proces $X_{hom} \sim$ Poisson(S, c), lad da X_{hom} være underlagt en uafhængig udtyndig med retention sandsynligheder $p(\xi) = \rho(\xi)/c$. Så bliver $\rho_{thin}(\xi) = p(\xi)c$. Da er den uafhængige udtyndning X_{thin} af X_{hom} en Poisson proces med intensitet $\rho_{thin}(\xi) = \rho(\xi)$ per Proposition 2.2.14.
X_{thin} har da void sandsynlighed

$$P(X_{thin} \cap B = \emptyset) = \exp\Big(-\int_B \rho_{thin}(\xi) d\xi\Big) = \exp\Big(-\int_B \rho(\xi) d\xi\Big) = \exp(-\mu(B)),$$

hvormed X er fordelt som X_{thin} per Sætning 2.2.8.

■

2.2.2 Simulation af Poisson processen

Afsnittet her vil fokusere på simulation af en Poisson proces $X \sim \text{Poisson}(\mathbb{R}^d, \rho)$ indenfor en begrænset mængde $B \subseteq \mathbb{R}^d$. Det homogene tilfælde hvor intensiteten $\rho(\xi) = \rho_0$ er konstant for alle $\xi \in B$ betragtes som det første.
For dimension $d \geq 1$ betragtes tre tilfælde:

1. I tilfældet hvor $B = b(0, r)$ er en kugle med centrum i 0 og radius r, følger simulationen af X_B direkte af Proposition 2.2.10: Simuler $S_1, \dots, S_m$ og $U_1, \dots, U_m$ hvor m er givet ved at $R_{m-1} \leq r < R_m$ og $R_0 = 0$ samt $R_i^d = R_{i-1}^d + S_i$, $i = 1, 2, \dots$. Da gives restriktionen af X til B ved $X_B = \{R_1U_1, \dots, R_{m-1}U_{m-1}\}$.

2. Hvis $B = [0, a_1] \times \cdots \times [0, a_d]$ er en d-dimensionel kasse da kan (i) og (ii) fra Definition 2.2.4 anvendes. Først kan antallet af punkter $N(B)$ genereres da $N(B) \sim \text{po}(\rho_o a_1 \cdots a_d)$, og herefter kan lokationerne af de $N(B)$ uniformt uafhængige fordelte punkter genereres.

3. Hvis B er begrænset, men hverken er en kugle eller kasse, da simuleres Poisson processen X på enten en kugle eller kasse B_0 hvor $B \subset B_0$. Herefter fjernes de genererede punkter som optræder i $B_0 \setminus B$.

Herefter antages, at X_B er inhomogen med intensitetsfunktion $\rho(\xi)$ med $\xi \in B$, som er begrænset af en konstant $\rho_0 > 0$. Da genereres først en homogen Poisson proces Y på B ved brug af ovenstående punkter, med tilhørende intensitet ρ_0. Herefter dannes X_B som en uafhængig udtynding af Y_B med retention sandsynligheder $p(\xi) = \rho(\xi)/\rho_0, \;\; \xi \in B$. Denne fremgangsmåde er muliggjort af Proposition 2.2.14.

2.2.3 Slivnyak-Meckes Sætning

En anden metode til at karakterisere en Poisson proces er ved Slivnyak-Meckes Sætning. Denne viser, at middelværdien over en sum af funktioner på S kan omskrives som integralet over middelværdien af denne funktion gange intensiteten.

Sætning 2.2.16 (Slivnyak-Meckes Sætning)
Hvis $X \sim \text{Poisson}(S, \rho)$ så gælder der for funktioner $h : S \times N_{le} \to [0, \infty)$ at

$$\mathbb{E}\Big[\sum_{\xi \in X} h(\xi, X \setminus \xi)\Big] = \int_S \mathbb{E}\big[h(\xi, X)\big]\rho(\xi)d\xi. \tag{2.12}$$

Bevis
Først betragtes tilfældet hvor $\sum_{\xi \in X} h(\xi, X \setminus \xi)$ kun afhænger af X gennem X_B for et $B \subseteq S$ med $\int_B \rho(\xi)d\xi < \infty$. Dermed kan Proposition 2.2.7 anvendes til at finde den forventede værdi

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{\xi\in X} h(\xi, X\setminus\xi)\Big]\\
&=\sum_{n=1}^{\infty}\frac{\exp(-\mu(B))}{n!}\int_B\cdots\int_B\sum_{i=1}^{n} h(x_i,\{x_1,\dots,x_n\}\setminus x_i)\prod_{i=1}^{n}\rho(x_i)dx_1\cdots dx_n. \qquad (2.13)
\end{aligned}$$

Idet at funktionerne integreres over samme integrationsområde, kan summen skrives som n gange funktionsværdien, hvormed ligning 2.13 kan omskrives til

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{\xi\in X} h(\xi, X\setminus\xi)\Big]\\
&=\sum_{n=1}^{\infty}\frac{\exp(-\mu(B))}{(n-1)!}\int_B\cdots\int_B h(x_n,\{x_1,\dots,x_{n-1}\})\prod_{i=1}^{n}\rho(x_i)dx_1\cdots dx_n\\
&=\int_B\sum_{k=0}^{\infty}\frac{\exp(-\mu(B))}{k!}\int_B\cdots\int_B h(x_{k+1},\{x_1,\dots,x_k\})\prod_{i=1}^{k+1}\rho(x_i)dx_1\cdots dx_{k+1}\\
&=\int_B\sum_{k=0}^{\infty}\frac{\exp(-\mu(B))}{k!}\mu(B)^k\int_B\cdots\int_B h(u,\{x_1,\dots,x_k\})\prod_{i=1}^{k}\frac{\rho(x_i)}{\mu(B)}dx_1\cdots dx_k\rho(u)du.
\end{aligned}\tag{2.14}$$

Det gælder, at $\frac{\rho(x_i)}{\mu(B)} = f(x_i)$ er tætheden for X betinget på $N(B)$, da X er en Poisson proces. Dermed kan

$$\int_B\cdots\int_B h(u,\{x_1,\dots,x_k\})\prod_{i=1}^{n}\frac{\rho(x_i)}{\mu(B)}dx_1\cdots dx_k$$

fra ligning (2.14) omskrives til middelværdien af h betinget på $N(B)$. Dermed fås, at ligning (2.14) kan skrives

$$\begin{aligned}
\mathbb{E}\Big[\sum_{\xi\in X} h(\xi, X\setminus\xi)\Big] &= \int_B\sum_{k=0}^{\infty}\frac{\exp(-\mu(B))}{k!}\mu(B)^k\mathbb{E}\big[h(u,\{x_1,\dots,x_k\})|N(B)=k\big]\rho(u)du\\
&=\int_B\mathbb{E}\Big[\mathbb{E}\big[h(u,\{x_1,\dots,x_k\})|N(B)=k\big]\Big]\rho(u)du
\end{aligned}\tag{2.15}$$

hvor det i ligning (2.15), anvendes at der summeres over tætheden for Poissonfordelingen. Til sidst anvendes loven om total forventning $E[Y] = E[E[Y|X]]$ [2][s. 185], og af ligning (2.15) fås

$$\mathbb{E}\Big[\sum_{\xi\in X} h(\xi, X\setminus\xi)\Big] = \int_B\mathbb{E}[h(u,\{x_1,\dots,x_k\})]\rho(u)du = \int_B\mathbb{E}[h(u,X)]\rho(u)du.$$

Herefter betragtes det generelle tilfælde, hvor $h: S\times N_{le}\to[0,\infty)$.
Hvis det kan vises, at ligning (2.12) er opfyldt, når $h(\xi,x) = \mathbb{1}[(\xi,x)\in F],\quad F\subseteq S\times N_{le}$ er en indikatorfunktion, så følger det af standardbeviset i Afsnit 2.1.2, at ligning (2.12) gælder for alle

ikke-negative funktioner h. Ideen er at anvende Lemma A.1.6, på højresiden og venstresiden af ligning (2.12), når h er en indikatorfunktion.
Det første der skal vises, er at hvis $h(\xi, x) = \mathbb{1}[(\xi, x) \in F], F \subseteq S \times N_{le}$, så er både højresiden og venstresiden af ligning (2.12) et mål. Lad

$$v_s(F) = \mathbb{E}\Big[\sum_{\xi \in X} \mathbb{1}[(\xi, x \backslash \xi) \in F]\Big], \tag{2.16}$$

$$h_s(F) = \int_S \mathbb{E}\big[\mathbb{1}[(\xi, x) \in F]\big]\rho(\xi)d\xi \tag{2.17}$$

for $F \subseteq S \times N_{le}$ betegne henholdsvis venstresiden og højresiden af (2.12), hvor h er erstattet af en indikatorfunktion. Det første der bemærkes, er at både indikatorfunktionen og ρ, per definition er ikke-negative, hvormed både v_s og h_s er ikke-negative. Dette opfylder det første kriterie i Definition A.1.2. Desuden ses at

$$v_s(\emptyset) = \mathbb{E}\Big[\sum_{\xi \in X} \mathbb{1}[(\xi, x \backslash \xi) \in \emptyset]\Big] = \mathbb{E}\Big[\sum_{\xi \in X} 0\Big] = 0$$

og

$$h_s(\emptyset) = \int_S \mathbb{E}\big[\mathbb{1}[(\xi, x) \in \emptyset]\big]\rho(\xi)d\xi = \int_S 0 \cdot \rho(\xi)d\xi = 0,$$

hvilket opfylder (1) i Definition A.1.2. Lad nu $S = \bigcup_{n=1}^{\infty} B_n$ for disjunkte delmængder B_n sådan at $S \times N_{le} = \big(\bigcup_{n=1}^{\infty} B_n\big) \times N_{le} = \bigcup_{n=1}^{\infty} F_n$ består af en disjunkt forening af elementer $B_n \bigcup N_{le} = F_n$. Da fås

$$\begin{aligned} v_s\big(\bigcup_{n=1}^{\infty} F_n\big) &= \mathbb{E}\Big[\sum_{\xi \in X} \mathbb{1}[(\xi, x \setminus \xi) \in \bigcup_{n=1}^{\infty} F_n]\Big] \\ &= \mathbb{E}\Big[\sum_{\xi \in X} \sum_{n=1}^{\infty} \mathbb{1}[(\xi, x \setminus \xi) \in F_n]\Big] = \sum_{n=1}^{\infty} \mathbb{E}\Big[\sum_{\xi \in X} \mathbb{1}[(\xi, x \setminus \xi) \in F_n]\Big] \end{aligned}$$

samt

$$\begin{aligned} h_s\big(\bigcup_{n=1}^{\infty} F_n\big) &= \int_{\bigcup_{n=1}^{\infty} B_n} \mathbb{E}\Big[\mathbb{1}\big[(\xi, x) \in \bigcup_{n=1}^{\infty} F_n\big]\Big]\rho(\xi)d\xi \\ &= \int_{B_1} \mathbb{E}\Big[\mathbb{1}\big[(\xi, x) \in \bigcup_{n=1}^{\infty} F_n\big]\Big]\rho(\xi)d\xi + \cdots \\ &= \sum_{n=1}^{\infty} \int_{B_n} \mathbb{E}\Big[\mathbb{1}\big[(\xi, x) \in \bigcup_{n=1}^{\infty} F_n\big]\Big]\rho(\xi)d\xi, \end{aligned}$$

hvilket viser (2) i Definition A.1.2. Dette viser altså at både v_s og h_s er mål.
Fra første del af beviset vides, at hvis h er på formen

$$h(\xi, x) = \mathbb{1}[\xi \in C, n(x_A) = 0]$$

hvor $A, C \subseteq S$ er begrænsede, så er ligning (2.12) opfyldt for $B = A \cup C$. Dermed kan Lemma A.1.6 anvendes og

$$\mathbb{E}\Big[\sum_{\xi \in X} h(\xi, X \setminus \xi)\Big] = \int_S \mathbb{E}\big[h(\xi, X)\big]\rho(\xi)d\xi$$

er opfyldt for $h(\xi, x) = \mathbb{1}[(\xi, x) \in F], \quad F \subseteq S \times N_{le}$. ■

Slivnyak-Meckes Sætning kan udvides til følgende sætning

Sætning 2.2.17 (Den udvidede Slivnyak-Meckes Sætning)
Hvis $X \sim$ Poisson(S, ρ), da gælder for alle $n \in \mathbb{N}$ og enhver funktion $h : S^n \times N_{le} \to [0, \infty)$ at

$$\begin{aligned} &\mathbb{E}\Big[\sum_{\xi_1,\dots,\xi_n \in X}^{\neq} h(\xi_1, \dots, \xi_n, X \setminus \{\xi_1, \dots, \xi_n\})\Big] \\ &= \int_S \cdots \int_S \mathbb{E}\big[h(\xi_1, \dots, \xi_n, X)\big] \prod_{i=1}^{n} \rho(\xi_i) d\xi_i \cdots d\xi_n \end{aligned} \tag{2.18}$$

(hvor $\neq$ angiver at de n punkter er parvist forskellige).

Bevis
For bevis se [14][s. 22]. ■

Summary Statistics 3

Dette kapitel introducerer *summary statistics* for rumlige punktprocesser X på $S = \mathbb{R}^d$. Summary statistics kan benyttes til modeltilpasning og validering. Desuden giver summary statistics et kvantiferbart mål for, hvordan punktmønstre er struktureret, hvilket opnås ved sammenligning med summary statistics for Poisson processen. Kapitlet indeholder ikke-parametriske estimater af summary statistics og væsentlige resultater.
Dette kapitel er baseret på [14].

3.1 Første- og anden ordens egenskaber

Første og anden ordens egenskaberne for de stokastiske tællevariable, $N(B)$ når $B \subseteq S$, er beskrevet ved intenstitetsmålet og det *anden ordens faktorielle momentmål*. Intensitetsmålet μ er i ligning (2.2) givet som $\mu(B) = \mathbb{E}[N(B)]$, og kaldes en første ordens egenskab. Intensitetsmålet for en punktproces X på S defineres generelt på denne måde.
I Definition 2.2.3 er intensitetsmålet μ givet som

$$\mu(B) = \int_B \rho(\xi) d\xi, \quad B \subseteq S.$$

Er det muligt at give μ på denne måde, kaldes ρ for intensitetsfunktionen for punktprocessen X på S. Hvis ρ er konstant, så er punktprocessen X homogen eller første ordens stationær med intensitet ρ, ellers siges X at være inhomogen. Når X siges at være stationær, menes at X er første ordens stationær.

Analogt med intensitetsmålet defineres det anden ordens faktorielle momentmål $\alpha^{(2)}$, som kaldes en anden ordens egenskab. For første ordens egenskaben μ undersøges antallet af punkter i $B \subseteq S$, men for anden ordens egenskaben $\alpha^{(2)}$ er det antallet af punktpar som undersøges.

Definition 3.1.1 (Anden ordens faktorielle momentmål)
Det anden ordens faktorielle momentmål $\alpha^{(2)}$ på $\mathbb{R}^d \times \mathbb{R}^d$ er givet ved

$$\alpha^{(2)}(C) = \mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} \mathbb{1}[(\xi, \eta) \in C] \Big], \quad C \subseteq \mathbb{R}^d \times \mathbb{R}^d.$$

Ud fra det anden ordens faktorielle momentmål er det muligt at definere *anden ordens produkttætheden*.

Definition 3.1.2 (Anden ordens produkttæthed)
Hvis det anden ordens faktorielle momentmål $\alpha^{(2)}$ kan skrives som

$$\alpha^{(2)}(C) = \iint \mathbb{1}[(\xi, \eta) \in C]\rho^{(2)}(\xi, \eta)d\xi d\eta, \quad C \subseteq \mathbb{R}^d \times \mathbb{R}^d,$$

hvor $\rho^{(2)}$ er en ikke-negativ funktion, så kaldes $\rho^{(2)}$ anden ordens produkttætheden.

Udtrykket $\rho^{(2)}(\xi, \eta)d\xi d\eta$ kan intuitivt ses som sandsynligheden for at observere et par af punkter fra X i både kuglen med centrum i ξ og volumen $d\xi$ og i kuglen med centrum i η og volumen $d\eta$.
For at kunne afgøre i hvilken grad en punktproces afviger fra en Poisson proces, så er det fordelagtigt at normalisere anden ordens produkttætheden $\rho^{(2)}(\xi, \eta)$, ved at dividere med $\rho(\xi)\rho(\eta)$.

Definition 3.1.3 (Parkorrelationsfunktionen)
Hvis både ρ og $\rho^{(2)}$ eksisterer, så er *parkorrelationsfunktionen* defineret ved

$$g(\xi, \eta) = \frac{\rho^{(2)}(\xi, \eta)}{\rho(\xi)\rho(\eta)},$$

med $g(\xi, \eta) = 0$, hvis $\rho(\xi)\rho(\eta) = 0$.

Intuitivt set er parkorrelationsfunktionen forholdet mellem den simultane sandsynlighed for at observere et par af punkter for punktprocessen og sandsynligheden for at observere det samme par af punkter uafhængigt af hinanden. Følgende proposition vil vise, at det for parkorrelationsfunktionen for en Poisson proces gælder at $g(\xi, \eta) = 1$. Dette stemmer overens med, at punkterne i en Poisson process optræder uafhængigt af hinanden.

Proposition 3.1.4
Lad $X \sim \text{Poisson}(S, \rho)$. Da gælder det for anden ordens produkttætheden $\rho^{(2)}$ at

$$\rho^{(2)}(\xi, \eta) = \rho(\xi)\rho(\eta). \tag{3.1}$$

Bevis
Fra Definition 3.1.2 vides at det anden ordens faktorielle momentmål for $C \subseteq \mathbb{R}^d \times \mathbb{R}^d$ kan skrives

$$\alpha^{(2)}(C) = \iint \mathbb{1}[(\xi, \eta) \in C]\rho^{(2)}(\xi, \eta)d\xi d\eta. \tag{3.2}$$

Ved at lade $C_1 \times C_2 = C$ hvor $C_1, C_2 \subseteq \mathbb{R}^d$ samt anvende Definition 3.1.1 fås af ligning (3.2) at

$$\begin{aligned}
\alpha^{(2)}(C) &= \int_{C_2}\int_{C_1} \rho^{(2)}(\xi,\eta)d\xi d\eta \\
&= \mathbb{E}\Big[\sum_{\xi,\eta\in X}^{\neq} \mathbb{1}[\xi\in C_1,\eta\in C_2]\Big] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{E}\Big[\mathbb{1}[\xi\in C_1,\eta\in C_2]\Big]\rho(\xi)\rho(\eta)d\xi d\eta. \qquad (3.3)
\end{aligned}$$

Ligning (3.3) fås ved at anvende den udvidede Slivnyak-Meckes Sætning 2.2.17, hvor h erstattes med indikatorfunktionen $\mathbb{1}[\xi \in C_1, \eta \in C_2]$. Middelværdien i ligning (3.3) giver altid 1, hvis (ξ,η) er i C. Dermed fås

$$\alpha^{(2)}(C) = \int_{C_2}\int_{C_1} \rho(\xi)\rho(\eta)d\xi d\eta,$$

hvormed det gælder at $\rho^{(2)}(\xi,\eta) = \rho(\xi)\rho(\eta)$ (μ-næsten overalt). ■

Parkorrelationen g kan ud fra Proposition 3.1.4 anvendes til at afgøre, om en punktproces er en Poisson proces. Hvis der for en punktproces X gælder at $g(\xi,\eta) > 1$, antyder dette at sandsynligheden for at observere et par af punkter i (ξ,η) er større end for en Poisson proces med samme intensitetsfunktion som X. Hvis $g(\xi,\eta) > 1$ vil punkterne, givet som udfaldene for punktprocessen, typisk forekomme tættere sammen end for Poisson processen. Dette kaldes for *clustering*. Omvendt antyder $g(\xi,\eta) < 1$, at sandsynligheden for at observere et par af punkter i (ξ,η) er mindre end for en Poisson Proces med samme intensitetsfunktion. Dermed vil $g(\xi,\eta) < 1$ typisk medføre, at punkterne givet som udfaldene for punktprocessen forekommer længere fra hinanden end for Poisson processen. Dette kaldes for *regularitet*.

Parkorrelationsfunktionen er invariant under translationer, hvis X er stationær, altså gælder $g(\xi,\eta) = g(\xi-\eta)$. Der findes tilfælde, hvor ρ er inhomogen, selvom g er translationsinvariant. Eksempelvis er dette opfyldt for en inhomogen Poisson proces hvor $g = 1$ per Proposition 3.1.4.

Følgende proposition minder om Slivnyak-Mecke Sætningerne 2.2.16 og 2.2.17, men disse gælder kun for Poisson processer. Følgende gælder generelt for punktprocesser.

Proposition 3.1.5 (Campells Sætning)
Antag at X har intensitetsfunktion ρ og anden ordens produkttæthed $\rho^{(2)}$. Så gælder for funktioner $h_1 : \mathbb{R}^d \to [0, \infty)$ og $h_2 : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ at

$$\mathbb{E}\Big[\sum_{\xi \in X} h_1(\xi)\Big] = \int h_1(\xi)\rho(\xi)d\xi \tag{3.4}$$

og

$$\mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} h_2(\xi,\eta)\Big] = \iint h_2(\xi,\eta)\rho^{(2)}(\xi,\eta)d\xi d\eta. \tag{3.5}$$

Bevis
Til dette bevis bruges teknikken bag standard beviset beskrevet i Afsnit 2.1.2. Funktionen $h_1(\xi)$ erstattes i ligning (3.4) med indikatorfunktionen $\mathbb{1}[\xi \in B]$ for passende $B \subseteq \mathbb{R}^d$. Fra ligning (2.2) og Definition 2.2.3 fås da

$$\mathbb{E}\Big[\sum_{\xi \in X} \mathbb{1}[\xi \in B]\Big] = \mathbb{E}[N(B)] = \mu(B) = \int \mathbb{1}[\xi \in B]\rho(\xi)d\xi.$$

På samme måde vises ligning (3.5), hvor $h_2(\xi,\eta)$ erstattes med indikatorfunktionen $\mathbb{1}[(\xi,\eta) \in C]$. Jævnfør Definition 3.1.1 og 3.1.2 fås

$$\mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} \mathbb{1}[(\xi,\eta) \in C]\Big] = \alpha^{(2)}(C) = \iint \mathbb{1}[(\xi,\eta) \in C]\rho^{(2)}(\xi,\eta)d\xi d\eta.$$

■

For en udtynding af punktprocessen X kan intensitetsfunktionen og anden ordens produkttætheden for X bruges til at bestemme intensitetsfunktionen og anden ordens produkttætheden for udtyndingen X_{thin}.

Proposition 3.1.6
Antag at X har intensitetsfunktion ρ og anden ordens produkttæthed $\rho^{(2)}$, og at X_{thin} er en uafhængig udtynding af X med retention sandsynligheder $p(\xi)$ hvor $\xi \in \mathbb{R}^d$. Så er intensitetsfunktionen og anden ordens produkttætheden for X_{thin} givet ved

$$\rho_{thin}(\xi) = p(\xi)\rho(\xi) \quad \text{og} \quad \rho^{(2)}_{thin}(\xi,\eta) = p(\xi)p(\eta)\rho^{(2)}(\xi,\eta).$$

Derudover gælder også at parkorrelationsfunktionen er invariant under uafhængig udtynding så $g = g_{thin}$.

Bevis
Per Definition 2.2.12 er $R(\xi) \sim \text{Uniform}([0,1]),\ \ \xi \in \mathbb{R}^d$ og gensidigt uafhængige samt uaf-

hængige af X. Ved anvendelse af ligning (2.2) og ved brug af notationen i Definition 2.2.12, så kan det for $B \subseteq \mathbb{R}^d$ skrives at

$$\begin{aligned}
\mu_{thin}(B) = \mathbb{E}[n(X_{thin} \cap B)] &= \mathbb{E}\big[\sum_{\xi \in X} \mathbb{1}[\xi \in B, R(\xi) \leq p(\xi)]\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[\sum_{\xi \in X} \mathbb{1}[\xi \in B, R(\xi) \leq p(\xi)]\big| X\big]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[\sum_{\xi \in X} \mathbb{1}[\xi \in B]\mathbb{1}[R(\xi) \leq p(\xi)]\big| X\big]\Big]. \qquad (3.6)
\end{aligned}$$

Loven om total forventning betinget på X er anvendt i ligning (3.6). Ved at betinge på X kan $\mathbb{1}[\xi \in B]$ trækkes udenfor den inderste middelværdi i ligning (3.6). Samtidig indses det at $\mathbb{E}\big[\mathbb{1}[R(\xi) \leq p(\xi)]\big] = p(\xi)$, og dette giver

$$\begin{aligned}
\mu_{thin}(B) &= \mathbb{E}\Big[\sum_{\xi \in X} \mathbb{1}[\xi \in B]\mathbb{E}\big[\mathbb{1}[R(\xi) \leq p(\xi)]\big]\Big] \\
&= \mathbb{E}\Big[\sum_{\xi \in X} p(\xi)\mathbb{1}[\xi \in B]\Big] \\
&= \int_B p(\xi)\rho(\xi) d\xi.
\end{aligned}$$

Her kommer sidste lighed fra Proposition 3.1.5 og dermed ses at $\rho_{thin}(\xi) = p(\xi)\rho(\xi)$. På samme måde kan $\rho^{(2)}_{thin}(\xi, \eta) = p(\xi)p(\eta)\rho^{(2)}(\xi, \eta)$ vises. Herfra følger at

$$g_{thin}(\xi, \eta) = \frac{\rho^{(2)}_{thin}(\xi, \eta)}{\rho_{thin}(\xi)\rho_{thin}(\eta)} = \frac{p(\xi)p(\eta)\rho^{(2)}(\xi, \eta)}{p(\xi)\rho(\xi)p(\eta)\rho(\eta)} = \frac{\rho^{(2)}(\xi, \eta)}{\rho(\xi)\rho(\eta)} = g(\xi, \eta).$$

■

3.1.1 Anden ordens reduceret momentmål

Det anden ordens faktorielle momentmål kan udtrykkes i form af intensitetsfunktionen og målet $\mathcal{K}$ på $\mathbb{R}^d$, som introduceres nu.

Definition 3.1.7 (Anden ordens reduceret momentmål)
Antag at X har intensitetsfunktion ρ og at målet

$$\mathcal{K}(B) = \frac{1}{|A|}\mathbb{E}\Big[\sum_{\xi, \eta \in X}^{\neq} \frac{\mathbb{1}[\xi \in A, \eta - \xi \in B]}{\rho(\xi)\rho(\eta)}\Big], \quad B \subseteq \mathbb{R}^d \qquad (3.7)$$

ikke afhænger af valget af $A \subseteq \mathbb{R}^d$ med $0 < |A| < \infty$. Hvis $\rho(\xi)\rho(\eta) = 0$ i ligning (3.7) sættes det tilsvarende led i summen lig 0. Så siges X at være *anden ordens intensitets-genvægtet stationær* og $\mathcal{K}$ kaldes det *anden ordens reducerede momentmål*.

Translationsinvarians af parkorrelationsfunktionen g for X medfører anden ordens intensitets-genvægtet stationaritet, hvilket følgende proposition viser.

Proposition 3.1.8
Lad X være en punktproces på $S = \mathbb{R}^d$. Givet at parkorrelationsfunktionen g for X er translationsinvariant, da gælder for det anden ordens reducerede momentmål at

$$\mathcal{K}(B) = \int_B g(\tilde{\eta}) d\tilde{\eta}, \quad B \subseteq \mathbb{R}^d \tag{3.8}$$

hvor $\tilde{\eta} = \xi - \eta$.

Bevis
Ved at anvende ligning (3.5) på ligning (3.7) og lade $\tilde{\eta} = \xi - \eta$ fås

$$\begin{aligned}
\mathcal{K}(B) &= \frac{1}{|A|}\mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} \frac{\mathbb{1}[\xi \in A, \eta - \xi \in B]}{\rho(\xi)\rho(\eta)}\Big] \\
&= \frac{1}{|A|}\iint \frac{\mathbb{1}[\xi \in A, \eta - \xi \in B]}{\rho(\xi)\rho(\eta)} \rho^{(2)}(\xi, \eta) d\xi d\eta \\
&= \frac{1}{|A|}\iint \mathbb{1}[\xi \in A, \eta - \xi \in B] g(\xi, \eta) d\xi d\eta \\
&= \frac{1}{|A|}\int_B \int_A g(\xi, \xi - \tilde{\eta}) d\xi d\tilde{\eta} \\
&= \frac{1}{|A|}\int_B \int_A g(\tilde{\eta}) d\xi d\tilde{\eta} \\
&= \frac{1}{|A|}\int_B g(\tilde{\eta}) \int_A 1 d\xi d\tilde{\eta} \\
&= \int_B g(\tilde{\eta}) d\tilde{\eta},
\end{aligned}$$

hvor translationsinvariansen af g er anvendt, ved at translatere med $-(\xi - \tilde{\eta})$.
Det anden ordens reducerede momentmål $\mathcal{K}$ afhænger ikke af A, hvormed X er anden ordens intensitetsgenvægtet stationær. ■

Generelt set er $\mathcal{K}$ ikke translationsinvariant, heller ikke hvis X er stationær. Hvis X er stationær, haves følgende proposition.

Proposition 3.1.9
Lad $\mathcal{K}(B)$ være det anden ordens reducerede momentmål for en homogen punktproces X. Da gælder for det anden ordens faktorielle momentmål $\alpha^{(2)}$ på $\mathbb{R}^d \times \mathbb{R}^d$ at

$$\alpha^{(2)}(B_1 \times B_2) = \rho^2 \int_{B_2} \mathcal{K}(B_1 - \eta) d\eta, \quad B_1, B_2 \subseteq \mathbb{R}^d.$$

Bevis
Når ρ er konstant, og g er invariant under translationer, fås for det anden ordens faktorielle

momentmål at

$$\begin{aligned}
\alpha^{(2)}(B_1 \times B_2) &= \iint \mathbb{1}[\xi \in B_1, \eta \in B_2]\rho^{(2)}(\xi,\eta)d\xi d\eta \\
&= \rho^2 \iint \frac{\mathbb{1}[\xi \in B_1, \eta \in B_2]\rho^{(2)}(\xi,\eta)}{\rho^2} d\xi d\eta \\
&= \rho^2 \iint \mathbb{1}[\xi \in B_1, \eta \in B_2] g(\xi,\eta) d\xi d\eta \\
&= \rho^2 \int_{B_2} \int_{B_1} g(\xi,\eta) d\xi d\eta \\
&= \rho^2 \int_{B_2} \mathcal{K}(B_1 - \eta) d\eta.
\end{aligned}$$

■

I det stationære tilfælde kan $\rho\mathcal{K}(B)$ tolkes som den betingede forventede værdi af antal punkter omkring Origo i B, givet at X har et punkt i Origo.

Et resultat relateret til summary statistics, er at $\mathcal{K}$ er invariant under uafhængig udtynding.

Proposition 3.1.10
Hvis X_{thin} er en uafhængig udtynding af en anden ordens intensitetsgenvægtet stationær punktproces X, så er X_{thin} anden ordens intensitetsgenvægtet stationær og $\mathcal{K}_{thin} = \mathcal{K}$.

Bevis
Det fås fra Definition 3.1.7, at

$$\begin{aligned}
\mathcal{K}_{thin}(B) &= \frac{1}{|A|}\mathbb{E}\Big[\sum_{\xi,\eta \in X_{thin}}^{\neq} \frac{\mathbb{1}[\xi \in A, \eta - \xi \in B]}{\rho_{thin}(\xi)\rho_{thin}(\eta)}\Big] \\
&= \frac{1}{|A|}\mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} \frac{\mathbb{1}[\xi \in A, \eta - \xi \in B]}{\rho_{thin}(\xi)\rho_{thin}(\eta)} p(\xi)p(\eta)\Big].
\end{aligned}$$

Andet lighedstegn følger af, at punkterne ξ og η indgår i X_{thin} med sandsynlighed henholdsvis $p(\xi)$ og $p(\eta)$. Det fås da af Proposition 3.1.6 at

$$\begin{aligned}
\mathcal{K}_{thin}(B) &= \frac{1}{|A|}\mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} \frac{\mathbb{1}[\xi \in A, \eta - \xi \in B]}{\rho(\xi)\rho(\eta)}\Big] \\
&= \mathcal{K}(B).
\end{aligned}$$

■

3.2 Summary statistics ud fra det andenordens reducerede momentmål

Det andenordens reducerede momentmål $\mathcal{K}$ benyttes her til at definere Ripleys K-*funktion*, ved at anvende $\mathcal{K}$ på 'test mængder' i form af d-dimensionelle kugler. Ud fra K-funktionen defineres yderligere en summary statistic, som er en transformation af K-funktionen.

Definition 3.2.1 (K- og L-funktionen)
K- og L-*funktionen* for en anden ordens intensitetsgenvægtet stationær punktproces X, er defineret ved

$$K(r) = \mathcal{K}(b(0,r)) \quad og \quad L(r) = \left(\frac{K(r)}{\omega_d}\right)^{1/d}$$

for $r > 0$.

I Definition 3.2.1 angiver $b(0,r)$ kuglen med radius r og centrum i 0, og $\omega_d = \pi^{d/2}/\Gamma(1+d/2)$ er volumen af den d-dimensionelle enhedskugle.
K er dermed defineret, ud fra værdien $\mathcal{K}$ antager på en kugle med centrum i 0 og radius r. Desuden ses at K- og L-funktionen afhænger af hinanden, idet L blot er en transformation af K. Det vises dog i Proposition 3.2.2, at L for en Poisson proces er lettere at genkende end K-funktionen. I praksis kan det derfor være lettere at anvende L-funktionen.
I det stationære tilfælde kan $\rho K(r)$ fortolkes som det forventede antal af punkter indenfor afstand r af et vilkårligt punkt i X.
Hvis $\mathcal{K}$ er invariant under rotationer, kan denne bestemmes ud fra K. Hvis parkorrelationsfunkionen g er invariant under rotationer, kan det ud fra ligning (3.8) vises at

$$K(r) = \sigma_d \int_0^r t^{d-1} g(t)dt \tag{3.9}$$

ifølge [14][s. 34].
Ligning (3.9) viser sammenhængen mellem K og g. Her er $\sigma_d = 2\pi^{d/2}/\Gamma(d/2)$ overfladearealet af den d-dimensionelle enhedskugle.

Følgende proposition giver et væsentligt resultat for K- og L-funktionen for en Poisson proces.

Proposition 3.2.2
Hvis X er en Poisson proces, samt andensordens intensitetsgenvægtet stationær, da gælder for K- og L-funktionen at

$$K(r) = \sigma_d \frac{1}{d} r^d, \;\; L(r) = r. \tag{3.10}$$

Bevis
Hvis X er en Poisson proces, samt andensordens intensitetsgenvægtet stationær, vides det at $g(\xi, \eta) = 1$, hvormed ligning (3.9) kan anvendes for at beregne $K(r)$ for X. Dermed fås

$$\begin{aligned} K(r) &= \sigma_d \int_0^r t^{d-1} g(t) dt \\ &= \sigma_d \int_0^r t^{d-1} dt \\ &= \sigma_d \Big[\frac{1}{d} t^d\Big]_{t=0}^r \\ &= \sigma_d \frac{1}{d} r^d. \end{aligned} \tag{3.11}$$

For at udregne $L(r)$ bemærkes det først, at for $d \in \mathbb{N}$ er

$$\Gamma(d/2) = \sqrt{\pi}\frac{(d-2)!!}{2^{\frac{d-1}{2}}},$$

hvilket giver en formel for udregning af gamma-funktionen [15]. Heraf fås at

$$\begin{aligned} \Gamma(1 + d/2) &= \Gamma(\frac{d+2}{2}) \\ &= \sqrt{\pi}\frac{d!!}{2^{\frac{d+1}{2}}}. \end{aligned}$$

Nu udregnes brøken

$$\begin{aligned} \frac{K(r)}{\omega_d} &= \frac{\sigma_d \frac{1}{d} r^d}{\pi^{d/2}/\Gamma(1+d/2)} \\ &= \frac{\frac{2\pi^{d/2}}{\sqrt{\pi}(d-2)!!(2^{\frac{d-1}{2}})^{-1}}}{\frac{\pi^{d/2}}{\sqrt{\pi}d!!(2^{\frac{d+1}{2}})^{-1}}} \frac{1}{d} r^d \\ &= \frac{2\pi^{d/2}\sqrt{\pi}d!!(2^{\frac{d+1}{2}})^{-1}}{\pi^{d/2}\sqrt{\pi}(d-2)!!(2^{\frac{d-1}{2}})^{-1}} \frac{1}{d} r^d \\ &= \frac{2d!!(2^{\frac{d+1}{2}})^{-1}}{(d-2)!!(2^{\frac{d-1}{2}})^{-1}} \frac{1}{d} r^d \\ &= \frac{d!!}{(d-2)!!} \frac{1}{d} r^d \\ &= r^d. \end{aligned} \tag{3.12}$$

Sidste lighed i ligning (3.12) følger af definitionen på dobbelt fakultet

$$n!! = \prod_{k=0}^{\lceil \frac{n}{2} \rceil - 1} (n - 2k)$$

for $n \in \mathbb{N}$ [3][s. 139]. Af definitionen på L-funktionen fås det dermed for en Poisson proces X at

$$L(r) = \left(\frac{K(r)}{\omega_d}\right)^{1/d} = r.$$

■

K-funktionen for en Poisson proces er dermed et d'te grads polynomium. Proposition 3.2.2 viser desuden, at det for en Poisson proces gælder, at $L(r)$ er identitetsafbildningen på $r > 0$, hvilket giver et væsentligt karakteristika for L-funktionen for en Poisson proces.
Hvis det for en punktproces gælder at $L(r) > r$, indikerer dette clustering for afstande mindre end r. Hvis det omvendt gælder at $L(r) < r$, indikerer dette regularitet for afstande mindre end r.

3.3 Summary statistics ud fra interne punktafstande

I dette afsnit indføres summary statistics, som afhænger af interne punktafstande i punktprocessen X, hvor X er antaget stationær. Herudover undersøges formen på disse summary statistics for en Poisson proces.
Først defineres *tomrumsfunktionen*.

Definition 3.3.1 (Tomrumsfunktionen)
Antag at X er stationær. Tomrumsfunktionen F er fordelingsfunktionen for afstanden fra Origo (eller et andet fast punkt i $\mathbb{R}^d$) til det nærmeste punkt i X

$$F(r) = P(X \cap b(0, r) \neq \emptyset) \tag{3.13}$$

for $r > 0$.

Bemærk at siden X er antaget stationær, vil F ikke være afhængig af valget af det faste punkt. Værdierne som F antager, er sandsynligheden for at der vil være mindst ét punkt i X indenfor afstand r af det faste punkt (Origo i definionen). Tomrumsfunktionen F er dermed voksende når r er voksende [4][s. 262].
En anden funktion, som også afhænger af afstanden mellem punkter, er den såkaldte *nærmeste nabo funktion* G.

Definition 3.3.2 (Nærmeste nabo funktionen)
Antag at X er stationær. Nærmeste nabo funktionen G er defineret ved

$$G(r) = \frac{1}{\rho|A|}\mathbb{E}\sum_{\xi \in X \cap A} \mathbb{1}[(X \setminus \xi) \cap b(\xi, r) \neq \emptyset] \tag{3.14}$$

for $r > 0$ og en vilkårlig mængde $A \subset \mathbb{R}^d$ hvor $0 < |A| < \infty$.

Nærmeste nabo funktionen G kan fortolkes som fordelingsfunktionen for afstanden fra et punkt ξ i X til det nærmeste nabopunkt i X. Idet X er stationær, afhænger G ikke af punktet ξ. Funktionen G antager sandsynligheder for at den mindste afstand mellem et fast punkt og et punkt i X, er lig r [4][s. 262].

Ud fra F- og G-funktionen er det muligt at definere en tredje summary statistic baseret på interne punktafstande.

Definition 3.3.3 (J-funktionen)
Antag at X er stationær. J*-funktionen* er da givet ved

$$J(r) = \frac{1 - G(r)}{1 - F(r)} \tag{3.15}$$

for $F(r) < 1$.

Den følgende proposition viser, at det for en Poisson proces gælder, at F- og G-funktionen er ens, hvilket giver et pænt udtryk for J-funktionen, når X er en Poisson proces.

Proposition 3.3.4
Lad X være en stationær Poisson proces. Da gælder for $\rho < \infty$ at

$$F(r) = G(r) = 1 - \exp(-\rho\omega_d r^d) \quad \text{og} \quad J(r) = 1.$$

Bevis
For en stationær Poisson proces med intensitet $\rho < \infty$ kan Tomrumsfunktionen F udregnes ved

$$\begin{aligned} F(r) &= P(X \cap b(0,r) \neq \emptyset) \\ &= 1 - P(X \cap b(0,r) = \emptyset) \\ &= 1 - v(b(0,r)). \end{aligned} \tag{3.16}$$

Sidste lighedstegn følger af at $P(X \cap b(0,r) = \emptyset)$ er void sandsynligheden for X. Idet $X \sim$ Poisson(S, ρ) kan Sætning 2.2.8 anvendes hvormed

$$\begin{aligned} F(r) &= 1 - \exp(-\mu(b(0,r)) \\ &= 1 - \exp\Big(-\int_{b(0,r)} \rho d\xi\Big) \\ &= 1 - \exp\Big(-\rho \int_{b(0,r)} 1 d\xi\Big) \\ &= 1 - \exp(-\rho\omega_d r^d). \end{aligned} \tag{3.17}$$

For G-funktionen anvendes Slivnyak-Meckes Sætning 2.2.16

$$\begin{aligned} G(r) &= \frac{1}{\rho|A|}\mathbb{E}\Big[\sum_{\xi\in X\cap A} \mathbb{1}[(X\setminus\xi)\cap b(\xi,r)\neq\emptyset]\Big] \\ &= \frac{1}{\rho|A|}\mathbb{E}\Big[\sum_{\xi\in X} \mathbb{1}[\xi\in A, (X\setminus\xi)\cap b(\xi,r)\neq\emptyset]\Big] \\ &= \frac{1}{\rho|A|}\int_A \mathbb{E}\big[\mathbb{1}[X\cap b(\xi,r)\neq\emptyset]\big]\rho d\xi \end{aligned} \tag{3.18}$$

I ligning (3.18) anvendes nu at $\mathbb{E}\big[\mathbb{1}[X\cap b(\xi,r)\neq\emptyset]\big] = P\big(X\cap b(\xi,r)\neq\emptyset\big)$, og det fås at

$$\begin{aligned} G(r) &= \frac{1}{|A|}\int_A P\big(X\cap b(\xi,r)\neq\emptyset\big)d\xi \\ &= \frac{1}{|A|}\int_A 1 - P\big(X\cap b(\xi,r)=\emptyset\big)d\xi. \end{aligned}$$

Ved at anvende ligning (3.16) og (3.17) fås da

$$\begin{aligned} G(r) &= \frac{1}{|A|}\int_A 1-\exp(-\rho\omega_d r^d)d\xi \\ &= \big(1-\exp(-\rho\omega_d r^d)\big)\frac{1}{|A|}\int_A 1 d\xi \\ &= 1-\exp(-\rho\omega_d r^d). \end{aligned}$$

Heraf følger det at $J(r)=1$. ■

Generelt indikerer $F(r) < G(r)$ at punktprocessen er clusteret. Bemærk at betingelsen $F(r) < G(r)$ er ækvivalent med $J(r) < 1$. Omvendt indikerer $F(r) > G(r)$ regularitet af punktprocessen. Betingelsen $F(r) > G(r)$ er ækvivalent med $J(r) > 1$. Dette gør sig gældende for 'tilpas små' værdier af $r > 0$ jævnfør [14][s. 36].

3.4 Ikke-parametrisk estimation af summary statistics

Da der nu er introduceret summary statistics for punktprocesser, gives herefter ikke-parametriske estimater af disse. Derudover introduceres punktvise envelopes, som giver en metode til at eftertjekke modeltilpasning.

I dette afsnit antages intensitetsfunktionen ρ at være konstant med $0 < \rho < \infty$, hvis X er stationær. Derudover så antages parkorrelationsfunktionen g og det andenordens reducerede momentmål $\mathcal{K}$ at eksistere.

I afsnittet begrænses der til tilfældet, hvor én realisation af punktprocessen $X_W = x$ er observeret på det afgrænsede område $W \subset \mathbb{R}^d$, hvor $|W| > 0$ er arealet af W. Det vil sige, at punktmønsteret x observeres på et afgrænset observationsvindue W.

3.4.1 Ikke-parametrisk estimation af intensitetsfunktioner

I det homogene tilfælde er et unbiased estimat af ρ givet som

$$\hat{\rho} = \frac{n(x)}{|W|}, \tag{3.19}$$

hvor $n(x)$ er antal punkter i x. At $\hat{\rho}$ er unbiased, ses ved at tage middelværdien herpå

$$\begin{aligned}\mathbb{E}[\hat{\rho}] &= \mathbb{E}\Big[\frac{n(x)}{|W|}\Big] = \frac{1}{|W|}\mathbb{E}[n(x)] = \frac{1}{|W|}\mu(W) \\ &= \frac{1}{|W|}\int_W \rho d\xi = \frac{1}{|W|}\rho\int_W 1 d\xi = \rho.\end{aligned}$$

For inhomogene punktprocesser vil begrebet *kerne* blive benyttet til at bestemme estimater.

Definition 3.4.1 (Kerne)
En funktion k kaldes en kerne, hvis k er en symmetrisk kontinuert tæthedsfunktion $k : \xi \to k(\xi)$.

Det betyder, at en kerne opfylder at $\int k(\xi)d\xi = 1$ og $k(\xi) = k(-\xi)$ [5][s. 26].
I det inhomogene tilfælde bruges et ikke-parametrisk kerne estimat af ρ

$$\hat{\rho}_b(\xi) = \sum_{\eta\in x} \frac{k_b(\xi - \eta)}{c_{W,b}(\eta)}, \tag{3.20}$$

med $\xi \in W$. Her er k_b en kerne, med *båndbredde* $b > 0$, givet som

$$k_b(\xi) = \frac{k(\xi/b)}{b^d},$$

hvor k er en tæthedsfunktion og båndbredden bestemmer glatheden af kernen. Derudover er $c_{W,b}$ en *kantkorrektionsfaktor* givet som

$$c_{W,b}(\eta) = \int_W k_b(\xi - \eta)d\xi.$$

Estimatet for intensitetsfunktionen kan desuden anvendes til at definere et estimat for intensitetsmålet $\int_W \hat{\rho}_b(\xi)d\xi$ ud fra definitionen af intensitetsmålet.

Lemma 3.4.2
Lad $\hat{\mu}(W) = \int_W \hat{\rho}_b(\xi)d\xi$ være et estimat af $\mu(W)$. Da gælder, at $\hat{\mu}(W)$ er et unbiased estimat.

Bevis
For at vise at estimatet $\hat{\mu}(W)$ er unbiased, tages middelværdien af estimatet. Ved at bruge udtrykket for $\hat{\rho}_b(\xi)$ i ligning (3.20) fås at

$$\begin{aligned}\mathbb{E}[\hat{\mu}(W)] = \mathbb{E}\Big[\int_W \sum_{\eta \in X_W} \frac{k_b(\xi - \eta)}{c_{W,b}(\eta)} d\xi\Big] &= \int_W \mathbb{E}\Big[\sum_{\eta \in X_W} \frac{k_b(\xi - \eta)}{c_{W,b}(\eta)}\Big] d\xi \\ &= \int_W \int_W \left(\frac{k_b(\xi - \eta)}{c_{W,b}(\eta)}\right) \rho(\eta) d\eta d\xi \\ &= \int_W \rho(\eta) d\eta = \mu(W),\end{aligned}$$

hvor ligning (3.4) fra Proposition 3.1.5 er brugt ved tredje lighedstegn. ■

3.4.2 Ikke-parametrisk estimation af *K*- og *L*-funktionen

For ikke-parametrisk estimation af $\mathcal{K}(B)$ er følgende lemma fordelagtigt at indføre.

Lad $W_\xi = \{\eta + \xi : \eta \in W\}$ være translationen af W med $\xi \in \mathbb{R}^d$.

Lemma 3.4.3
Antag at $|W \cap W_\xi| > 0$ for alle $\xi \in B$ hvor $B \subseteq S$, og at X er andenordens intensitetsgenvægtet stationær. Så gælder, at

$$\sum_{\xi, \eta \in x}^{\neq} \frac{\mathbb{1}[\eta - \xi \in B]}{\rho(\xi)\rho(\eta)|W \cap W_{\eta - \xi}|} \tag{3.21}$$

er et unbiased estimat af $\mathcal{K}(B)$.

Bevis
Antag, for simpelhedens skyld, at X har en translationsinvariant parkorrelationsfunktion g, sådan at ligning (3.5) kan bruges med $\rho^{(2)}(\xi, \eta) = \rho(\eta)\rho(\xi)g(\eta - \xi)$. Ved at lade $h(\xi, \eta) = \mathbb{1}[\eta - \xi \in B]/(\rho(\xi)\rho(\eta)|W \cap W_{\eta-\xi}|)$ i ligning (3.5) og $\tilde{\eta} = \eta - \xi$ så fås

$$\begin{aligned}\mathbb{E}\Big[\sum_{\xi, \eta \in X_W}^{\neq} \frac{\mathbb{1}[\eta - \xi \in B]}{\rho(\xi)\rho(\eta)|W \cap W_{\eta - \xi}|}\Big] &= \iint \frac{\mathbb{1}[\xi \in W, \eta \in W, \eta - \xi \in B]}{\rho(\xi)\rho(\eta)|W \cap W_{\eta - \xi}|} \rho^{(2)}(\xi, \eta) d\eta d\xi \\ &= \iint \frac{\mathbb{1}[\xi \in W, \eta \in W, \eta - \xi \in B]}{|W \cap W_{\eta - \xi}|} g(\eta - \xi) d\eta d\xi. \end{aligned} \tag{3.22}$$

Det ses, at $\eta \in W$ er ækvivalent med $\xi \in W_{-\tilde{\eta}}$, da translationen med $-\tilde{\eta}$ giver at

$\eta - \tilde{\eta} = \eta - (\eta - \xi) = \xi$. Dermed fås at

$$\begin{aligned}
\mathbb{E}\Big[\sum_{\xi,\eta\in X_W}^{\neq} \frac{\mathbb{1}[\eta-\xi\in B]}{\rho(\xi)\rho(\eta)|W\cap W_{\eta-\xi}|}\Big] &= \iint \frac{\mathbb{1}[\xi\in W, \xi\in W_{-\tilde{\eta}}, \tilde{\eta}\in B]}{|W\cap W_{\tilde{\eta}}|} g(\tilde{\eta}) d\xi d\tilde{\eta} \\
&= \iint \frac{\mathbb{1}[\xi\in W\cap W_{-\tilde{\eta}}]\mathbb{1}[\tilde{\eta}\in B]}{|W\cap W_{\tilde{\eta}}|} g(\tilde{\eta}) d\xi d\tilde{\eta} \\
&= \int \frac{\mathbb{1}[\tilde{\eta}\in B]}{|W\cap W_{\tilde{\eta}}|} g(\tilde{\eta}) \int_{W\cap W_{-\tilde{\eta}}} 1 d\xi d\tilde{\eta} \\
&= \int_B g(\tilde{\eta}) \frac{1}{|W\cap W_{\tilde{\eta}}|} |W\cap W_{-\tilde{\eta}}| d\tilde{\eta} \\
&= \mathcal{K}(B),
\end{aligned}$$

ved brug af ligning (3.8) og at $|W\cap W_{\tilde{\eta}}| = |W\cap W_{-\tilde{\eta}}|$ i sidste ligning. ■

Lemma 3.4.3 giver et unbiased estimat af $\mathcal{K}(B)$ givet at intensiteten ρ kendes. I praksis er ρ ikke kendt, så $\rho(\xi)\rho(\eta)$ i (3.21) skal erstattes med et estimat $\widehat{\rho(\xi)\rho(\eta)}$. Det sammensatte estimat bliver da

$$\hat{\mathcal{K}}(B) = \sum_{\xi,\eta\in x}^{\neq} \frac{\mathbb{1}[\eta-\xi\in B]}{\widehat{\rho(\xi)\rho(\eta)}|W\cap W_{\eta-\xi}|}, \tag{3.23}$$

som er biased. Dette kan ses ud fra beviset for Lemma 3.4.3 i ligning (3.22), hvor $\rho(\xi)\rho(\eta)/\rho^{(2)}(\xi,\eta)$ bliver til g. Hvis estimatet $\widehat{\rho(\xi)\rho(\eta)}$ indsættes i ligning (3.22), giver det ikke g.

Estimatet af L-funktionen, som fremkommer gennem transformation af estimatet af K-funktionen, er generelt også biased. Faktisk er mange estimatorer indenfor rumlig statistik biased. Ofte er de til gengæld *ratio-unbiased*.

Et estimat for θ siges at være ratio-unbiased, hvis estimatoren $\hat{\theta} = Y/Z$ opfylder at $\mathbb{E}[Y]/\mathbb{E}[Z] = \theta$.

Se eksempelvis på det homogene tilfælde. Hvis $\widehat{\rho(\xi)\rho(\eta)} = \widehat{\rho^2}$ er unbiased, så er ligning (3.23) ratio-unbiased jævnfør [14][s. 39].

3.4.3 Kantkorrektion

Vægten $1/|W\cap W_{\xi-\eta}|$, som indgår i ligningerne (3.21) og (3.23), kaldes en kantkorrektionsfaktor, hvilke der findes flere eksempler på. En anden kantkorrektion, som kaldes grænse kantkorrektion, er givet for $r > 0$ som

$$W' = \{\xi\in W : b(\xi, r)\subseteq W\}. \tag{3.24}$$

I ligning (3.24) er W' mængden af punkter i W, hvis afstand til randen af W er større end r. Der gælder at

$$\sum_{\xi \in x, \eta \in x \cap W'}^{\neq} \frac{\mathbb{1}[\eta - \xi \in B]}{\rho(\xi)\rho(\eta)} \tag{3.25}$$

er et unbiased estimat af $\mathcal{K}(B)$. At udtrykket i (3.25) er et unbiased estimat af $\mathcal{K}(B)$ ses ved at tage middelværdien heraf, hvorved

$$\mathbb{E}\Big[\sum_{\xi \in x, \eta \in x \cap W'}^{\neq} \frac{\mathbb{1}[\eta - \xi \in B]}{\rho(\xi)\rho(\eta)}\Big] = \mathbb{E}\Big[\sum_{\xi, \eta \in x}^{\neq} \frac{\mathbb{1}[\eta \in W', \eta - \xi \in B]}{\rho(\xi)\rho(\eta)}\Big]. \tag{3.26}$$

Det ses, at sidste udtryk i ligning (3.26) stemmer overens med ligning (3.7) i Definition 3.1.7. Estimatet i (3.25) indeholder mindre information end summen $\hat{\mathcal{K}}(B)$ i ligning (3.23), idet nogle punktpar fjernes gennem grænse kantkorrektionen W'. Dette gør estimatet uegnet for mindre datasæt.

Ved anvendelse af grænse kantkorrektionen W' fås estimatet for $K(r)$ som

$$\hat{K}_{bord}(r) = \frac{\sum_{\xi \in x \cap W'} t(\xi, r, x)}{\hat{\rho} n(x \cap W')},$$

hvor $t(\xi, r, x) = \sum_{\eta \in x} \mathbb{1}[0 < \|\xi - \eta\| \leq r]$. For store datasæt er grænse kantkorrektionen konsistent og approksimativt unbiased under regularitetsbetingelser [4][s. 215].

3.4.4 Ikke-parametrisk estimation af parkorrelationsfunktionen

For at estimere g antages det, at $g(\xi, \eta) = g(\|\xi - \eta\|)$ er isotropisk. Ud fra ligning (3.8) kan g udtrykkes som den afledede af $\mathcal{K}(B)$, men dette kræver, at $\mathcal{K}(B)$ er differentiabel, eller rettere at $\hat{\mathcal{K}}(B)$ er differentiabel. Dog er $\hat{\mathcal{K}}(B)$ typisk en trin-funktion, som besværliggør at bestemme $\hat{\mathcal{K}}'(B)$ ud fra $\hat{\mathcal{K}}(B)$. Et muligt estimat for parkorrelationsfunktionen er i stedet givet ved et kantkorrigeret kerne estimat

$$\hat{g}(r) = \frac{1}{\sigma_d r^{d-1} |W|} \sum_{\xi, \eta \in x}^{\neq} \frac{k_b(r - \|\eta - \xi\|)}{\widehat{\rho(\xi)\rho(\eta)} |W \cap W_{\eta - \xi}|}. \tag{3.27}$$

I ligning (3.27) er $k_b = k(u/b)/b, \ u \in \mathbb{R}$ en kerne k med båndbredde $b > 0$, og $\sigma_d = 2\pi^{d/2}/\Gamma(d/2)$ er overfladearealet af den d-dimensionelle enhedskugle. Desuden er det antaget at $|W \cap W_\xi| > 0$ for $\|\xi\| \leq r$.

3.4.5 Ikke-parametrisk estimation af *F*-, *G*- og *J*-funktionen

Ved anvendelse af minus-sampling kan reducerede-sample estimatorer for F og G opstilles. Den korteste afstand fra $\xi \in \mathbb{R}^d$ til en mængde $B \subset \mathbb{R}^d$ skrives

$$d(\xi, B) = \inf\{\|\xi - \eta\| : \eta \in B\}.$$

Vælg, uafhængigt af punktprocessen X, et endeligt regulært gitter af punkter $I \subset W$. Regulær skal forstås på den måde, at der er en fast afstand mellem alle punkter i gitteret. Lad da $\#I_r$ betegne kardinaliteten af $I_r = I \cap W'$ for $r > 0$, hvor W' er givet som i ligning (3.24). Da gives det reducerede-sample estimat for F som

$$\hat{F}_{RS}(r) = \sum_{\xi \in I_r} \frac{\mathbb{1}[d(\xi, x) \leq r]}{\#I_r}, \tag{3.28}$$

hvor $\#I_r > 0$.
Det reducerede-sample estimat for G gives som

$$\hat{G}_{RS}(r) = \sum_{\xi \in x \cap W'} \frac{\mathbb{1}[d(\xi, x \setminus \xi) \leq r]}{\hat{\rho}|W'|} \tag{3.29}$$

for $|W'| > 0$. Det gælder at $\hat{F}_{RS}(r)$ er et unbiased estimat og at $\hat{G}_{RS}(r)$ er et ratio-unbiased estimat ifølge [14][s. 46].
Når estimater for F og G er givet, er det ud fra Definition 3.3.3 muligt at få følgende estimat for J

$$\hat{J}(r) = \frac{1 - \hat{G}(r)}{1 - \hat{F}(r)}.$$

3.4.6 Envelopes for summary statistics

I dette speciale anvendes *envelopes* til at undersøge hvor godt modellerne, som bliver introduceret i Kapitel 4, passer på data. Grundlæggende benyttes envelopes som en form for hypotesetest, hvor en punktproces simuleres under antagelsen af en nulhypotese. Ved hjælp af estimater for summary statistics kan de pågældende envelopes opnås ud fra førnævnte simulationer.

I det følgende betragtes L-funktionen, men de pågældende teknikker i afsnittet kan anvendes på hver af de definerede summary statitistics fra Afsnit 3.1, 3.2 og 3.3. Ud fra en simpel hypotese H, kan konfidensintervaller og andre fordelingsmæssige karakteristika for det ikke-parametriske estimat $\hat{L}(r)$ bestemmes ud fra simulation under H.

For en givet afstand $r > 0$ lad $T_0(r) = T(X, r)$ notere $\hat{L}(r)$ opnået fra en punktproces X indenfor observationsvinduet W. Lad $T_1(r) = T(X_1, r), \ldots, T_n(r) = T(X_n, r)$ være opnået fra uafhængige identisk fordelte simulationer $X_1, \ldots, X_n$ under H. Ud fra den empiriske fordeling af $T_1(r), \ldots, T_n(r)$ kan enhver fraktil for fordelingen af $T_0(r)$ under H bestemmes. Bemærk at selvom $T_1(r), \ldots, T_n(r)$ er uafhængige, er de stokastiske vektorer $\big(T_1(r), \ldots, T_n(r)\big)$ afhængige for forskellige værdier af $r > 0$. Dette er værd at bemærke ved sammenligning af resultater for forskellige r-værdier. Hvis beregningen af $T_i(r), i = 1, \ldots, n$ er tidskrævende for store værdier af n, kan følgende envelopes anvendes. Antallet af simulationer holdes her forholdvis småt - for eksempel $n = 39$. Lad

$$T_{\min}(r) = \min\{T_1(r), \ldots, T_n(r)\} \text{ og } T_{\max}(r) = \max\{T_1(r), \ldots, T_n(r)\}$$

være henholdsvis minimum og maksimum af de n punktprocesser. Under hypotesen H gælder

$$P(T_0(r) < T_{\min}(r)) = P(T_0(r) > T_{\max}(r)) \leq 1/(n+1)$$

på baggrund af symmetrien af de $n + 1$ uafhængige identiske fordelte $T_0(r), T_1(r), \ldots, T_n(r)$, med lighed hvis $T_0(r), T_1(r), \ldots, T_n(r)$ er forskellige. Grænserne $T_{\min}(r)$, $T_{\max}(r)$ kaldes henholdsvis $100/(n+1)\%$-nedre og $100n/(n+1)\%$-øvre envelopen ved afstand $r > 0$. Eksempelvis kaldes disse for $n = 19$ henholdsvis 5%- nedre og 95%-øvre envelopen.

Envelopes kan fortolkes som en statistisk signifikanstest på følgende måde. Antag X er en givet mængde data, som er stokastisk, og har samme antal punkter som $X_1, \ldots, X_n$. Under H vil X og de simulerede $X_1, \ldots, X_n$ være statistisk ækvivalente. Dermed vil $T_0(r)$ og $T_1(r), \ldots, T_n(r)$ alle være statistisk ækvivalente for et fast r. Givet at $n = 19$ vil der være 5% sandsynlighed for at $T_0(r)$ vil være den mindste værdi, og 5% sandsynlighed for at $T_0(r)$ vil være den største værdi ud af $T_0(r), T_1(r), \ldots, T_{19}(r)$. Dermed vil der være 10% sandsynlighed for at $T_0(r)$, vil ligge udenfor de simulerede værdier $T_1(r), \ldots, T_{19}(r)$ for hvert fast r. Hvis dette forekommer, kan nulhypotesen afvises med signifikans niveau 0.10. Dog ønskes som regel et signifikans niveau på 0.05 (39 simulationer) eller 0.01 (199 simulationer) [4][s. 371].

Envelopes af ovenstående type kaldes også punktvise envelopes, idet maksimum og minimumsværdierne udregnes separat for hvert r. Dette betyder, at envelopes kun kan fortolkes som en statistisk signifikanstest, hvis afstanden r fastsættes på forhånd. Et plot af den empirisk bestemte funktion $T_0(r)$ sammen med tilhørende envelopes opnået ud fra de simulerede $T_1(r), \ldots, T_n(r)$ kan dermed ikke anvendes til at afvise nulhypotesen (med $2/(n + 1)$ signifikansniveau), da plottet ikke forudsætter en bestemt værdi for r. I stedet kan plottet anvendes til at bestemme, hvad resultatet af testen ville have været givet en bestemt værdi af r. Envelopes som statistisk signifikanstest virker dermed kun, hvis valget af r tages på forhånd [4][s.234-235].

Cox processer 4

Klassen af Poisson processer er ofte for simple modeller til at kunne beskrive virkelige data. I stedet kan Poisson processerne bruges til at konstruere andre mere fleksible modelklasser, som bedre beskriver virkelige data. En af disse klasser er *Cox processerne*, som introduceres i dette kapitel. Cox processer er modeller for clusterede punktmønstre, hvor clusteringen er forårsaget af en underliggende stokastisk proces. Et specialtilfælde af Cox processen fås ved at lade intensiteten være givet som eksponentialfunktionen taget på et Gaussisk stokastisk felt. Disse processer kaldes *log Gaussiske Cox processer*. I dette speciale anvendes de log Gaussiske Cox processer til at modellere hussalg i Aalborg Kommune, som ses i Kapitel 6.
Kapitel 4 er baseret på [14].

4.1 Definitioner og egenskaber for Cox processer

En Cox proces er en naturlig udvidelse af en Poisson proces, som opnås ved at betragte intensitets funktionen for en Poisson proces, som en realisation af et stokastisk felt.
I overensstemmelse med Poisson processerne i Afsnit 2.2 er $S \subseteq \mathbb{R}^d$ i dette afsnit.
Først defineres et stokastisk felt.

Definition 4.1.1 (Stokastisk felt)
Lad et sandsynlighedsrum $(\Omega, \mathcal{F}, P)$ og en parametermængde T være givet. Et stokastisk felt er da enten en endelig funktion, eller en funktion med reelle værdier $Z(\omega, \mathbf{t})$, som for et hvert fast $\mathbf{t}$, er en målelig funktion med variable $\omega \in \Omega$ [1][s. 3].

At Z er et stokastisk felt, betyder dermed, at $Z(\xi)$ er en stokastisk variabel for alle $\xi \in S$. Yderligere information relateret til stokastiske felter kan findes i Appendiks B.1.
Da stokastiske felter nu er indført, er det muligt at definere Cox processer.

Definition 4.1.2 (Cox proces)
Antag at $Z = \{Z(\xi) : \xi \in S\}$ er et ikke-negativt stokastisk felt, sådan at der gælder med sandsynlighed et, at $\xi \to Z(\xi)$ er en lokal integrabel funktion. Hvis den betingede fordeling af X givet Z, er en Poisson proces på S med intensitetsfunktion Z, så siges X at være en Cox proces drevet af Z.

Hvis $\rho(\xi) = \mathbb{E}[Z(\xi)]$ eksisterer, og er lokal integrabel, så er $Z(\xi)$, med sandsynlighed et, en lokal integrabel funktion jævnfør [14][s. 58].

Intensitetsmålet for Poisson processen $X|Z$ er

$$M(B) = \int_B Z(\xi)d\xi, \;\; B \subseteq S.$$

I tilfældet hvor Z er deterministisk, bliver X en Poisson proces med intensitetsfunktion $\rho = Z$.

I det følgende vises det, at visse fordelingsmæssige egenskaber for Cox processen følger af at betinge på Z, hvorved egenskaberne for Poisson processen $X|Z$ kan benyttes. Først undersøges intensitetsmålet og herigennem intensistetsfunktionen af en Cox proces.

Proposition 4.1.3

Lad X være en Cox proces drevet af Z på S. Da er intentsitetsmålet for X givet ved

$$\mu(B) = \int_B \mathbb{E}[Z(\xi)]d\xi$$

for $B \subseteq S$. Intentsitetsfunktionen er givet

$$\rho(\xi) = \mathbb{E}[Z(\xi)].$$

Bevis

For at bestemme $\mu(B)$ for $B \subseteq S$ anvendes definitionen på intentsitetsmålet i Definition 2.2.3. Loven om total forventning giver da

$$\begin{aligned}\mu(B) &= \mathbb{E}[N(X_B)] = \mathbb{E}\big[\mathbb{E}[N(X_B)|Z]\big] \\ &= \mathbb{E}\Big[\int_B Z(\xi)d\xi\Big] = \int_B \mathbb{E}[Z(\xi)]d\xi,\end{aligned}$$

hvor tredje lighed følger af Definition 4.1.2. ■

Med samme bevismetode er det muligt at udlede void sandsynligheden for en Cox proces. Dette giver følgende proposition.

Proposition 4.1.4

For en Cox proces X er void sandsynligheden $v(B)$, $B \subseteq S$ givet som

$$v(B) = \mathbb{E}\Big[\exp\Big(-\int_B Z(\xi)d\xi\Big)\Big].$$

Bevis

Ud fra definitionen af void sandsynligheder og loven om total forventning fås

$$\begin{aligned}v(B) &= P(X \cap B = \emptyset) = \mathbb{E}\big[\mathbb{1}[X_B = \emptyset]\big] \\ &= \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}[X_B = \emptyset]|Z\big]\Big].\end{aligned}$$

Idet at

$$\mathbb{E}\big[\mathbb{1}[X_B = \emptyset]|Z\big] = 1 \cdot P(X_B = \emptyset|Z) + 0 \cdot P(X_B = \emptyset|Z) = P(X_B = \emptyset|Z)$$

anvendes nu at $X|Z$ er en Poisson proces og af Sætning 2.2.8 fås dermed

$$\begin{aligned} v(B) &= \mathbb{E}\big[P(X_B = \emptyset|Z)\big] = \mathbb{E}\big[\exp(-\mu(B))\big] \\ &= \mathbb{E}\Big[\exp\Big(-\int_B Z(\xi)d\xi\Big)\Big]. \end{aligned}$$

■

Følgende proposition giver udtryk for produkttætheden og for parkorrelationen for en Cox proces.

Proposition 4.1.5
For en Cox proces X er produktætheden givet som

$$\rho^{(2)}(\xi,\eta) = \mathbb{E}\big[Z(\xi)Z(\eta)\big].$$

Parkorrelationen for X er givet ved

$$g(\xi,\eta) = \frac{\mathbb{E}[Z(\xi)Z(\eta)]}{\mathbb{E}[Z(\xi)]\mathbb{E}[Z(\eta)]}.$$

Bevis
Ved at betragte det andenordens faktorielle momentmål for $C \subseteq \mathbb{R}^d \times \mathbb{R}^d$ fås af loven om total forventning at

$$\begin{aligned} \alpha^{(2)}(C) &= \mathbb{E}\Big[\sum_{\xi,\eta\in X}^{\neq} \mathbb{1}[(\xi,\eta)\in C]\Big] \\ &= \mathbb{E}\Bigg[\mathbb{E}\Big[\sum_{\xi,\eta\in X}^{\neq} \mathbb{1}[(\xi,\eta)\in C]|Z\Big]\Bigg]. \end{aligned}$$

Ved at anvende Proposition 3.1.5 ligning (3.5) og lade $\rho^{(2)}_{X|Z}(\xi,\eta)$ være andenordens produkttætheden for Poisson processen $X|Z$ fås at

$$\begin{aligned} \alpha^{(2)}(C) &= \mathbb{E}\Big[\iint \mathbb{1}[(\xi,\eta)\in C]\rho^{(2)}_{X|Z}(\xi,\eta)d\xi d\eta\Big] \\ &= \mathbb{E}\Big[\iint \mathbb{1}[(\xi,\eta)\in C]Z(\xi)Z(\eta)d\xi d\eta\Big] \\ &= \iint_C \mathbb{E}\big[Z(\xi)Z(\eta)\big]d\xi d\eta. \end{aligned}$$

Proposition 3.1.4 er anvendt i andet lighedstegn.

Ud fra Definition 3.1.2 fås da at $\rho^{(2)}(\xi,\eta) = \mathbb{E}[Z(\xi)Z(\eta)]$ for Cox processen X. Det følger da af Proposition 4.1.3, at parkorrelationen er givet

$$g(\xi,\eta) = \frac{\mathbb{E}[Z(\xi)Z(\eta)]}{\mathbb{E}[Z(\xi)]\mathbb{E}[Z(\eta)]}.$$

■

Bemærk at $\rho^{(2)}(\xi,\eta) = \mathbb{E}[Z(\xi)Z(\eta)]$ kun er opfyldt μ-næsten overalt, det vil sige udenfor en nulmængde.
Hvis g er translationsinvariant fås andenordens intentsitetsgenvægtet stationaritet af Proposition 3.1.8, som også giver en formel for udregning af både $\mathcal{K}$ og K (ligning (3.9)).

Antallet af punkter i $A \subseteq S$ skrives som tidligere $N(A)$. Følgende proposition omtaler, at variansen af antallet af punkter i $A \subseteq S$ er større end (eller lig med) det forventede antal punkter for en Cox proces.

Proposition 4.1.6
For en Cox proces X gælder

$$\text{Var}\big(N(A)\big) \geq \mathbb{E}\big[N(A)\big]$$

for $A \subseteq S$.

Bevis
Først anvendes loven om total varians [2][s. 193].

$$\begin{aligned}\text{Var}\big(N(A)\big) &= \text{Var}\big(\mathbb{E}[N(A)|Z]\big) + \mathbb{E}\big[\text{Var}(N(A)|Z)\big] \\ &\geq \mathbb{E}\big[\text{Var}(N(A)|Z)\big]\end{aligned}$$

Ved at bemærke at $X|Z$ er en Poisson proces må $N(A)|Z$ være Poisson fordelt. Dermed er $N(A)|Z \sim \text{po}\big(\int_A Z(\xi)d\xi\big)$, hvorom det gælder at $\text{Var}(N(A)|Z) = \mathbb{E}[N(A)|Z]$. Da fås

$$\begin{aligned}\text{Var}\big(N(A)\big) &\geq \mathbb{E}\Big[\int_A Z(\xi)d\xi\Big] \\ &= \int_A \mathbb{E}\big[Z(\xi)\big]d\xi \\ &= \mathbb{E}\big[N(A)\big].\end{aligned}$$

■

Bemærk at for en Poisson proces er $\text{Var}(N(A)) = \mathbb{E}[N(A)]$, hvormed $N(A)$ er Poisson fordelt. Cox processen udviser altså, hvad der kaldes *overspredning* sammenlignet med Poisson processen. Overspredning skal forstås på den måde, at variansen af antallet af punkter typisk vil være større end middelværdien for antallet af punkter i en Cox proces. Dette betyder, at antallet af punkter typisk vil variere mere i en Cox proces end en Poisson proces. Som følge heraf vil punkterne i en Cox proces som regel optræde mere clusterede end i en Poisson process.

Følgende proposition giver et udtryk for kovariansen af antal punkter i $A, B \subseteq S$, når X er en Cox proces.

Proposition 4.1.7
For en Cox proces X gælder at

$$\text{Cov}\big(N(A), N(B)\big) = \int_A \int_B \text{Cov}\big(Z(\xi), Z(\eta)\big) d\xi d\eta + \mu(A \cap B)$$

hvor $A, B \subseteq S$.

Bevis
Først anvendes følgende formel for kovarians [2][s. 197]

$$\begin{aligned}
\text{Cov}\big(N(A), N(B)\big) &= \mathbb{E}\big[N(A)N(B)\big] - \mathbb{E}\big[N(A)\big]\mathbb{E}\big[N(B)\big] \\
&= \mathbb{E}\Big[\sum_{\xi,\eta \in X} \mathbb{1}[\xi \in A, \eta \in B]\Big] - \mu(A)\mu(B) \\
&= \mathbb{E}\Big[\sum_{\xi,\eta \in X}^{\neq} \mathbb{1}[\xi \in A, \eta \in B]\Big] + \mathbb{E}\Big[\sum_{\xi=\eta \in X} \mathbb{1}[\xi \in A, \eta \in B]\Big] - \mu(A)\mu(B) \\
&= \alpha^{(2)}(A \times B) + \mathbb{E}[N(A \cap B)] - \mu(A)\mu(B).
\end{aligned}$$

Sidste lighed følger af Definition 3.1.1. Ved at anvende Definition 3.1.2 sammen med Proposition 4.1.5 og Proposition 4.1.3 fås at

$$\begin{aligned}
\text{Cov}\big(N(A), N(B)\big) &= \int_B \int_A \mathbb{E}\big[Z(\xi)Z(\eta)\big] d\xi d\eta + \mu(A \cap B) - \int_A \mathbb{E}\big[Z(\xi)\big] d\xi \int_B \mathbb{E}\big[Z(\eta)\big] d\eta \\
&= \int_B \int_A \mathbb{E}\big[Z(\xi)Z(\eta)\big] d\xi d\eta - \int_B \int_A \mathbb{E}\big[Z(\xi)\big] d\xi \mathbb{E}\big[Z(\eta)\big] d\eta + \mu(A \cap B) \\
&= \int_B \int_A \mathbb{E}\big[Z(\xi)Z(\eta)\big] d\xi d\eta - \int_B \int_A \mathbb{E}\big[Z(\xi)\big]\mathbb{E}\big[Z(\eta)\big] d\xi d\eta + \mu(A \cap B) \\
&= \int_B \int_A \mathbb{E}\big[Z(\xi)Z(\eta)\big] - \mathbb{E}\big[Z(\xi)\big]\mathbb{E}\big[Z(\eta)\big] d\xi d\eta + \mu(A \cap B) \\
&= \int_B \int_A \text{Cov}\big(Z(\xi), Z(\eta)\big) d\xi d\eta + \mu(A \cap B).
\end{aligned}$$

■

Kovariansen af $N(A)$ og $N(B)$ kan dermed udtrykkes ud fra kovariansen af $Z(\xi)$ og $Z(\eta)$, når X er en Cox proces på S.

4.2 Log Gaussiske Cox processer

I dette afsnit introduceres log Gaussiske Cox processer, der er et specialtilfælde af Cox processen X. Cox processerne blev indført i Afsnit 4.1, hvor X blev drevet af et stokastisk felt Z. Ved at lade $Y = \log(Z)$ være et Gaussisk stokastisk felt, fås en log Gaussisk Cox proces. Først defineres et Gaussisk stokastisk felt.

Definition 4.2.1 (Gaussisk stokastisk felt)
Et Gaussisk stokastisk felt Y er et stokastisk felt, som opfylder, at dets endelig dimensionelle fordelinger $F_{\mathbf{t}_1}, \dots, F_{\mathbf{t}_k}$ følger en multivariat normalfordeling for alle k og $(\mathbf{t}_1, \dots, \mathbf{t}_k)$ [1][s. 7].

De endelig dimensionelle fordelinger relateret til Gaussisk stokatiske felter kan findes i Appendiks B.1.

Da Gaussisk stokastiske felter nu er defineret, kan den log Gaussiske Cox proces indføres.

Definition 4.2.2 (Log Gaussisk Cox proces)
Lad X være en Cox proces på $\mathbb{R}^d$ drevet af $Z = \exp(Y)$, hvor Y er et Gaussisk stokastisk felt. Der siges X at være en log Gaussisk Cox proces.

Mere generelt kan andre specialtilfælde af Cox processer defineres, ved at lade intensitetsfunktionen for $X|Y$ være på formen $h(Y)$, hvor h er en ikke negativ funktion. For log Gaussiske Cox processer er $h(Y) = \exp(Y)$.

Fordelingen af (X, Y) viser sig at kunne bestemmes udelukkende ud fra middelværdifunktion $m(\xi)$ og kovariansfunktionen $c(\xi, \eta)$ for det Gaussisk stokastiske felt Y.

Proposition 4.2.3
For en log Gaussisk Cox proces X drevet af $Z = \exp(Y)$ er fordelingen af (X, Y) fuldstændig afgjort af

$$m(\xi) = \mathbb{E}\big(Y(\xi)\big) \quad \text{og} \quad c(\xi, \eta) = \text{Cov}\big(Y(\xi), Y(\eta)\big) \tag{4.1}$$

for det Gaussisk stokastiske felt Y.

Bevis
Idet X er drevet af det Gaussisk stokastiske felt, er fordelingen af X givet ud fra det Gaussisk stokastiske felt.

Det følger af Afsnit B.1.1, at et Gaussisk stokastisk felt er givet ved dets endelig dimensionelle fordelinger. Disse endelig dimensionelle fordelinger er netop den multivariate normalfordeling, som udelukkende er bestemt af m og c for det Gaussisk stokastiske felt. ■

Proposition 4.2.3 viser dermed, at Cox processer kan bestemmes ud fra middelværdi- og kovariansfunktioner. Dog er der visse betingelser til $c(\xi, \eta)$, som gør det til en kovariansfunktion.

4.2.1 Betingelser på kovariansfunktionen

I dette afsnit antages det, for simpelhedens skyld, at $c(\xi, \eta) = c(\xi - \eta)$ er translationsinvariant og på formen

$$c(\xi) = \sigma^2 r\left(\frac{\xi}{\alpha}\right).$$

Værdien $\sigma^2 > 0$ er variansen, r er en kendt korrelationsfunktion og $\alpha > 0$ er en korrelationsparameter. En funktion $r : \mathbb{R}^d \to [-1, 1]$ med $r(0) = 1$ for alle $\xi \in \mathbb{R}^d$ er en korrelationsfunktion for et Gaussisk stokastisk felt, hvis og kun hvis r er positiv semidefinit. Se bevis herfor i [1][s. 7]. Funktionen r er positiv semidefinit hvis $\sum_{i,j} a_i a_j r(\xi_i, \xi_j) \geq 0$ for alle $a_1, \ldots, a_n \in \mathbb{R}$ og alle $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$ for $n \in \mathbb{N}$. Om en funktion er positiv semidefinit besvares bedst ud fra spektralanalyse, og kan eksempelvis ses i [1][s. 28].

I dette speciale anvendes isotropiske korrelationsfunktioner. Det vil sige at $r(\xi) = r(\|\xi\|)$, hvorfra kovariansen mellem $Y(\xi)$ og $Y(\eta)$ kun afhænger af afstanden $\|\xi - \eta\|$. Korrelationsfunktionerne der anvendes er fra den *power eksponentielle* familie, som er på formen

$$r(\xi) = \exp(-\|\xi\|^\delta), \ \ 0 \leq \delta \leq 2. \tag{4.2}$$

Parameteren δ kontrollerer smoothing af realisationerne af det Gaussisk stokastiske felt, og er typisk en fast værdi.
I specialtilfældet hvor $\delta = 1$ kaldes korrelationsfunktionen en eksponentiel korrelationsfunktion, hvis $\delta = 1/2$ kaldes den en stabil korrelationsfunktion og hvis $\delta = 2$ kaldes den en Gaussisk korrelationsfunktion.

Følgende proposition giver et udtryk for intensiteten og parkorrelationen for en log Gaussisk Cox proces. Disse udtrykkes kun ud fra middelværdien og kovariansfunktionen for det tilhørende Gaussisk stokastiske felt.

Proposition 4.2.4
Lad X være en log Gaussisk Cox proces drevet af $Z = \exp(Y)$. Da er henholdsvis intensitetsfunktionen og parkorrelationsfunktionen givet som

$$\rho(\xi) = \exp(m(\xi) + c(\xi, \xi)/2) \ \text{ og } \ g(\xi, \eta) = \exp(c(\xi, \eta)). \tag{4.3}$$

Desuden gælder at X er stationær, hvis og kun hvis det underliggende Gaussisk stokastiske felt Y er stationært.

Bevis
Først vises at ligning (4.3) er opfyldt. Idet X er en Cox proces, kan Proposition 4.1.3 anvendes sådan at

$$\rho(\xi) = \mathbb{E}[Z(\xi)] = \mathbb{E}[\exp(Y(\xi))]$$

hvilket er den moment genererende funktion for $Y(\xi)$. Denne giver ud fra eksempel 3.66 i [2][s. 238] for $t = 1$ at

$$M_Y(t) = \exp(t\mu + \sigma^2 t^2/2)$$

med middelværdi μ og varians σ^2.
Dermed fås

$$\rho(\xi) = \exp(m(\xi) + c(\xi, \xi)/2). \tag{4.4}$$

Herefter anvendes Proposition 4.1.5 sådan at

$$\begin{aligned}\mathbb{E}[Z(\xi)Z(\eta)] &= \mathbb{E}[\exp(Y(\xi))\exp(Y(\eta))] \\ &= \mathbb{E}[\exp(Y(\xi) + Y(\eta))].\end{aligned}$$

Ved at anvende ligning (4.4), og at

$$\begin{aligned}\text{Cov}(Y(\xi) + Y(\eta), Y(\xi) + Y(\eta)) &= \text{Cov}(Y(\xi), Y(\xi)) + \text{Cov}(Y(\eta), Y(\eta)) \\ &\quad + 2\text{Cov}(Y(\xi), Y(\eta))\end{aligned}$$

samt

$$\mathbb{E}\big[Y(\xi) + Y(\eta)\big] = \mathbb{E}[Y(\xi)] + \mathbb{E}[Y(\eta)]$$

fås at

$$\begin{aligned}\mathbb{E}[Z(\xi)Z(\eta)] &= \exp(m(\xi) + m(\eta) + c(\xi,\xi)/2 + c(\eta,\eta)/2 + c(\xi,\eta)) \\ &= \exp(m(\xi) + c(\xi,\xi)/2)\exp(m(\eta) + c(\eta,\eta)/2)\exp(c(\xi,\eta)) \\ &= \rho(\xi)\rho(\eta)\exp(c(\xi,\eta)). \end{aligned} \tag{4.5}$$

Af Proposition 4.1.5 fås da

$$g(\xi, \eta) = \frac{\mathbb{E}[Z(\xi)Z(\eta)]}{\rho(\xi)\rho(\eta)} = \exp(c(\xi, \eta)).$$

Antag nu at X er stationær. Da er intensitetsfunktionen $\rho(\xi) = \rho$ konstant og parkorrelationen $g(\xi, \eta) = g(\xi - \eta)$ translations invariant. Da ligning (4.3) er bevist, gælder at

$$g(\xi - \eta) = \exp(c(\xi - \eta))$$

hvormed kovariansfunktionen for Y er translationsinvariant. For $\xi = \eta$ er parkorrelationen

$$g(\xi, \xi) = g(0) = \exp(c(\xi, \xi)) = \exp(\sigma^2(\xi)) = \exp(\sigma^2)$$

konstant, hvormed σ^2 er konstant. Da gælder at middelværdien er givet ved

$$m(\xi) = \log(\rho) - \sigma^2/2 = m,$$

som er konstant [8][s.456-457]. Antag omvendt at Y er stationær. Da gælder at middelværdien er konstant og kovariansen er tranlationsinvariant jævnfør [1]. Det vil sige

$$c(\xi, \eta) = c(\xi - \eta).$$

Dermed er $\sigma^2(\xi) = c(\xi - \xi) = c(0) = \sigma^2$ konstant. Da fås af ligning (4.4)

$$\rho(\xi) = \exp(m + \sigma^2/2) = \rho, \tag{4.6}$$

hvormed X er stationær.

$\blacksquare$

Bemærk at Proposition 4.2.4 etablerer en én-til-én korrespondance mellem (m, c) og (ρ, g), hvormed Proposition 4.2.3 giver, at fordelingen af (X, Y) er entydig bestemt af (ρ, g). Desuden kan ikke-parametriske estimater af m og c, via Proposition 4.2.4, opnås ud fra ikke-parametriske estimater af ρ og g givet i Afsnit 3.4.
Under antagelsen af en translationsinvariant korrelationsfunktion $c(\xi, \eta) = c(\xi - \eta)$, fås translations invarians af g. Dette giver andenordens intensitetsgenvægtet stationaritet per Proposition 3.1.8. Dermed kan $\mathcal{K}$ udregnes via ligning (3.8).

4.2.1.1 Simulation af log Gaussiske Cox processer

For at simulere en log Gaussisk Cox proces er det første, der skal simuleres det underliggende Gaussisk stokastiske felt. Dette skyldes, at fordelingen af en log Gaussisk Cox proces kun afhænger af fordelingen af det Gaussisk stokastiske felt $Y_B = \{Y(\xi) : \xi \in B\}$ for et begrænset område B.
For at approksimere Y_B gives en trinfunktion Y_B^{step}, som antager konstant værdi $Y(\eta)$ indenfor disjunkte områder C_η centreret rundt om en endelig mængde af lokationer $\eta \in B$, sådan at $B = \cup_{\eta \in I} C_\eta$. Ud fra en simulation af Y_B^{step} kan metoderne i Afsnit 2.2.2 anvendes til at simulere en Poisson proces med intensitetsfunktion $\exp(Y_B^{step})$. Dermed skal det Gaussisk stokastiske felt $\tilde{Y} = (Y(\eta))_{\eta \in I}$ simuleres. Simulation af Gaussisk stokastiske felter gennemgås ikke i dette speciale, dog kan metoder hertil findes i [14][s. 257-259] og [7][s. 278-279].

Composite Likelihood 5

For en Cox proces X på en mængde $B \subseteq S$ hvor $|B| < \infty$ er tætheden for X_B givet med hensyn til Poisson processen som

$$f(x) = \mathbb{E}\left[\exp\left(|B| - \int_B Z(\xi)d\xi\right)\prod_{\xi\in x} Z(\xi)\right].$$

Tætheden er givet for endelige punktkonfigurationer $x \subset B$ jævnfør [14][s. 60-61]. Et eksplicit udtryk for $f(x)$ er ofte ukendt, og integralet $\int_B Z(\xi)d\xi$ optræder som regel ikke på lukket form i følge [6][s. 1502].

Af denne grund introduceres *composite likelihood* metoden som en form for approksimation til tætheden for Cox processen. Metoden indføres med henblik på at anvende den til at tilpasse log Gaussiske Cox processer til data. Bemærk at metoden introduceres generelt for rumlige punktprocesser, men i [6] antydes det, at metoden anvendt på Poisson og log Gaussiske Cox processer resulterer i gode parameterestimater.
Dette kapitel er baseret på [6].

5.1 Composite likelihood estimation

I dette afsnit redefineres først visse summary statistics for senere at anvende disse til at konstruere en composite likelihood.

Lad X være en punktproces på $S \subseteq \mathbb{R}^2$.
Først defineres andenordens produkttætheden for punktprocessen X direkte som

$$\rho^{(2)}(\xi,\eta) = \lim_{|d\xi||d\eta|\to 0}\left(\frac{\mathbb{E}\big[N(d\xi)N(d\eta)\big]}{|d\xi||d\eta|}\right). \tag{5.1}$$

Her er N som tidligere antallet af punkter, og $d\xi$ er et infinitesimalt område, som indeholder punktet ξ.
Det ses, at udtrykket for $\rho^{(2)}$ i ligning (5.1) stemmer overens med Definition 3.1.2 idet

$$\frac{\partial^2\alpha^{(2)}(C)}{\partial\xi\partial\eta} = \frac{\partial^2\mathbb{E}\left[\sum_{\xi,\eta\in X}^{\neq}\mathbb{1}[(\xi,\eta)\in C]\right]}{\partial\xi\partial\eta} = \rho^{(2)}(\xi,\eta)$$

under antagelsen at $\xi \neq \eta$ i ligning (5.1). Hvis det samme punkt ikke forekommer i X mere end én gang, kan produkttætheden $\rho^{(2)}$ fortolkes som, at $\rho^{(2)}(\xi, \eta)d\xi d\eta$ er sandsynligheden for, at $d\xi$, $d\eta$ hver indeholder et punkt fra X hvor $\xi \neq \eta$.
Herefter defineres K-funktionen på anden vis som

$$K(r) = \frac{1}{\rho}\mathbb{E}[N_0(r)], \tag{5.2}$$

hvor $N_0(r)$ er antallet af punkter indenfor afstand r af et vilkårligt punkt.

5.1.1 Composite likelihood metoden

I dette afsnit vil der blive gennemgået en composite likelihood metode, som bruges til at fitte rumlige punktproces modeller til punktmønstre. En composite likelihood er en likelihood dannet af individuelle log-likelihoods, som hver er en marginal eller betinget log-likelihood. For at konstruere en composite likelihood er det nødvendigt først at definere en mængde af marginale eller betingede hændelser, for hvilke log-likelihooden nemt kan opnåes. Lad W være observationsvinduet. For at opnå førnævnte mængde af hændelser i forbindelse med rumlige punktprocesser, da betragtes to lokationer, ξ og η som begge er i W. Tæthedsfunktionen for at to punkter fra punktprocessen X optræder i ξ og η bliver da

$$f(\xi, \eta; \theta) = \frac{\rho^{(2)}(\xi - \eta; \theta)}{\int\int_W \rho^{(2)}(u - v; \theta)dudv}. \tag{5.3}$$

Dette følger af at produkttætheden er sandsynligheden for at to punkter i $\mathbb{R}^2$ er indeholdt i X. Produkttætheden normeres med dets integrale for at danne en tæthed.
I ovenstående indgår θ i udtrykket for $f(\xi, \eta; \theta)$ og $\rho^{(2)}(\xi; \theta)$. Dette betyder, at f og $\rho^{(2)}$ begge afhænger af den ukendte parameter θ, som bestemmer fordelingen af X. For funktionen $f(\xi, \eta; \theta)$ givet ved ligning (5.3) ses det, at denne er en tæthedsfunktion. Det følger af, at $\rho^{(2)}$ er en streng ikke-negativ funktion, og integration af f over ξ og η i W er lig 1. Dette ses ved

$$\begin{aligned}
\int\int_W f(\xi, \eta; \theta)d\xi d\eta &= \int\int_W \frac{\rho^{(2)}(\xi - \eta; \theta)}{\int\int_W \rho^{(2)}(u - v; \theta)dudv} d\xi d\eta \\
&= \frac{1}{\iint_W \rho^{(2)}(u - v; \theta)dudv} \iint_W \rho^{(2)}(\xi - \eta; \theta)d\xi d\eta \\
&= 1.
\end{aligned}$$

Dermed bliver den parvise log-likelihood

$$L(\xi, \eta; \theta) = \log[\rho^{(2)}(\xi - \eta; \theta)] - \log\Big[\iint_W \rho^{(2)}(u - v; \theta)dudv\Big] \tag{5.4}$$

en log-likelihood funktion. Lad nu $N^{(2)}(d\xi, d\eta) = N(d\xi)N(d\eta) \cdot \mathbb{1}[\xi \neq \eta]$ være antallet af punktpar i $d\xi \cup d\eta$, hvor der gælder at $\xi \neq \eta$. Alle log-likelihoods for par af ξ og η i $X \cap W$ summeres, og composite likelihood opnås ved

$$\iint_W \log[\rho^{(2)}(\xi - \eta; \theta)] N^{(2)}(d\xi, d\eta) - \iint_W N^{(2)}(d\xi, d\eta) \log\Big[\iint_W \rho^{(2)}(u - v; \theta) du dv\Big].$$

I praksis er par med stor afstand mellem sig ikke af interesse. Disse par er ofte næsten uafhængige, og derfor vil det ikke medføre betydelig tab af information, at eliminere sådanne par. Bidragene fra par med kortere afstand mellem sig kan få mindre betydning ved at inkludere for mange par af punkter med stor afstand mellem sig. Derudover kan fjernelsen af par med stor afstand mellem sig føre til stor reducering i beregningsarbejdet. Derfor vil fokus være på de par, som ligger indenfor en specificeret afstand r af hinanden. Betinget på at et observeret par af punkter ligger inden for afstanden r af hinanden, da fås følgende tæthedsfunktion for, at punkter i X forekommer i både ξ og η

$$f(\xi, \eta; \theta) = \frac{\rho^{(2)}(\xi - \eta; \theta)}{\iint_W \mathbb{1}[\|u - v\| < r] \rho^{(2)}(u - v; \theta) du dv}$$

hvor $\|\xi - \eta\| < r$. Tætheden leder til følgende parvise log-likelihood tilsvarende ligning (5.4)

$$L(\xi, \eta; \theta) = \log[\rho^{(2)}(\xi - \eta; \theta)] - \log\Big[\iint_W \mathbb{1}[\|u - v\| < r] \rho^{(2)}(u - v; \theta) du dv\Big] \tag{5.5}$$

hvor $\|\xi - \eta\| < r$. Bemærk at ligning (5.5) kan ses som en vægtet version af ligning (5.4), på den måde at hvert par af punkter får en vægt på 1, hvis de ligger inden for en afstand af r og en vægt på 0 ellers.

Lad $w(\xi)$ være en prædefineret ikke-negativ vægtfunktion som tager værdier forskellig fra nul kun hvis $\|\xi\| < r$ for et givet r. Følgende kan da ses som en generaliseret version af ligning (5.5)

$$L(\xi, \eta; \theta) = \Big(\log[\rho^{(2)}(\xi - \eta; \theta)] - \log\Big[\iint_W w(u - v) \rho^{(2)}(u - v; \theta) du dv\Big]\Big) \cdot w(\xi - \eta). \tag{5.6}$$

Det ses, at hvis $w(\xi) = \mathbb{1}[\|\xi\| < r]$, da bliver ligning (5.6) til ligning (5.5). Ved at bruge en bestemt vægtfunktion, som tilføjer mindre vægt til par af punkter med stor afstand mellem sig, opnås i praksis mindre påvirkning fra disse par. Dermed bliver estimationen mere robust i forhold til valget af r. Ud fra den parvise log-likelihood defineret i ligning (5.6) kan den følgende generelle vægtede composite likelihood defineres som

$$\begin{aligned} CL(r; \theta) = &\iint_W w(\xi - \eta) \log\left[\rho^{(2)}(\xi - \eta; \theta)\right] N^{(2)}(d\xi, d\eta) \\ &- \iint_W w(\xi - \eta) N^{(2)}(d\xi, d\eta) \cdot \log\Big[\iint_W w(u - v) \rho^{(2)}(u - v; \theta) du dv\Big]. \end{aligned} \tag{5.7}$$

Composite likelihood metoden benyttes dermed til at estimere den ukendte parametervektor θ, som ligesom i maksimum likelihood estimation, er defineret som den værdi af θ som maksimerer $CL(r, \theta)$. Et maksimum-likelihood estimat af θ er givet som $\hat{\theta}$. For at udregne $\hat{\theta}$ skal integralet med u og v i ligning (5.7) udregnes. Dette kan i princippet gøres ved brug af numeriske metoder jævnfør [6][s. 1504]. Dette kan dog kræve en del beregninger, hvormed en alternativ metode introduceres.

5.1.2 Alternativ composite likelihood

Dette afsnit giver et alternativ til formen af composite likelihood funktionen introduceret i forrige afsnit. Lad W' være defineret som i Afsnit 3.4.3. Herefter defineres den vægtede K-funktion som

$$K_w(r;\theta) = \frac{1}{\rho^2} \int_{\|\xi\|<r} w(\xi)\rho^{(2)}(\xi;\theta)d\xi \tag{5.8}$$

for en givet vægtfunktionen w.
Alternativet til composite likelihood formen opnås ved at modificere ligning (5.7). Den faktor fra ligning (5.7) som modificeres er $\iint_W w(u-v)\rho^{(2)}(u-v;\theta)dudv$. Ved at u integreres over W', og v integreres over W fås det følgende

$$\int_W \int_{W'} w(u-v)\rho^{(2)}(u-v,\theta)dudv = \int_{W'} \int_W w(u-v)\rho^{(2)}(u-v,\theta)dvdu. \tag{5.9}$$

Ved at lade $x = u - v$ og $W_{-u} = \{z : z + u \in W\}$ da fås, at ligning (5.9) bliver

$$\int_{W'} \int_W w(u-v)\rho^{(2)}(u-v,\theta)dvdu = \int_{W'} \int_{W_{-u}} w(x)\rho^{(2)}(x,\theta)dxdu. \tag{5.10}$$

Da $u \in W'$ så er $b(0,r) \subseteq W_{-u}$, og der gælder at $w(x) = 0$ på $W_{-u} \setminus b(0,r)$, hvormed ligning (5.10) bliver

$$\begin{aligned}\int_{W'} \int_{b(0,r)} w(x)\rho^{(2)}(x,\theta)dxdu &= |W'| \int_{b(0,r)} w(x)\rho^{(2)}(x,\theta)dx \\ &= \rho^2|W'|K_w(r,\theta). \end{aligned} \tag{5.11}$$

Hvis den teoretiske værdi for $K_w(r;\theta)$ er kendt, kan den modificerede udgave af composite likelihooden, hvor ligning (5.11) indgår som faktor, hurtigt udregnes. Bemærk at hvis vægtfunktionen for den vægtede K-funktion K_w er givet $w(\xi) = \mathbb{1}\left[\|\xi\| < r\right]$, da bliver $K_w(r;\theta) = K(r;\theta)$. Idet at udtrykket for K er kendt for både Poisson processen og Cox-processer (gennem g), vil dette lette beregningen i disse tilfælde. Denne beregningsmetode benævnes *indre område* metoden. Bemærk, at det i ovenstående udregninger er antaget, at integrationsgrænserne er ombyttelige.

5.2 Konsistens for $\hat{\theta}$

Som tidligere nævnt er målet med composite likelihood at estimere den ukendte parametervektor θ. Hvis denne estimeres som maksimum af $CL(r;\theta)$, viser det sig, at der gælder konsistens for estimatoren $\hat{\theta}$.

Lad θ_0 være den sande parametervektor for modellen og Θ være parameterrummet for θ. Antag, at Θ er kompakt (lukket og begrænset), og at θ_0 er et indre punkt i Θ. Bemærk, at antagelsen om at θ_0 er indre punkt i Θ ikke altid er opfyldt, da Θ er lukket.

De asymptotiske egenskaber for $\hat{\theta}$ undersøges for en følge af observationsområder W_n. Lad ∂W_n være randen af W_n, $|W_n|$ være arealet af W_n og $|\partial W_n|$ være længden af randen af W_n. Det antages at størrelsesordenen på W_n og ∂W_n er

$$|W_n| = O(n^2) \text{ og } |\partial W_n| = O(n). \tag{5.12}$$

Denne antagelse medfører at observationsområdet vokser i alle retninger, når n vokser. Lad $W'_n = \{\xi \mid \xi + \eta \in W_n,\ \xi \in W_n,\ \|\eta\| < r\}$. Ved at tage udgangspunkt i ligning (5.6), og integrere u over W'_n og v over W_n fås

$$\begin{aligned}
L(\xi,\eta;\theta) &= \Big(\log[\rho^{(2)}(\xi-\eta;\theta)] - \log\Big[\int_{W_n}\int_{W'_n} w(u-v)\rho^{(2)}(u-v;\theta)dudv\Big]\Big)\cdot w(\xi-\eta)\\
&= \Big(\log[\rho^{(2)}(\xi-\eta;\theta)] - \log\Big[\rho^2|W'_n|K_w(r,\theta)\Big]\Big)\cdot w(\xi-\eta)\\
&= \Big(\log[\rho^{(2)}(\xi-\eta;\theta)] - \log\Big[\rho^2|W'_n|\Big] - \log\Big[K_w(r,\theta)\Big]\Big)\cdot w(\xi-\eta)
\end{aligned} \tag{5.13}$$

hvor anden lighed følger af ligning (5.11).

Der kan ses bort fra konstanten $\log\Big[\rho^2|W'_n|\Big]\cdot w(\xi-\eta)$ i ligning (5.13), da den ikke har nogen konsekvens for maksimering af L med hensyn til θ. Dermed fås det at

$$L(\xi,\eta;\theta) = \Big(\log[\rho^{(2)}(\xi-\eta;\theta)] - \log\Big[K_w(r,\theta)\Big]\Big)\cdot w(\xi-\eta). \tag{5.14}$$

Ved at differentiere udtrykket i ligning (5.14) med hensyn til θ fås parvis composite score funktionen som

$$U(\xi,\eta;\theta) = \left(\frac{\rho^{(2)'}(\xi-\eta;\theta)}{\rho^{(2)}(\xi-\eta;\theta)} - \frac{K'_w(r;\theta)}{K_w(r;\theta)}\right)\cdot w(\xi-\eta). \tag{5.15}$$

Notationen $f'(\xi,\theta)$ for en funktion f angiver differentation af f med hensyn til θ. Det vil sige $f'(\xi,\theta) = \frac{df}{d\theta}(\xi,\theta)$.

Ved at 'addere' alle parvise composite score funktioner fås da composite score funktionen

$$\begin{aligned}
U_n(\theta) = &\frac{1}{|W'_n|}\int_{W_n}\int_{W'_n} w(\xi-\eta)\frac{\rho^{(2)'}(\xi-\eta;\theta)}{\rho^{(2)}(\xi-\eta;\theta)}N^{(2)}(d\xi,d\eta)\\
&-\frac{1}{|W'_n|}\int_{W_n}\int_{W'_n} w(\xi-\eta)N^{(2)}(d\xi,d\eta)\frac{K'_w(r;\theta)}{K_w(r;\theta)}.
\end{aligned} \tag{5.16}$$

Her multipliceres med faktoren $\frac{1}{|W_n'|}$ for at normalisere dobbelt integration over W_n og W_n'. Estimatoren givet som $\hat{\theta}_n$ for θ på W_n er da defineret som løsningen til $U_n(\theta) = 0$. Middelværdien af $U_n(\theta)$ kan udregnes ved

$$\begin{aligned}\mathbb{E}\big[U_n(\theta)\big] = \mathbb{E}\Bigg[&\frac{1}{|W_n'|}\int_{W_n}\int_{W_n'} w(\xi-\eta)\frac{\rho^{(2)'}(\xi-\eta;\theta)}{\rho^{(2)}(\xi-\eta;\theta)}N^{(2)}(d\xi,d\eta)\\ &-\frac{1}{|W_n'|}\int_{W_n}\int_{W_n'} w(\xi-\eta)N^{(2)}(d\xi,d\eta)\frac{K_w'(r;\theta)}{K_w(r;\theta)}\Bigg].\end{aligned} \tag{5.17}$$

Ved at anvende Proposition 3.1.5 ligning (3.5) og lade $x = \xi - \eta$ og $W_n^{-\xi} = \{z : z + \xi \in W_n\}$ fås

$$\begin{aligned}\mathbb{E}\big[U_n(\theta)\big] &= \frac{1}{|W_n'|}\int_{W_n}\int_{W_n'} w(\xi-\eta)\frac{\rho^{(2)'}(\xi-\eta;\theta)}{\rho^{(2)}(\xi-\eta;\theta)}\rho^{(2)}(\xi-\eta;\theta_0)d\xi d\eta\\ &\quad-\frac{1}{|W_n'|}\int_{W_n}\int_{W_n'} w(\xi-\eta)\rho^{(2)}(\xi-\eta;\theta_0)d\xi d\eta\frac{K_w'(r;\theta)}{K_w(r;\theta)}\\ &= \frac{1}{|W_n'|}\int_{W_n'}\int_{W_n^{-\xi}} w(x)\frac{\rho^{(2)'}(x;\theta)}{\rho^{(2)}(x;\theta)}\rho^{(2)}(x;\theta_0)dxd\xi\\ &\quad-\frac{1}{|W_n'|}\int_{W_n'}\int_{W_n^{-\xi}} w(x)\rho^{(2)}(x;\theta_0)dxd\xi\frac{K_w'(r;\theta)}{K_w(r;\theta)}\\ &= \int_{\|x\|<r} w(x)\frac{\rho^{(2)'}(x;\theta)}{\rho^{(2)}(x;\theta)}\rho^{(2)}(x;\theta_0)dx\\ &\quad-\int_{\|x\|<r} w(x)\rho^{(2)}(x;\theta_0)dx\frac{K_w'(r;\theta)}{K_w(r;\theta)}\\ &= \int_{\|x\|<r} w(x)\frac{\rho^{(2)'}(x;\theta)}{\rho^{(2)}(x;\theta)}\rho^{(2)}(x;\theta_0)dx - \rho^2K_w(r;\theta_0)\frac{K_w'(r;\theta)}{K_w(r;\theta)}.\end{aligned}$$

Sidste ligning følger af definitionen af den vægtede K-funktion i ligning (5.8).
Hvis $\theta = \theta_0$ så fås af udtrykket for $\mathbb{E}\big[U_n(\theta)\big]$ at

$$\begin{aligned}\mathbb{E}\big[U_n(\theta_0)\big] &= \int_{\|x\|<r} w(x)\frac{\rho^{(2)'}(x;\theta_0)}{\rho^{(2)}(x;\theta_0)}\rho^{(2)}(x;\theta_0)dx - \rho^2K_w(r;\theta_0)\frac{K_w'(r;\theta_0)}{K_w(r;\theta_0)}\\ &= \int_{\|x\|<r} w(x)\rho^{(2)'}(x;\theta_0)dx - \rho^2K_w'(r;\theta_0)\\ &= \int_{\|x\|<r} w(x)\rho^{(2)'}(x;\theta_0)dx - \Big(\int_{\|x\|<r} w(x)\rho^{(2)}(x;\theta_0)dx\Big)'\\ &= \int_{\|x\|<r} w(x)\rho^{(2)'}(x;\theta_0)dx - \int_{\|x\|<r} w(x)\rho^{(2)'}(x;\theta_0)dx\\ &= 0.\end{aligned}$$

Der gælder følgende om estimatoren $\hat{\theta}_n$ for θ på W.

Sætning 5.2.1
Antag at betingelsen i ligning (5.12) er opfyldt, at X er *ergodisk*, at $\mathbb{E}[U_n(\theta)] = 0$ kun hvis $\theta = \theta_0$ samt at $\rho^{(2)'}(x,\theta)/\rho^{(2)}(x,\theta)$ og $K'_w(r,\theta)/K_w(r,\theta)$ er begrænsede og kontinuerte funktioner. Så er $\hat{\theta}_n$ en konsistent estimator for θ.

Bevis
For bevis se [6][s. 1511]. ■

En definition af ergodisk kan findes i Definition A.2.1.
Det undersøges, om betingelserne i Sætning 5.2.1 er opfyldt netop for den log Gaussiske Cox proces. Følgende sætning giver et kriterie til at afgøre hvorvidt en log Gaussisk Cox proces er ergodisk.

Sætning 5.2.2
Lad Y være et stationært Gaussisk stokastisk felt på $\mathbb{R}^2$ med en korrelationsfunktion som går mod nul

$$r(\xi) \to 0 \quad \text{for} \quad \|\xi\| \to \infty. \tag{5.18}$$

Da er den tilsvarende log Gaussiske Cox proces ergodisk [8][s. 459].

Bevis
For bevis se [8][s. 459-460] ■

Bemærk at den power eksponentielle familie af korrelationsfunktioner på formen givet i ligning (4.2) opfylder betingelse (5.18) i Sætning 5.2.2. Dermed er homogene log Gaussiske Cox processer, med denne form for korrelation, ergodiske. Bemærk desuden at betingelse (5.18) svarer til betingelsen

$$g(\xi) \to 1 \quad \text{for} \quad \|\xi\| \to \infty \tag{5.19}$$

på baggrund af Proposition 4.2.4.
For en log Gaussisk Cox proces er $\rho^{(2)}(\xi,\eta) = \rho(\xi)\rho(\eta)\exp(c(\xi,\eta))$ per ligning (4.5) hvor $\rho(\xi) = \exp(m(\xi) + c(\xi,\xi)/2)$. Hvis det underliggende Gaussisk stokastiske felt Y er stationært, er middelværdien $m(\xi)$ og variansen $\sigma^2(\xi)$ konstante per definition. Det ses ud fra Proposition 4.2.4, at hvis det Gaussisk stokatiske felt er stationært, er den tilhørende log Gaussiske Cox proces homogen. Så er $\rho^{(2)}(\xi,\eta) = \rho^2 \exp(c(\xi,\eta))$. Ved at lade den translationsinvariante kovariansfunktion være givet ud fra den power eksponentielle familie af korrelationsfunktioner fra ligning (4.2) fås

$$c(\xi,\eta) = c(\xi-\eta,0) = c(x) = \sigma^2\exp(-\|x/\alpha\|^\delta), \quad 0 \le \delta \le 2.$$

Da fås andenordens produktætheden som

$$\rho^{(2)}(x,\alpha) = \rho^2 \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big).$$

Ved at differentiere med hensyn til α fås af kædereglen

$$\frac{\partial \rho^{(2)}(x,\alpha)}{\partial \alpha} = \rho^2 \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)(\delta|\alpha|^{-(\delta+1)}\|x\|^\delta \frac{\alpha}{|\alpha|}). \quad (5.20)$$

Dermed fås

$$\begin{aligned}\frac{\rho^{(2)\prime}(x,\alpha)}{\rho^{(2)}(x,\alpha)} &= \frac{\rho^2 \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)(\delta|\alpha|^{-(\delta+1)}\|x\|^\delta \frac{\alpha}{|\alpha|})}{\rho^2 \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)} \\ &= \sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)(\delta|\alpha|^{-(\delta+1)}\|x\|^\delta \frac{\alpha}{|\alpha|}) \\ &= \frac{\sigma^2\delta\alpha|\alpha|^{-2}\|x/\alpha\|^\delta}{\exp(\|x/\alpha\|^\delta)}. \quad (5.21)\end{aligned}$$

Det ses, at udtrykket i ligning (5.21) er kontinuert både med hensyn til x og α ved hjælp af Sætning 6.19, Sætning 6.20, Sætning 6.21 samt Sætning 5.4 og 5.5 i [10][s.71-94]. Kontinuiteten er naturligvis kun gældende for $\alpha \neq 0$, som også er tilfældet i resten af afsnittet.
Idet at Θ er antaget kompakt (i $\mathbb{R}^n$), er ligning (5.21) begrænset med hensyn til α på Θ på grund af kontinuiteten per Sætning 6.26 i [10][s.96]. Desuden gælder det, at udtrykket i ligning (5.21) er begrænset med hensyn til x, da eksponentialfunktioner går hurtigere mod ∞ end potensfunktioner per Sætning 7.23 i [10][s. 125].

Den vægtede K-funktion er defineret som

$$K_w(r;\theta) = \frac{1}{\rho^2}\int_{\|\xi\|<r} w(\xi)\rho^{(2)}(\xi;\theta)d\xi.$$

For den homogene log Gaussiske Cox proces bliver dette

$$K_w(r;\alpha) = \int_{\|x\|<r} w(x)\exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)dx.$$

I dokumentationen for `spatstat`-pakken er vægtfunktionen givet som $w(x) = \mathbb{1}[\|x\| \leq R]$, hvor $R = \frac{1}{2}r_{\max}$. Værdien $r_{\max}$ kan specificeres, men er per default sat til $\frac{1}{4}$ af længden af den korteste side af observationsvinduet. Dermed bliver den vægtede K-funktion

$$K_w(R;\alpha) = \int_{\|x\|\leq R} \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)dx, \quad (5.22)$$

hvor funktionen indenfor integralet er kontinuert per Sætning 6.19, Sætning 6.20, Sætning 6.21 og Sætning 5.5 i [10][s. 71-94]. Da Θ er lukket, følger det af Sætning 10.13 sammen med Sætning 8.15 i [10][s. 210 & s. 143], at $K_w(R;\alpha)$ er differentiabel med hensyn til α, som medfører kontinuitet per Sætning 7.15 i [10][s.113].
Udtrykket for $\frac{\partial K_w(R;\alpha)}{\partial \alpha}$ kan findes ved at differentiere $K_w(R;\alpha)$ fra ligning (5.22) med hensyn til α. Ved anvendelse af ligning (5.20) fås

$$\begin{aligned}
\frac{\partial K_w(R;\alpha)}{\partial \alpha} &= \frac{\partial\Big(\int_{\|x\|\le R} \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)dx\Big)}{\partial \alpha} \\
&= \int_{\|x\|\le R} \frac{\partial\Big(\exp\Big(\sigma^2\exp\big(-\|x/\alpha\|^\delta\big)\Big)\Big)}{\partial\alpha} dx \\
&= \int_{\|x\|\le R} \rho^2 \exp\Big(\sigma^2 \exp\big(-\|x/\alpha\|^\delta\big)\Big)\sigma^2\exp\big(-\|x/\alpha\|^\delta\big)(\delta|\alpha|^{-(\delta+1)}\|x\|^\delta \frac{\alpha}{|\alpha|})dx. \qquad (5.23)
\end{aligned}$$

Det ses, at integranden i ligning (5.23) er kontinuert, på baggrund af samme sætninger som medførte kontinuiteten for integranden fra ligning (5.22). Da Θ er lukket, følger det igen af Sætning 10.13 sammen med Sætning 8.15 i [10][s. 210 & s. 143], at $\frac{\partial K_w(R;\alpha)}{\partial \alpha}$ er differentiabel med hensyn til α. Dette medfører kontinuitet per Sætning 7.15 i [10][s.113].
Da fås, at udtrykket fra Sætning 5.2.1

$$\frac{K_w'(R;\alpha)}{K_w(R;\alpha)} \tag{5.24}$$

er kontinuert med hensyn til α. Da (5.24) er kontinuert, og Θ er lukket og begrænset, så følger det af Sætning 6.26 i [10][s. 96], at udtrykket i (5.24) er begrænset med hensyn til α.

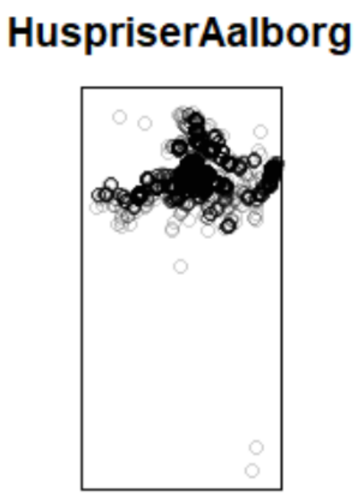
Det er dermed vist, at betingelserne i Sætning 5.2.1 er opfyldt for en homogen log Gaussisk Cox proces med power eksponentiel korrelationsfunktion. Det betyder, at estimatoren $\hat{\theta}$ er konsistent for den homogene log Gaussiske Cox proces.

Databehandling 6

I dette kapitel bliver specialets databehandling gennemgået. Databehandlingen er lavet på baggrund af data, som stammer fra ejendomsmæglerfirmaet Home. Alt bearbejdningen af data er lavet i R via platformen RStudio. I kapitlet undersøges først datasættes rumlige struktur ud fra summary statistics, hvorefter ni forskellige log Gaussiske Cox processer tilpasses datasættet ved hjælp af composite likelihood metoden. Denne modeltilpasning eftertjekkes med punktvise envelopes. Alle funktioner til databearbejdningen er i forvejen implementeret i R-pakken `spatstat` [4].

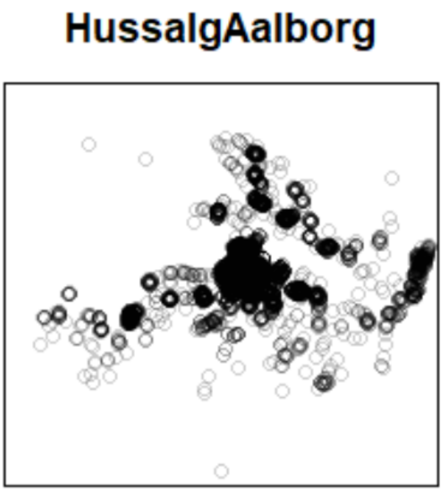
6.1 Databeskrivelse

I dette speciale arbejdes der ud fra et datasæt over hussalg i Danmark i perioden 2004 til 2016. Datasættet består af 163888 observationer over 48 variable. Grundet den teoretiske rammesætning tager dette speciale kun højde for husenes geografiske placering gennem UTM-koordinater som variable. For at begrænse beregningstiden for selve modelfitting algoritmen, restringeres der til boliger indenfor Aalborg Kommune. Dermed består det behandlede datasæt kun af 7223 observationer over 2 variable i form af x- og y-koordinater. Første trin i databehandlingen er at indlæse datasættet, og omdanne længde- og breddegradskoordinater til UTM koordinater. R-koden for dette kan ses i Listing C.1 i Appendiks C. En gennemgang af hvordan datasættet bliver omdannet til et punktmønster kan findes i Appendiks C. Et plot af punktmønsteret, som repræsenterer hussalg i Aalborg Kommune, kan ses på Figur 6.1.



Figur 6.1: *Punktmønster bestående af hussalg i Aalborg Kommune.*

Det bemærkes at der er to outliers i punktmønsteret set på Figur 6.1. Disse to punkter viser sig at ligge i Aarhus Kommune, så disse fjernes og observationsvinduet justeres. R-koden for denne procedure kan findes i Listing C.6. Et plot for punktmønsteret for det tilpassede datasæt ses på Figur 6.2.

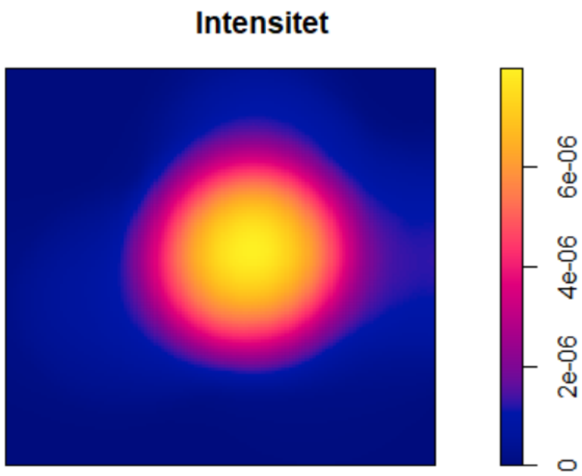


Figur 6.2: *Reduceret punktmønster bestående af hussalg i Aalborg Kommune.*

Bemærk at punktmønsteret på Figur 6.2 ser ud til at være clusteret, det vil sige at parkorrelationen $g > 1$. I det næste afsnit vil det blive undersøgt, om punktmønsteret udviser clustering. Hertil anvendes de summary statistics, som blev introduceret i Kapitel 3. For at præcisere hvilket punktmønster der undersøges fremover, er det punktmønsteret tilsvarende det reducerede datasæt bestående af hussalg i Aalborg Kommune set på Figur 6.2. I punktmønsteret er der korrigeret for dobbeltpunkter. Det vil sige, at de boliger som er solgt flere gange over perioden 2004-2016, kun repræsenteres en gang.

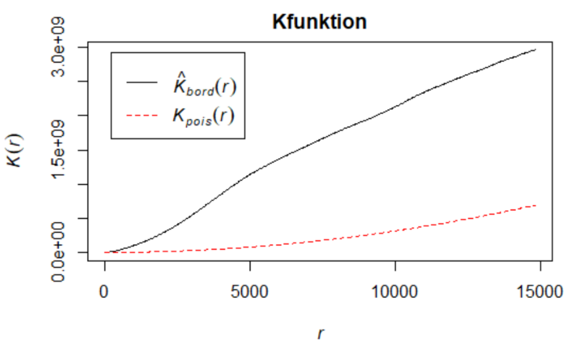
6.2 Indledende databehandling

I dette afsnit anvendes summary statistics til at undersøge punktstrukturen i punktmønsteret over hussalg i Aalborg Kommune. Først gives et grafisk plot af intensiteten gennem `density.ppp` kommandoen i `spatstat`. Funktionen `density.ppp` anvender det ikke-parametriske kerneestimat for intensitetsfunktionen givet i ligning (3.20).

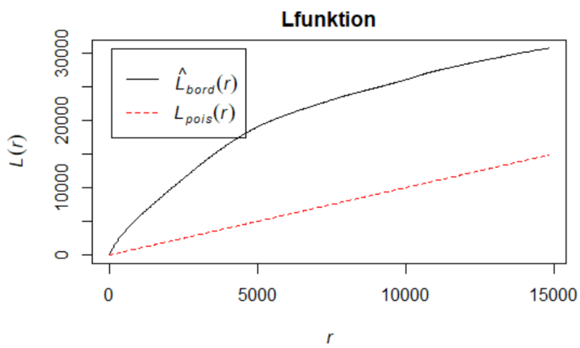


Figur 6.3: *Intensitet for punktmønsteret bestående af hussalg i Aalborg Kommune.*

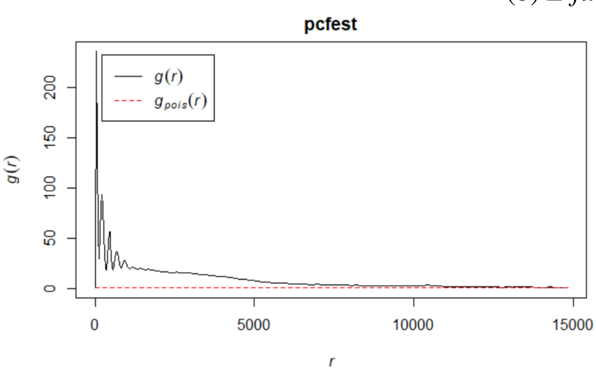
Intensiteten udviser samme clustertendenser som punktmønsteret for datasættet, men for at undersøge om dette gør sig gældende, estimeres andre summary statistics. Først undersøges K- og L-funktionen samt parkorrelationen g. Estimater for K-funktionen fås gennem `Kest` kommandoen i R. Denne giver et ikke-parametrisk estimat for K-funktionen introduceret i Afsnit 3.4.3. `Lest`-funktionen giver et tilsvarende estimat for L-funktionen, og `pcf`-kommandoen giver ikke-parametriske estimater af parkorrelationsfunktionen introduceret i Afsnit 3.4.4.



(a) *K-funktionen.*



(b) *L-funktionen.*

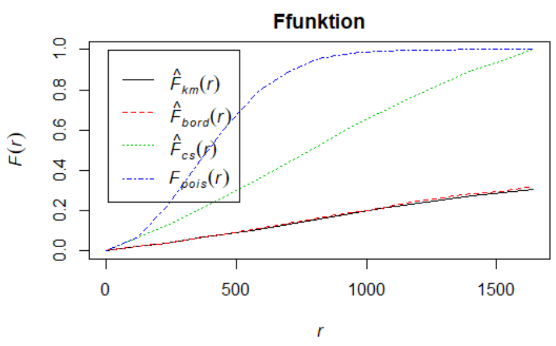


(c) *Parkorrelationsfunktionen.*

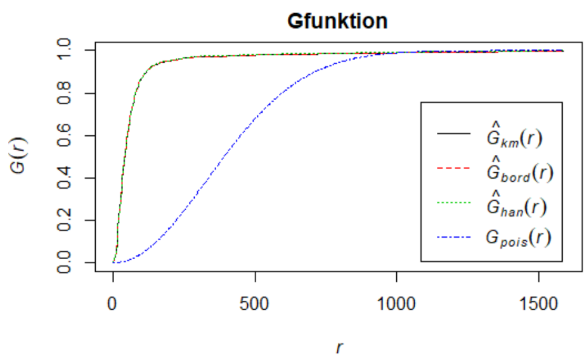
Figur 6.4: *Plot af K-, L- og Parkorrelationsfunktionen for punktmønsteret bestående af hussalg i Aalborg Kommune.*

Det ses på Figur 6.4a, at den empiriske K-funktion ligger over den teoretiske K-funktion for Poisson processen. Dette indikerer clustering i punktmønsteret. Desuden ligger den empiriske L-funktion også over den teoretiske L-funktion for en Poisson proces set på Figur 6.4b. Dermed fås det også af L-funktionen, at punktmønsteret udviser clustertendenser. På Figur 6.4c ses parkorrelationsfunktionen for punktmønsteret over hussalg i Aalborg Kommune. Parkorrelationen er estimeret ud fra estimatet af K-funktionen gennem ligning (3.9). Det ses, at estimatet af g ligger over den teoretiske værdi for Poisson processen, hvilket indikerer at punktmønsteret er clusteret.

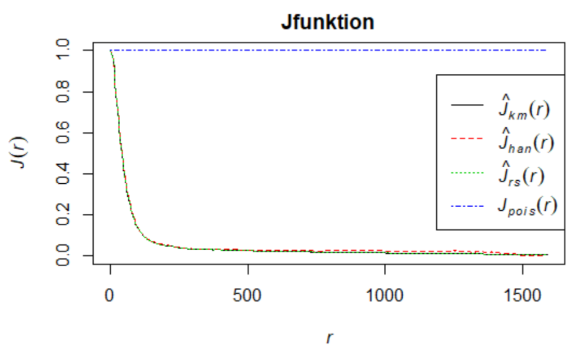
For yderligere at undersøge hvorvidt punktmønsteret udviser clustering, udregnes summary statistics baseret på interne punktafstande. Den empiriske F-funktion kan bestemmes ud fra kommandoen `Fest`, og dens tilhørende plot ses på Figur 6.5a. `Fest`-funktionen giver forskellige estimater af F-funktionen. Værdien $\hat{F}_{bord}$ er det reducerede sample estimat givet i Afsnit 3.4.5. På samme måde gives estimater af G- og J-funktionen gennem kommandoerne `Gest` og `Jest`, hvor $\hat{G}_{bord}$ og $\hat{J}_{rs}$ er det reducerede sample estimat for henholdsvis G- og J-funktionen. De resterende estimater set på Figur 6.5 er ikke introduceret i specialet.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.5: *Plot af F-, G- og J-funktionen for punktmønsteret bestående af hussalg i Aalborg Kommune.*

På Figur 6.5a og Figur 6.5b ses den teoretiske værdi for F- og G-funktionen under antagelse af en Poisson proces. Det vides per Proposition 3.3.4, at disse værdier teoretisk set er lig hinanden for en homogen Poisson proces. Dermed ses, at det for estimaterne af F- og G-funktionen for datasættet er opfyldt at $F(r) < G(r)$, hvilket per Afsnit 3.3 indikerer, at punktmønsteret er clusteret. Det ses at J-funktionen er mindre end 1 for alle værdier af r, hvilket også indikerer clustering.

Alle de anvendte summary statistics antyder dermed, at punktmønsteret bestående af hussalg i Aalborg Kommune er mere clusteret end en Poisson proces. Det vil sige, at hussalgene i Aalborg Kommune i den pågældende periode påvirker hinanden positivt. Dette skal forstås sådan, at et hussalg har større sandsynlighed for at forekomme i forbindelse med andre hussalg, end at hussalgene forekommer uafhængigt af hinanden.

I diskussionen efter Proposition 4.1.6 blev det etableret at en Cox proces generelt vil forekomme mere clusteret end en Poisson proces. Det kan altså vise sig interessant at undersøge, hvorvidt en Cox proces (log Gaussisk Cox proces) er en bedre model for punktmønsteret end Poisson processen. I det efterfølgende tilpasses forskellige log Gaussiske Cox processer til punktmønsteret for hussalg i Aalborg Kommune.

6.3 Model fitting af log Gaussiske Cox processer

I dette afsnit påbegyndes opsætningen af den model, som er introduceret gennem teoriafsnittene netop den log Gaussiske Cox proces. I `spatstat` pakken findes en indbygget funktion kaldet `kppm`, som fitter en Cox eller Cluster proces til datasættet. Dette udføres ved brug af forskellige modeltilpasnings metoder. Den pågældende metode vælges på baggrund af teorien indført i Kapitel 5 til composite likelihood metoden. I `kppm`-funktionen kan typen af korrelationsfunktion for det underliggende Gaussisk stokastiske felt fastsættes. I Afsnit 4.2.1 er der indført tre korrelationsfunktioner: Den eksponentielle, den Gaussiske og den stabile korrelationsfunktion. Bemærk, at hvis modellen der fittes er en log Gaussisk Cox proces, skal `kppm`-funktionen simulere det underliggende Gaussisk stokastiske felt gennem pakken `RandomFields`.

Afsnittet indeles i tre underafsnit, hvor log Gaussiske Cox processer med tre forskellige intensitetsfunktioner fittes til datasættet. For hver intensitetsfunktion tilpasses tre log Gaussiske Cox processer, hvor det underliggende Gaussisk stokastiske felt har eksponentiel, Gaussisk eller stabil korrelationsfunktion. Afslutningsvist sammenlignes korrelationsfunktionerne for de i alt ni forskellige modeller.

6.3.1 Homogen log Gaussisk Cox proces

I dette delafsnit fittes tre homogene log Gaussiske Cox processer til datasættet, og disse evalueres med punktvise envelopes for summary statistics. Afsnittet inddeles i tre underafsnit; et for hver korrelationsfunktion.

Den første indgang i `kppm` specificerer formen på log af intensiteten. I det homogene tilfælde angiver ~ 1, at log af intensiteten er konstant. Denne model vil da antage log-konstant formen

$$\log \rho(\xi) = a(\xi) = a.$$

Idet at modellen er en log Gaussisk Cox proces, følger det af Proposition 4.2.4 at

$$\rho(\xi) = \exp(m(\xi) + c(\xi, \xi)/2).$$

Dermed fås en formel for middelværdien for konstant intensitetsfunktion $\rho = \exp(a)$

$$m(\xi) = -\sigma^2/2 + a$$

jævnfør [4][s. 473-474].

6.3.1.1 Homogen log Gaussisk Cox proces med eksponentiel korrelationsfunktion

Først fittes en homogen log Gaussisk Cox proces med eksponentiel korrelation $r(\xi) = \exp(-\|\xi\|)$ til punktmønsteret. Koden i Listing 6.1 fitter en log Gaussisk Cox proces med eksponentiel korrelationsfunktion og konstant intensitet til punktmønsteret på Figur 6.2.

Listing 6.1: Homogen model med eksponentiel korrelation.

```
husfit <- kppm(HussalgAalborg~1, "LGCP", statistic="pcf",
method="clik2")
  if(require("RandomFields")) {
    husfit <- kppm(HussalgAalborg~1, "LGCP", statistic="pcf",
    method="clik2",
                    control=list(maxit=100), algorithm="BFGS")
  }
```

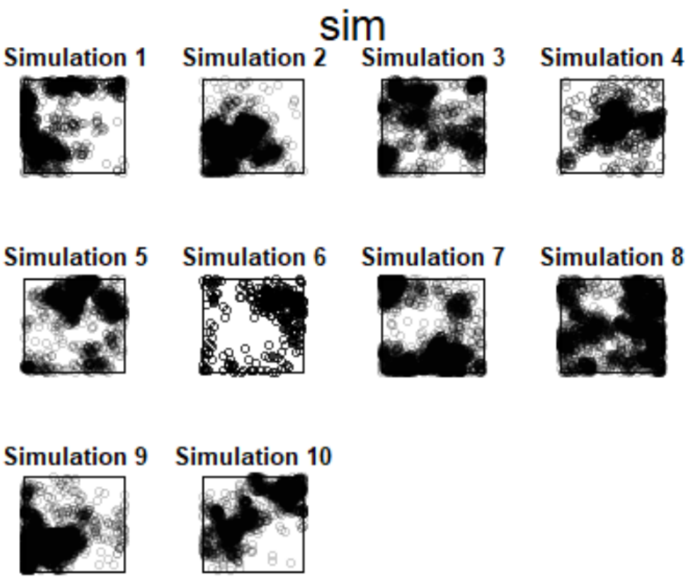
Koden producerer en model for en homogen log Gaussisk Cox proces med følgende parametre.

```
$trend
[1] 1.447179e-06

$var
[1] 6.563197

$scale
[1] 18993.71

$mu
[1] -16.72749
```

Figur 6.6: *Parametre for den fittede log Gaussiske Cox proces med eksponentiel korrelation.*

Parametrene specificerer en log Gaussisk Cox proces med eksponentiel korrelation, hvor kovariansfunktionen er givet som $c(\xi) = 6.563197 \exp(-\|\xi/18993.71\|)$, middelværdien $m = -16.72749$ og intensiteten $\rho = 1.447179 \cdot 10^{-6}$. Det ses, at modellen følger en homogen proces. Dette betyder, at processen har konstant intensitetsfunktion, hvilket leder til en konstant middelværdi for det underliggende Gaussisk stokastiske felt.

Ved anvendelse af `simulate`-funktionen dannes 10 simulationer af modellen givet ved parametrene på Figur 6.6. Simulationerne producerer punktmønstre hvor antallet af punkter varierer fra 406 punkter for Simulation 6 til 10757 punkter for Simulation 2. Dette ses på Figur 6.7.

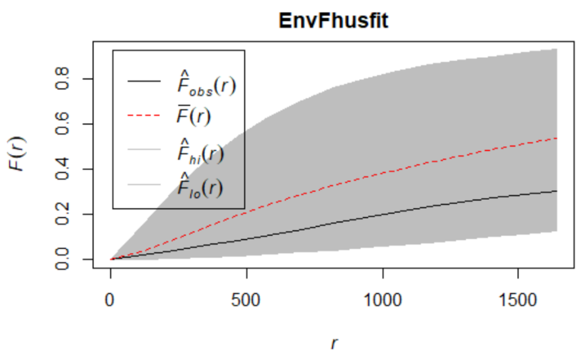


Figur 6.7: *10 simulationer af den fittede homogene log Gaussiske Cox proces med eksponentiel korrelationsfunktion.*

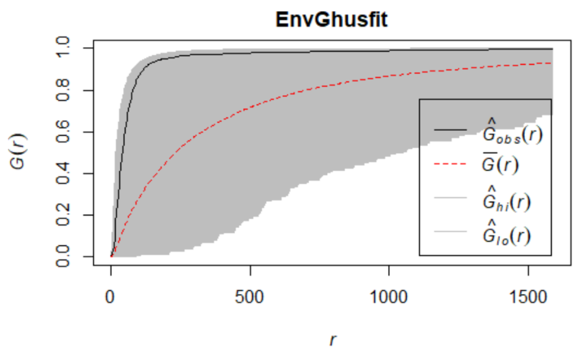
Det ses på Figur 6.7, hvordan de 10 simulationer ligner clusterede punktmønstre ligesom punktmønsteret bestående af hussalg i Aalborg Kommune. Det ses dog, at simulationerne ikke virker til at være clusteret omkring centrum af observationsvinduet i alle tilfælde. Simulationerne adskiller sig på den måde, fra den clustering struktur der er i punktmønsteret for hussalg i Aalborg

Kommune.

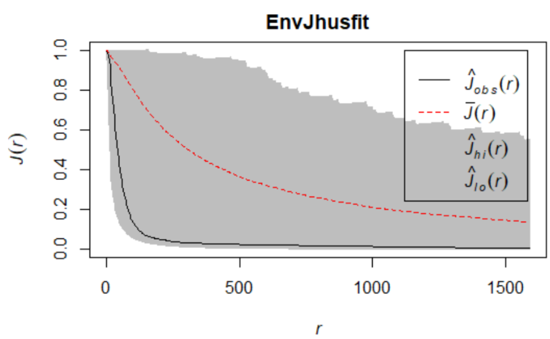
For at afgøre hvor godt den homogene log Gaussiske Cox proces med eksponentiel korrelation passer på data, kan envelopes ud fra summary statistics anvendes. Bemærk at modelfitting funktionen `kppm` har anvendt parkorrelationsfunktionen g og herigennem K-funktionen til at tilpasse modellen. Derfor benyttes i stedet de summary statistics, som er baseret på afstande mellem punkter, givet som F-, G- og J-funktionen, til at eftertjekke modellen. F-, G- og J-funktionen for modellen med tilhørende 2.5% -nedre og 97.5%-øvre envelopes udregnes ved hjælp af `envelope`-funktionen over 39 simulationer. Disse simulationer er under antagelse af en homogen log Gaussisk Cox proces givet ved parametrene i Figur 6.6. Dette giver nulhypotesen: At punktmønsteret for hussalg i Aalborg Kommune følger en homogen log Gaussisk Cox proces specificeret ved parametrene fra Figur 6.6. Bemærk at de nulhypoteser som anvendes fremover i kapitlet, antager en log Gaussisk Cox proces, hvor intensitets- og korrelationsfunktionen er specificeret i det pågældende underafsnit. Plottet for disse envelopes ses på Figur 6.8. Bemærk at `envelope`-funktionen for F, G og J giver både et estimat af den empiriske F-, G- og J-funktion og et gennemsnit af F-, G- og J-funktionen lavet ud fra de 39 simulationer. Gennem hele kapitlet anvendes 39 simulationer til simulere envelopes for de forskellige modeller.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.8: *Plot af F-, G- og J-funktionen for den homogene log Gaussiske Cox proces med eksponentiel korrelation med tilhørende envelopes.*

Som nævnt i Afsnit 3.4.6 kan envelopes af denne type kun anvendes som signifikanstest, hvis afstanden r på forhånd er fastsat. Dermed kan plottene i Figur 6.8 ikke anvendes til at bekræfte, at punktmønsteret bestående af hussalg i Aalborg Kommune følger den log Gaussiske Cox proces defineret ud fra parametrene i Figur 6.6. Dog kan disse envelopes anvendes til at se, for hvilke værdier af r at nulhypotesen vil blive afvist. Ud fra plottet på Figur 6.8a ses det for F-funktionen, at ingen værdier af r ville resultere i, at nulhypotesen afvises, givet at r er valgt på forhånd.
Som for F-funktionen i Figur 6.8a, ses det for G-funktionen på Figur 6.8b, at ingen værdier af r ville resultere i, at nulhypotesen vil blive afvist.
Ud fra Figur 6.8c ses, at den empiriske J-funktion er indenfor de tilhørende envelopes for J-funktionen for alle værdier af r. Dermed vil ingen af disse værdier forårsage at nulhypotesen afvises på baggrund af envelopes for J-funktionen, givet r er valgt på forhånd.

Overordnet set viser disse tre summary statistics og deres tilhørende envelopes, at den homogene log Gaussiske Cox proces med eksponentiel korrelationsfunktion passer godt på punktmønsteret, som repræsenterer hussalg i Aalborg kommune. Løst sagt er fittet af den log Gaussiske Cox proces 'passende' på data.

6.3.1.2 Homogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion

I dette underafsnit ændres korrelationsfunktionen til en Gaussisk på formen

$r(\xi) = \exp(-\|\xi/\alpha\|^2)$. Koden i Listing 6.2 fitter en homogen log Gaussisk Cox proces med Gaussisk korrelation til punktmønsteret over hussalg i Aalborg Kommune.

Listing 6.2: Homogen model med Gaussisk korrelation.

```
husfit <- kppm(HussalgAalborg~1, "LGCP", statistic="pcf",
method="clik2")
  if(require("RandomFields")) {
    husfit <- kppm(HussalgAalborg~1, "LGCP", statistic="pcf",
    model="gauss", method="clik2",
                    control=list(maxit=100), algorithm="BFGS")
  }
```

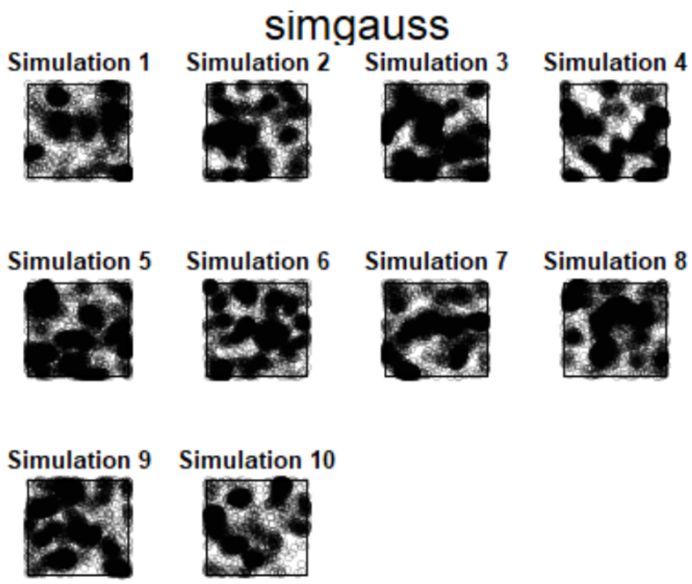
Modellen følger en homogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion, hvis parametre er givet på Figur 6.9.

```
$trend
[1] 1.447179e-06

$var
[1] 2.484551

$scale
[1] 6564.21

$mu
[1] -14.68817
```

Figur 6.9: *Parametre for den fittede log Gaussiske Cox proces med Gaussisk korrelation.*

Parametrene i Figur 6.9 specificerer en homogen log Gaussisk Cox proces med kovariansfunktion givet som $c(\xi) = 2.484551 \exp(-\|\xi/6564.21\|^2)$, middelværdi $m = -16.72749$ og intensitet $\rho = 1.447179 \cdot 10^{-6}$.
Bemærk, at intensiteten for modellen er identisk med intensiteten for den homogene log Gaussiske Cox proces med eksponentiel korrelationsfunktion. Dette skyldes, at estimatet for ρ i det homogene tilfælde givet i ligning (3.19) ikke afhænger af kovariansen, men kun af antallet af punkter og størrelsen på observationsvinduet.

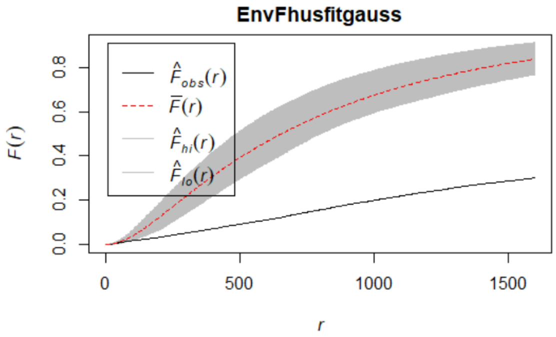
Modellen givet ud fra parametrene i Figur 6.9 simuleres 10 gange, hvilket ses på Figur 6.10.



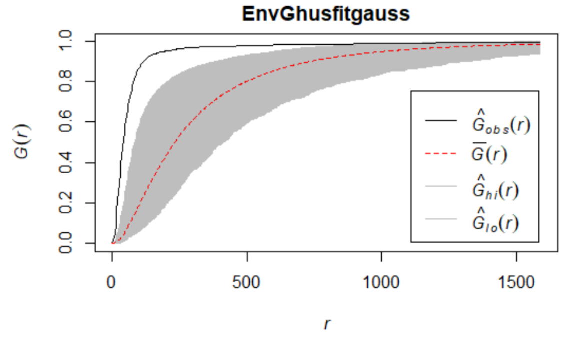
Figur 6.10: *10 simulationer af den fittede homogene log Gaussiske Cox proces med Gaussisk korrelationsfunktion.*

Simulationerne producerer punktmønstre med et antal af punkter, som varierer fra 3332 punkter i Simulation 1 til 7167 punkter i Simulation 6. Simulationerne udviser clustering, men som i Afsnit 6.3.1.1 er punkterne ikke kun clusterede omkring centrum af observationsvinduet.

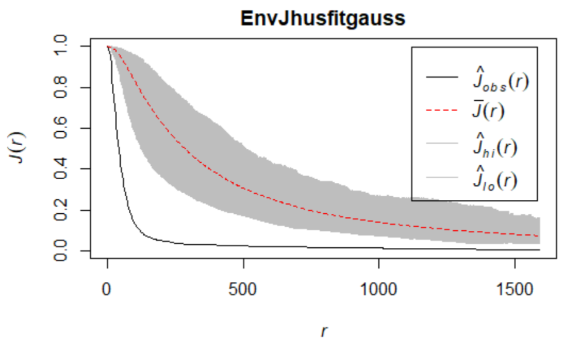
Herefter anvendes envelopes for summary statistics til at undersøge, hvorvidt modellen givet ud fra parametrene i Figur 6.9 passer på data. Envelopes for F-, G- og J-funktionen under antagelse af modellen ses på Figur 6.11.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.11: *Plot af F-, G- og J-funktionen for den homogene log Gaussiske Cox proces med Gaussisk korrelation med tilhørende envelopes.*

På Figur 6.11 ses det, at de empiriske F-,G- og J-funktioner er udenfor deres tilhørende envelopes for de fleste værdier af r. Af denne grund vil de fleste valg af r resultere i, at nulhypotesen vil blive afvist med signifikansniveau 0.05, givet at r vælges på forhånd. Dermed indikerer disse summary statistics og deres tilhørende envelopes, at en homogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion ikke fitter punktmønsteret specielt godt.

6.3.1.3 Homogen log Gaussisk Cox proces med stabil korrelationsfunktion

I dette underafsnit ændres korrelationsfunktionen nu til den stabile korrelationsfunktion $r(\xi) = \exp(-||\xi||^{1/2})$. I Listing 6.3 ses koden for at fitte en homogen log Gaussisk Cox proces med stabil korrelation til punktmønsteret over hussalg i Aalborg Kommune. I Listing 6.3 er korrelationsfunktionen specificeret med en α-værdi, svarende til δ i ligning (4.2), på 0.5 som angiver at korrelationsfunktionen er stabil.

Listing 6.3: Homogen model med stabil korrelation.

```
husfit <- kppm(HussalgAalborg~1, "LGCP", statistic="pcf",
```

```
method="clik2", covmodel=list(model="stable", alpha=0.5))
  if(require("RandomFields")) {
    husfit <- kppm(HussalgAalborg~1, "LGCP", statistic="pcf",
    covmodel=list(model="stable", alpha=0.5), method="clik2",
                       control=list(maxit=100), algorithm="BFGS")
  }
```

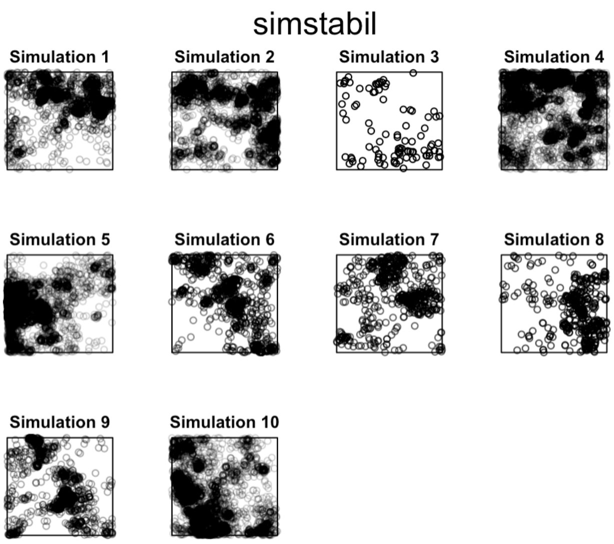
Parametrene for modellen i Listing 6.3 ses på Figur 6.12.

```
$trend
[1] 1.447179e-06

$var
[1] 6.836904

$scale
[1] 23666.05

$mu
[1] -16.86435
```

Figur 6.12: *Parametre for den fittede log Gaussiske Cox proces med stabil korrelation.*

Kovariansfunktionen for modellen er givet som $c(\xi) = 6.836904 \exp(\|\xi/23666.05\|^{1/2})$, middelværdien er givet som $m = -16.86435$ og intensiteten er givet som $\rho = 1.447179 \cdot 10^{-6}$. Det ses, at intensiteten for modellen er den samme, som den var for de homogene log Gaussiske Cox processer med eksponentiel og Gaussisk korrelationsfunktion.

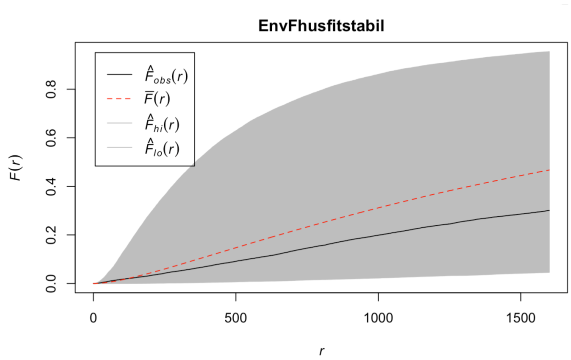
På Figur 6.13 er 10 simulationer af den fittede log Gaussiske Cox proces med stabil korrelation vist.



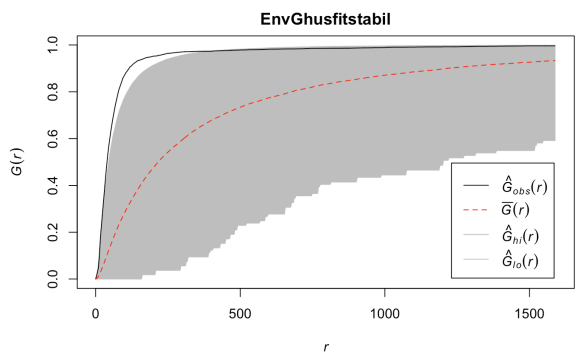
Figur 6.13: *10 simulationer af den fittede homogene log Gaussiske Cox proces med stabil korrelationsfunktion.*

Simulationerne viser punktmønstre, hvor antallet af punkter varierer mellem 96 punkter i Simulation 3 og 12186 punkter i Simulation 5. De 10 simulationer udviser, ligesom for de to andre korrelationstyper, clustertendenser. Desuden ses, at punktmønstrene på Figur 6.13 ikke kun er clusterede omkring centrum af observationsvinduet. Ligesom for den eksponentielle og Gaussiske model, afviger dette fra strukturen i punktmønsteret for hussalg i Aalborg Kommune.

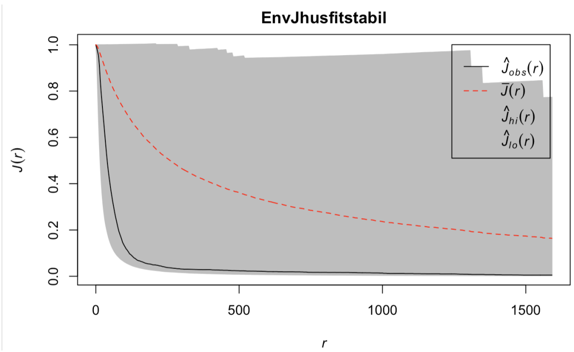
For at vurdere hvorvidt modellen specificeret ved parametrene på Figur 6.12 er passende for data, anvendes som tidligere de tre summary statistics F-, G- og J-funktionen. Hertil benyttes `envelope` kommandoen, som udregner F-, G- og J-funktionen med tilhørende 2.5%-nedre og 97.5%-øvre envelopes. Dette kan ses på Figur 6.14.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.14: *Plot af F-, G- og J-funktionen for den homogene log Gaussiske Cox proces med stabil korrelation med tilhørende envelopes.*

Det ses på Figur 6.14a og 6.14c at for enhver værdi af r ligger F- og J-funktionen indenfor de simulerede envelopes. Det betyder, at ingen værdier af r ville resultere i at nulhypotesen afvises, givet at r er valgt på forhånd.

Det ses på Figur 6.14b, at G-funktionen ikke er indenfor dens tilhørende envelopes for enhver værdi af r. Dermed vil nogle valg af r medføre at nulhypotesen afvises med signifikansniveau 0.05, givet at r er valgt på forhånd.

Ud fra F- og J-funktionen passer modellen godt på data, da funktionerne begge ligger inden for envelopes for alle værdier af r. Dog indikerer G-funktionen, at modellen ikke er specielt passende på data for bestemte værdier af r.

6.3.2 Inhomogen log Gaussisk Cox proces med lineær intensitet

Efter at have fittet tre forskellige homogene log Gaussiske Cox processer til punktmønsteret for Hussalg i Aalborg Kommune er det set, hvordan modellen med eksponentiel korrelationsfunktion passer godt på datasættet. Samtidig ses, at processen med Gaussisk korrelation, ikke fitter data specielt godt. Modellen med stabil korrelation er passende for bestemte værdier af r.

Det kan ses, hvordan en stor del af observationsvinduet på Figur 6.2 er uden punkter, som kan tyde på inhomogenitet. Et andet tegn på inhomogenitet ses ved estimatet af intensiteten på Figur 6.3, som har varierende værdier over observationsvinduet. Dermed kan intensiteten for den log Gaussiske Cox proces tilpasses. I dette delafsnit gives log intensiteten på en første ordens polynomisk form. Dette specificeres gennem første indgang i `kppm` ved $\sim x + y$, som giver formen af log intensiteten.

Som for de homogene log Gaussiske Cox processer fittes nu tre inhomogene log Gaussiske Cox processer med eksponentiel, Gaussisk og stabil korrelationsfunktion. Desuden undersøges gennem envelopes, hvor godt disse modeller fitter punktmønsteret for hussalg i Aalborg Kommune. Som afslutning på delafsnittet sammenlignes og undersøges middelværdifunktionerne for de tilhørende Gaussisk stokastiske felter.

6.3.2.1 Inhomogen log Gaussisk Cox proces med eksponentiel korrelationsfunktion og lineær intensitet

Nu fittes en inhomogen log Gaussisk Cox Proces med eksponentiel korrelationsfunktion $r(\xi) = \exp(-\|\xi/\alpha\|)$ til punktmønsteret. Listing 6.4 indeholder koden som fitter den inhomogene log Gaussiske Cox proces til punktmønsteret.

Listing 6.4: Inhomogen model med eksponentiel korrelation.

```
husfit2 <- kppm(HussalgAalborg~x+y, "LGCP", statistic="pcf",
method="clik2")
  if(require("RandomFields")) {
    husfit2 <- kppm(HussalgAalborg~x+y, "LGCP",
    statistic="pcf", method="clik2",
                             control=list(maxit=100), algorithm="BFGS")
```

Dette giver en inhomogen log Gaussisk Cox proces med følgende parametre.

```
$trend
  (Intercept)             x             y
-6.792226e+01  1.427365e-05  7.364943e-06

$var
[1] 6.461898

$scale
[1] 18822.19
```

Figur 6.15: *Parametre for den fittede inhomogene log Gaussiske Cox proces med eksponentiel korrelation.*

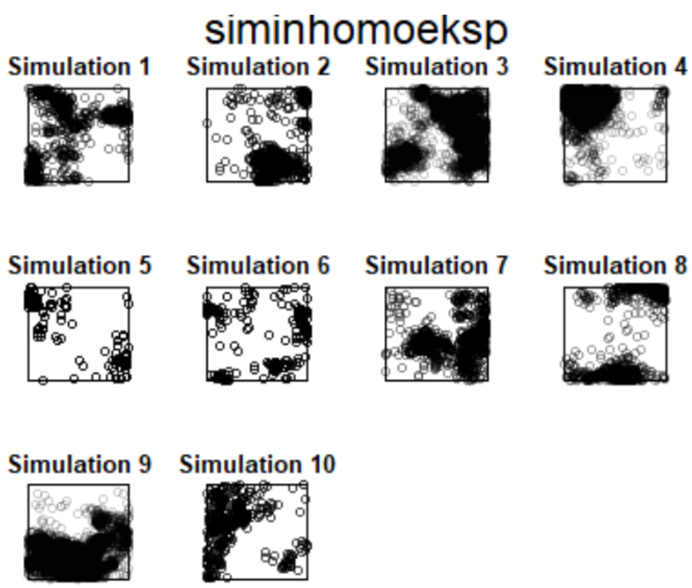
Disse parametre specificerer en inhomogen log Gaussisk Cox proces med eksponentiel korrelationsfunktion. Den tilhørende kovariansfunktion er givet som

$$c(\xi) = 6.461898 \exp(-\|\xi/18822.19\|),$$

og intensitetsfunktionen er

$$\rho(\xi) = 1.427365 \cdot 10^{-5}x + 7.364943 \cdot 10^{-6}y - 67.92226.$$

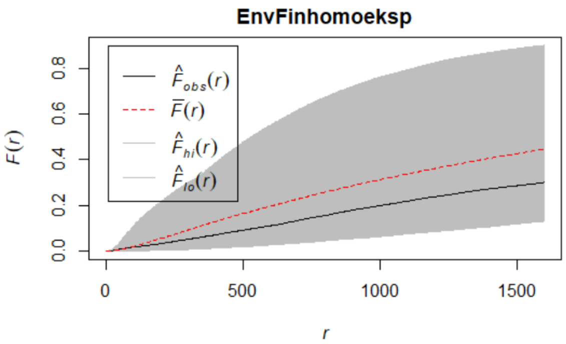
Gennem `simulate`-funktionen kan den inhomogene log Gaussiske Cox proces givet ved parametrene i Figur 6.15 simuleres. 10 simulationer af modellen kan ses på Figur 6.16.



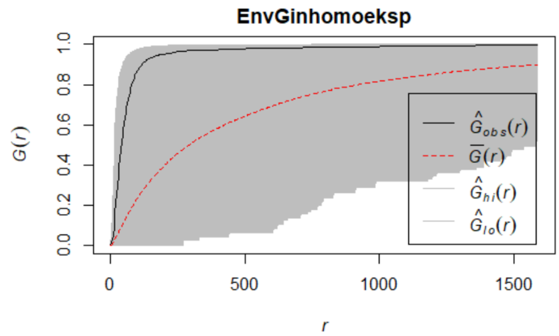
Figur 6.16: *10 simulationer af den inhomogene log Gaussiske Cox proces med eksponentiel korrelation.*

Disse 10 simulationer producerer punktmønstre indeholdende mellem 161 punkter i Simulation 5 til 22722 punkter i Simulation 4. De simulerede punktprocesser fra Figur 6.16 udviser ikke clustertendenser udelukkende omkring centrum af observationsvinduet, som er tilfældet for punktmønsteret bestående af hussalg i Aalborg Kommune.

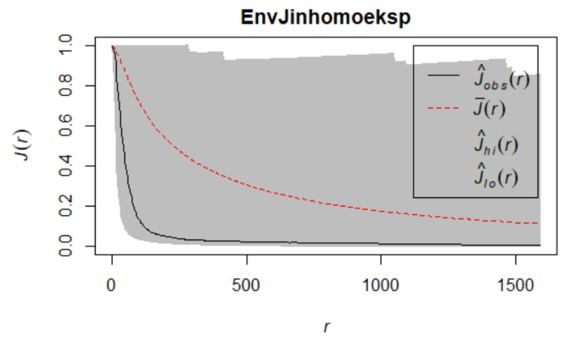
Som tidligere undersøges gennem envelopes ud fra F-, G- og J-funktionen, om modellen fitter punktmønsteret for hussalg i Aalborg Kommune godt. Igen anvendes `envelope`-funktionen til at udregne F, G og J for punktmønsteret for hussalg i Aalborg Kommune samt 2.5%-nedre og 97.5%-øvre envelopes. Et plot for dette ses på Figur 6.17.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.17: *Plot af F-, G- og J-funktionen for den inhomogene log Gaussiske Cox proces med eksponentiel korrelation med tilhørende envelopes.*

På Figur 6.17 ses, at både F-,G- og J-funktionen udregnet for punktmønsteret for hussalg i Aalborg Kommune ligger indenfor de tilhørende envelopes. Da vil intet valg af r resultere i at nulhypotesen bliver afvist, givet at r er valgt på forhånd. Dermed er også den inhomogene log Gaussiske Cox proces specificeret ud fra værdierne på Figur 6.15 'et godt fit' for datasættet.

6.3.2.2 Inhomogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion og lineær intensitet

Herefter fittes en inhomogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion $r(\xi) = \exp(-\|\xi/\alpha\|^2)$ til punktmønsteret. I Listing 6.5 ses koden for at fitte en inhomogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion til punktmønsteret for hussalg i Aalborg Kommune.

Listing 6.5: Inhomogen model med Gaussisk korrelation.

```
husfit3 <- kppm(HussalgAalborg~x+y, "LGCP", statistic="pcf",
```

```
method="clik2")
  if(require("RandomFields")) {
    husfit2 <- kppm(HussalgAalborg~x+y, "LGCP",
    statistic="pcf", model="gauss", method="clik2",
                         control=list(maxit=100), algorithm="BFGS")
```

Den fittede log Gaussiske Cox proces har parametrene set på Figur 6.18.

```
$trend
  (Intercept)             x             y
-6.792226e+01  1.427365e-05  7.364943e-06

$var
[1] 2.522723

$scale
[1] 6717.566
```

Figur 6.18: *Parametre for den fittede inhomogene log Gaussiske Cox proces med Gaussisk korrelation.*

Disse parametre specificerer en inhomogen log Gaussisk Cox proces med Gaussisk korrelation. Kovariansen er givet som

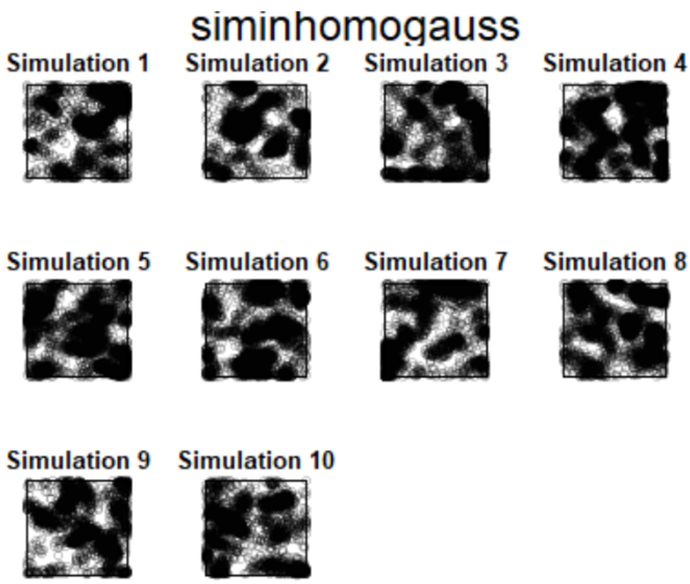
$$c(\xi) = 2.522723 \exp(-\|\xi/6717.566\|^2),$$

og intensiteten er givet som

$$\rho(\xi) = 1.427365 \cdot 10^{-5}x + 7.364943 \cdot 10^{-6}y - 67.922226.$$

Bemærk, at intensitetsfunktionen for denne model er lig intensitetsfunktionen for modellen med eksponentiel korrelation fra Afsnit 6.3.2.1.

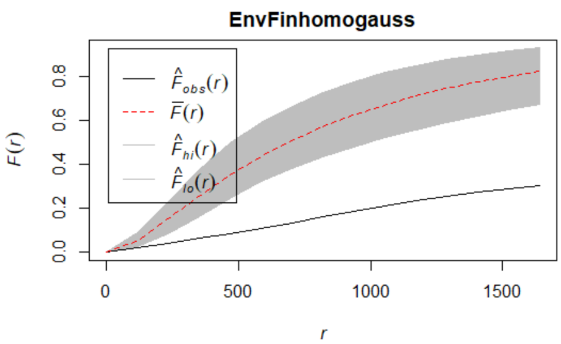
På Figur 6.19 ses 10 simulationer af den inhomogene log Gaussiske Cox proces givet ved parametrene på Figur 6.18. Disse simulationer giver punktmønstre, hvor antallet af punkter er mellem 2990 punkter for Simulation 1 og 9667 punkter for Simulation 3.



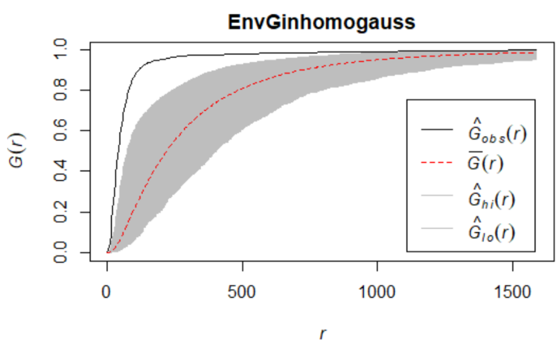
Figur 6.19: *10 simulationer af den inhomogene log Gaussiske Cox proces med Gaussisk korrelation.*

Igen bemærkes, hvordan de simulerede punktmønstre udviser clustering, som ikke kun forekommer omkring centrum af observationsvinduet.

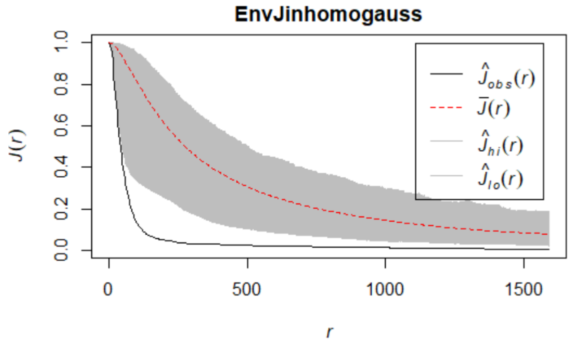
Ved hjælp af envelopes for F-, G- og J-funktionen undersøges hvor godt modellen givet ud fra parametrene i Figur 6.18 fitter punktmønsteret. De pågældende envelopes dannes ud fra simulationer under antagelse af modellen specificeret ved disse parametre. F-, G- og J-funktionen med tilhørende envelopes ses på Figur 6.20.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.20: *Plot af F-, G- og J-funktionen for den inhomogene log Gaussiske Cox proces med Gaussisk korrelation med tilhørende envelopes.*

På Figur 6.20a og 6.20c ses det, at F- og J-funktionen for puntktmønsteret ligger uden for de tilhørende envelopes undtagen for tilpas små værdier af r. Det vil sige, at for tilpas store værdier af r ville nulhypotesen afvises med signifikansniveau 0.05, givet at r er valgt på forhånd. Ud fra Figur 6.20b ses det, at G-funktionen er udenfor envelopes for alle værdier af r. Dette ville resultere i at nulhypotesen afvises med signifikansniveau 0.05 for ethvert r, hvis r er valgt på forhånd.

Dermed passer den inhomogene log Gaussiske Cox proces med Gaussisk korrelation ikke godt på data.

6.3.2.3 Inhomogen log Gaussisk Cox proces med stabil korrelationsfunktion og lineær intensitet

I dette afsnit fittes en log Gaussisk Cox proces med stabil korrelationsfunktion $r(\xi) = \exp(-\|\xi/\alpha\|^{1/2})$ til punktmønsteret for Hussalg i Aalborg Kommune. Koden for dette ses i følgende listing.

Listing 6.6: Inhomogen model med stabil korrelation.

```
husfit <- kppm(HussalgAalborg~x+y, "LGCP", statistic="pcf",
method="clik2", covmodel=list(model="stable", alpha=0.5))
  if(require("RandomFields")) {
    husfit <- kppm(HussalgAalborg~x+y, "LGCP", statistic="pcf",
    method="clik2",control=list(maxit=100), algorithm="BFGS",
                         covmodel=list(model="stable", alpha=0.5))}
```

Denne log Gaussiske Cox proces er da specificeret ved parametrene angivet på Figur 6.21.

```
$trend
  (Intercept)             x             y
-6.792226e+01  1.427365e-05  7.364943e-06

$var
[1] 6.499979

$scale
[1] 21024.93
```

Figur 6.21: *Parametre for den fittede inhomogene log Gaussiske Cox proces med stabil korrelation.*

Dette giver en log Gaussisk Cox proces med kovariansfunktion

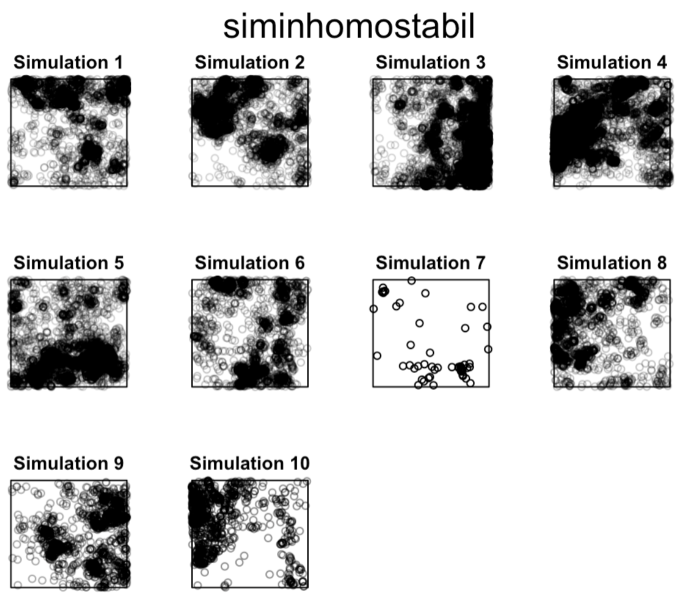
$$c(\xi) = 6.499979 \exp(-\|\xi/21024.93\|^{1/2})$$

og intensitetsfunktion

$$\rho(\xi) = 1.427365 \cdot 10^{-5}x + 7.364943 \cdot 10^{-6}y - 67.922226.$$

Det ses, at intensitetsfunktionen igen er den samme som for den eksponentielle og Gaussiske model med lineær intensitet.

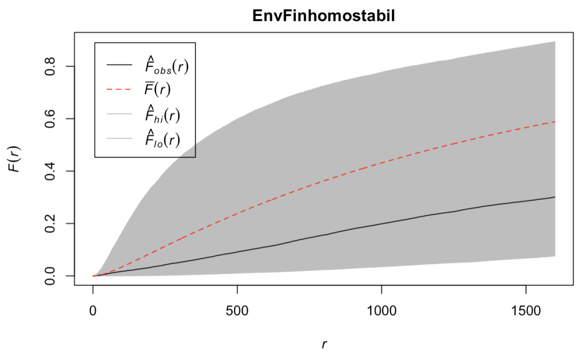
Gennem `simulate`-funktionen laves 10 simulationer af den inhomogene log Gaussiske Cox proces specificeret ud fra værdierne på Figur 6.21. Simulationerne ses på Figur 6.22.



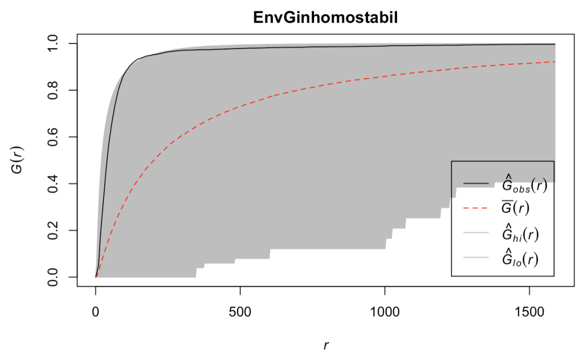
Figur 6.22: *10 simulationer af den inhomogene log Gaussiske Cox proces med stabil korrelation.*

De 10 simulationer er punktmønstre, hvor antallet af punkter går fra 58 for Simulation 7 til 18759 for Simulation 4. Ligesom i de andre tilfælde udviser punktmønstrene clustering, men er ikke kun clusterede omkring centrum af observationsvinduet.

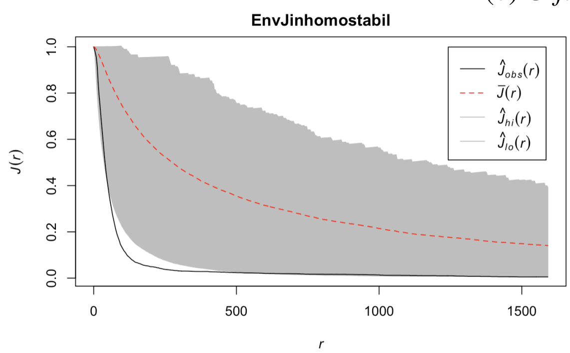
Til sidst undersøges hvor godt den fittede model passer på punktmønsteret gennem envelopes. F-, G- og J-funktionen udregnes med tilhørende 2.5%-nedre og 97.5%-øvre envelopes. Dette ses på Figur 6.23.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.23: *Plot af F-, G- og J-funktionen for den inhomogene log Gaussiske Cox proces med stabil korrelation med tilhørende envelopes.*

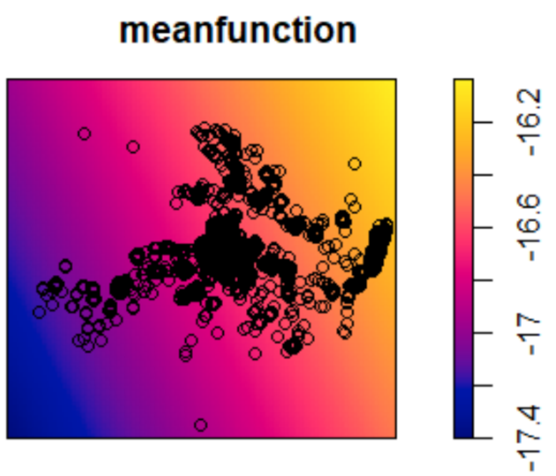
Det ses, at F- og G-funktionen er inden for deres tilhørende envelopes for alle værdier af r. Dette betyder, at nulhypotesen ikke vil blive afvist for nogle værdier af r, givet at r er valgt på forhånd.

På Figur 6.23c ses, at J-funktionen for datasættet er udenfor de tilhørende envelopes for bestemte værdier af r. Dermed vil valget af disse r-værdier resultere i at nulhypotesen afvises med signifikansniveau på 0.05, givet at valget af r foretages før envelopes udregnes. Dette betyder, at den log Gaussiske Cox proces med stabil korrelation og lineær intensitet passer godt på data for bestemte værdier af r.

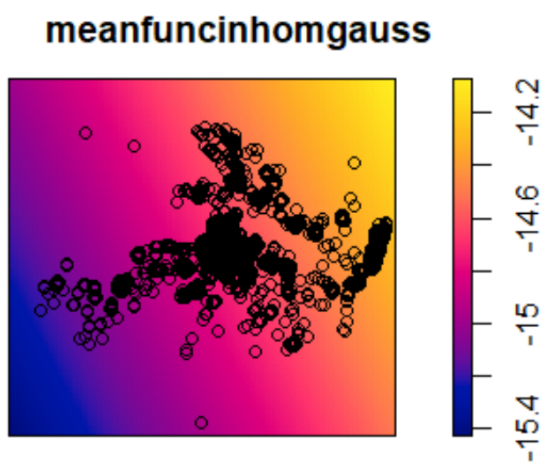
6.3.2.4 Middelværdifunktioner for modellerne med lineær intensitet

I dette delafsnit undersøges middelværdifunktionerne for de Gaussisk stokastiske felter tilhørende de log Gaussiske Cox processer med lineær intensitet. Middelværdien for hver af de

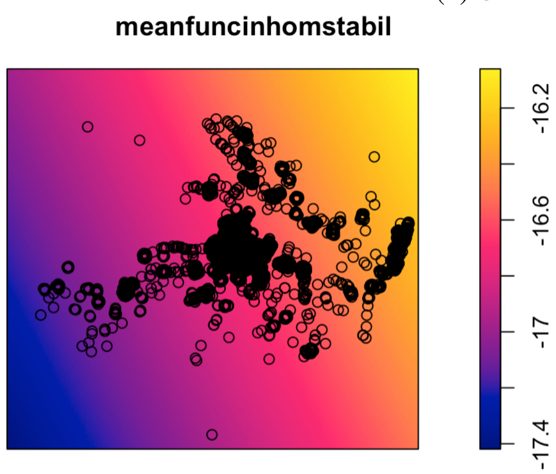
Gaussisk stokastiske felter er givet som et `pixel image`. Disse ses afbildet på Figur 6.24 sammen med punktmønsteret for hussalg i Aalborg Kommune.



(a) *Eksponentiel korrelation.*



(b) *Gaussisk korrelation.*



(c) *Stabil korrelation.*

Figur 6.24: *Plot af middelværdifunktionerne for den log Gaussiske Cox proces med lineær intensitet.*

Først bemærkes, at middelværdifunktionerne på Figur 6.24 stort set er ens. Dette skal forstås på den måde, at deres struktur, i forhold til i hvilken retning de vokser, stort set er identisk. Dog varierer middelværdien for den Gaussiske model fra middelværdien for den eksponentielle og stabile model, med en værdi på 2. Det ses på Figur 6.24a-6.24c, at middelværdien vokser mod øverste højre hjørne af observationsvinduet. Dette stemmer ikke overens med, hvordan intensiteten for punktmønsteret for hussalg i Aalborg Kommune på Figur 6.3 ser ud. Det ses desuden, at middelværdifunktionerne for alle tre modeller varierer med 1.2 fra den højeste til den laveste værdi, for hver middelværdifunktion. Givet størrelsesordenen af observationsvinduet betyder dette, at middelværdifunktionen stort set ikke ændrer sig. Det vil sige, at modellerne ikke opfører sig specielt inhomogent.

6.3.3 Inhomogen log Gaussisk Cox proces med 3. grads polynomisk intensitet

For de inhomogene modeller med lineær intensitet, viste det sig at middelværdifunktionerne set på Figur 6.24a, 6.24b og 6.24c, ikke stemmer overens med intensiteten for datasættet set på Figur 6.3. For at forsøge at imødekomme dette problem, anvendes en anden intensitetsfunktion for modellen. For at finde en passende intensitetsfunktion er forskellige funktioner blevet afprøvet. Ud fra envelopes er (log) intensitetsfunktionen blevet valgt til $x^3 + y^3 + x^2 + y^2$.
Som i Afsnit 6.3.1 og 6.3.2 fittes tre inhomogene log Gaussiske Cox processer med henholdsvis eksponentiel, Gaussisk og stabil korrelation. Som tidligere anvendes envelopes til at vurdere, hvor godt de individuelle modeller passer på punktmønsteret over hussalg i Aalborg Kommune. Formen af log intensiteten speciferes gennem første indgang i `kppm` ved $\sim x^3 + y^3 + x^2 + y^2$. Som afslutning på delafsnittet undersøges middelværdifunktionerne for de tilhørende Gaussisk stokastiske felter.

6.3.3.1 Inhomogen log Gaussisk Cox proces med eksponentiel korrelationsfunktion og 3. grads intensitet

I dette afsnit fittes en log Gaussisk Cox proces med eksponentiel korrelationsfunktion $r(\xi) = \exp(-\|\xi/\alpha\|)$ til punktmønsteret. Koden for dette ses i Listing 6.7.

Listing 6.7: Inhomogen model med eksponentiel korrelation.

```
husfit5 <- kppm(HussalgAalborg~I(x^3)+I(y^3)+I(x^2)+I(y^2), "LGCP",
statistic="pcf", method="clik2")
  if(require("RandomFields")) {
    husfit5 <- kppm(HussalgAalborg~I(x^3)+I(y^3)+I(x^2)+I(y^2),
    "LGCP", statistic="pcf", method="clik2",
                              control=list(maxit=100), algorithm="BFGS")}
```

Den log Gaussiske Cox proces har parametrene set på Figur 6.25.

```
$trend
  (Intercept)         I(x^3)         I(x^2)         I(y^3)         I(y^2)
-2.229980e+05 -7.494374e-15  6.268562e-09 -1.759690e-15  1.668769e-08

$var
[1] 5.582861

$scale
[1] 17168.65
```

Figur 6.25: *Parametre for den fittede inhomogene log Gaussiske Cox proces med eksponentiel korrelation.*

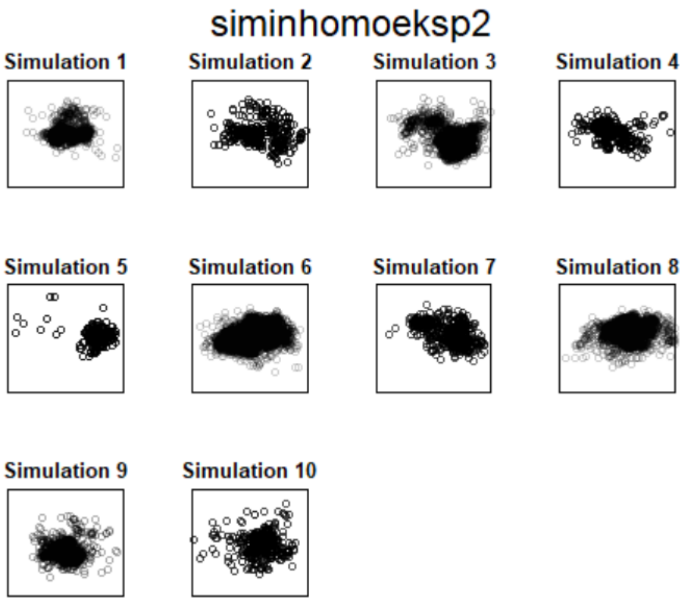
Disse parametre specificerer en log Gaussisk Cox proces med kovariansfunktion

$$c(\xi) = 5.582861 \exp(-\|\xi/171768.65\|)$$

og intensitet

$$\begin{aligned}\rho(\xi) = &-7.494374 \cdot 10^{-15}x^3 - 1.759690 \cdot 10^{-15}y^3 + 6.268562 \cdot 10^{-9}x^2 \\ &+ 1.668769 \cdot 10^{-8}y^2 - 2.229980 \cdot 10^5.\end{aligned}$$

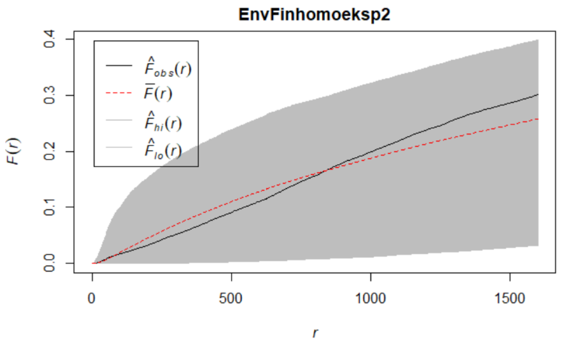
På Figur 6.26 ses 10 simulationer af den inhomogene log Gaussiske Cox proces givet ved parametrene i 6.25.



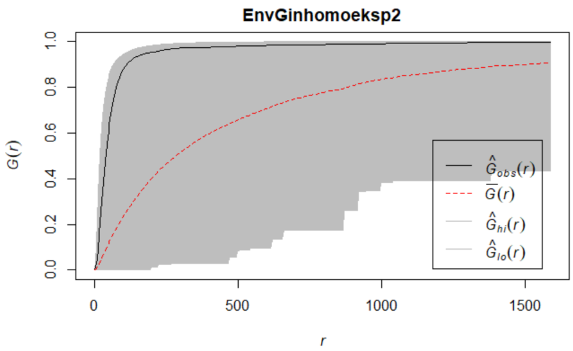
Figur 6.26: *10 simulationer af den inhomogene log Gaussiske Cox proces med eksponentiel korrelation.*

Simulationerne er punktmønstre, hvor antallet af punkter ligger mellem 229 for Simulation 5 og 40521 for Simulation 6. Bemærk at disse simulationer udviser clustertendenser, sådan at antallet af punkter er størst omkring centrum af observationsvinduet. Disse simulationer minder dermed om punktmønsteret for hussalg i Aalborg Kommune, som ses på Figur 6.2.

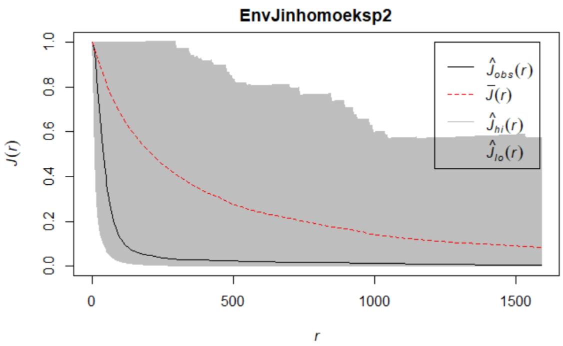
Herefter anvendes envelopes for F-, G- og J-funktionen til at undersøge hvor godt modellen specificeret ved parametrene på Figur 6.25 passer på data. Disse envelopes ses på Figur 6.27.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.27: *Plot af F-, G- og J-funktionen for den inhomogene log Gaussiske Cox proces med eksponentiel korrelation med tilhørende envelopes.*

Som for de andre modeller med eksponentiel korrelationsfunktion, ses det på Figur 6.27, hvordan F-,G- og J-funktionen er fuldstændig indeholdt i deres tilhørende envelopes. Dermed vil intet valg af r resultere i at nulhypotesen bliver afvist, givet at r er valgt på forhånd. Dette indikerer, at denne model også er 'et godt fit' på data.

6.3.3.2 Inhomogen log Gaussisk Cox proces med Gaussisk korrelationsfunktion og 3. grads intensitet

Nu tilpasses en log Gaussisk Cox proces med Gaussisk korrelationsfunktion $r(\xi) = \exp(-\|\xi/\alpha\|^2)$ til punktmønsteret for hussalg i Aalborg Kommune. Listing 6.8 angiver koden til at fitte en model af denne type til data.

Listing 6.8: Inhomogen model med Gaussisk korrelation.

```
husfit5 <- kppm(HussalgAalborg~I(x^3)+I(y^3)+I(x^2)+I(y^2), "LGCP",
statistic="pcf", model="gauss", method="clik2")
 if(require("RandomFields")) {
   husfit5 <- kppm(HussalgAalborg~I(x^3)+I(y^3)+I(x^2)+I(y^2),
   "LGCP", statistic="pcf", model="gauss", method="clik2",
                             control=list(maxit=100), algorithm="BFGS")}
```

Den log Gaussiske Cox proces, der fås ved at køre koden fra Listing 6.8, er specificeret ud fra parametrene på Figur 6.28.

```
$trend
  (Intercept)        I(x^3)        I(x^2)        I(y^3)        I(y^2)
-2.229980e+05 -7.494374e-15  6.268562e-09 -1.759690e-15  1.668769e-08

$var
[1] 2.044861

$scale
[1] 5924.249
```

Figur 6.28: *Parametre for den fittede inhomogene log Gaussiske Cox proces med Gaussisk korrelation.*

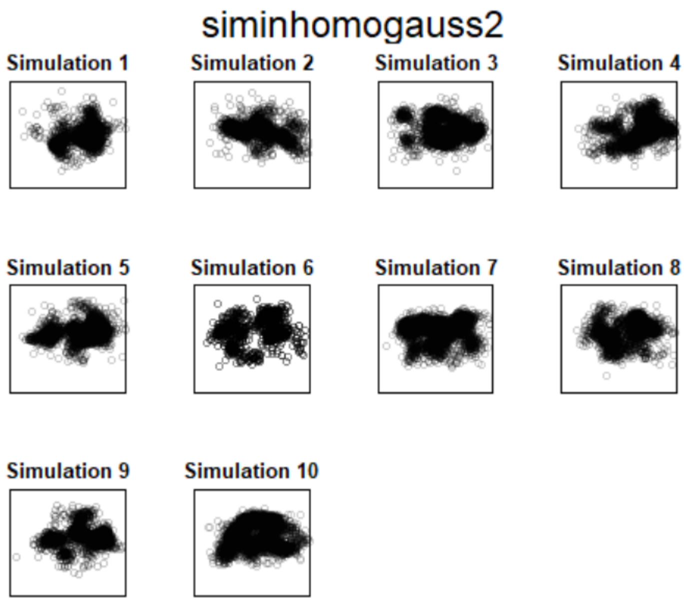
Parametrene giver en log Gaussisk Cox proces med kovariansfunktion

$$c(\xi) = 2.044861 \exp(-\|\xi/5924.249\|)$$

og intensitet

$$\begin{aligned}\rho(\xi) = &-7.494374 \cdot 10^{-15}x^3 - 1.759690 \cdot 10^{-15}y^3 + 6.268562 \cdot 10^{-9}x^2 \\ &+ 1.668769 \cdot 10^{-8}y^2 - 2.229980 \cdot 10^5.\end{aligned}$$

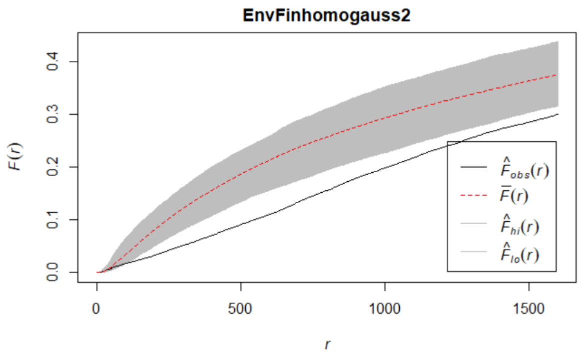
Ved anvendelse af `simulate`-funktionen laves 10 simulationer af modellen specificeret ved parametrene i Figur 6.28. Et plot af disse simulationer ses på Figur 6.29.



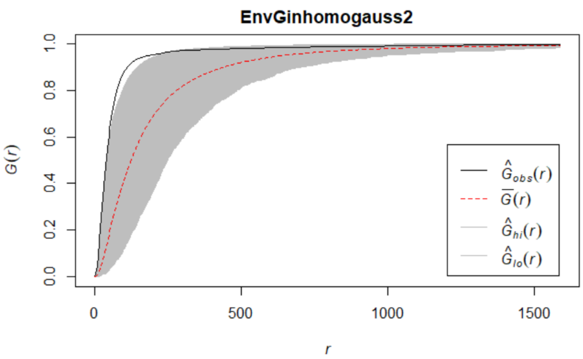
Figur 6.29: *10 simulationer af den inhomogene log Gaussiske Cox proces med Gaussisk korrelation.*

Simulationerne giver punktmønstre, hvor antallet af punkter varierer mellem 1082 punkter for Simulation 6 og 7021 punkter for Simulation 7. Disse punktmønstre udviser også clustertendenser, hvor punkterne centreres omkring midten af observationsvinduet. Dette afspejler strukturen i punktmønsteret for hussalg i Aalborg Kommune.

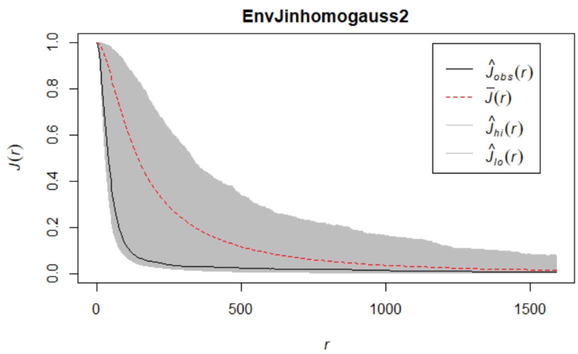
Via envelopes for F-, G- og J-funktionen undersøges det, hvor godt modellen passer på punktmønsteret for hussalg i Aalborg Kommune. F-, G- og J-funktionen med tilhørende 2.5%-nedre og 97.5%-øvre envelopes udregnes. Et plot for dette ses på Figur 6.30.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.30: *Plot af F-, G- og J-funktionen for den inhomogene log Gaussiske Cox proces med Gaussisk korrelation med tilhørende envelopes.*

På Figur 6.30a ses, at F-funktionen er uden for de tilhørende envelopes undtagen for små værdier af r. På baggrund af dette vil nulhyposen blive afvist for de fleste værdier af r, med et signifikansniveau på 0.05, givet at r er valgt på forhånd. Dog ses det på Figur 6.30b, at G-funktionen er udenfor de tilhørende envelopes for få værdier af r. Dermed vil nulhypotesen, i dette tilfælde, kun blive afvist for få værdier af r, givet at r er valgt på forhånd. På Figur 6.30c ses at J-funktionen er indenfor de tilhørende envelopes for alle værdier af r. Nulhypotesen vil dermed ikke blive afvist for nogle værdier af r, givet r er valgt på forhånd. Envelopes for F-funktionen viser, at modellen ikke fitter data specielt godt, hvorimod envelopes for G- og J-funktionen indikerer, at modellen er 'passende' for data for bestemte værdier af r.

6.3.3.3 Inhomogen log Gaussisk Cox proces med stabil korrelationsfunktion og 3. grads intensitet

I dette afsnit fittes en log Gaussisk Cox proces med stabil korrelationsfunktion $r(\xi) = \exp(-\|\xi/\alpha\|^{1/2})$ til punktmønsteret. Koden for dette kan ses i Listing 6.9.

Listing 6.9: Inhomogen model med stabil korrelation.

```
husfit5 <- kppm(HussalgAalborg~I(x^3)+I(y^3)+I(x^2)+I(y^2), "LGCP",
statistic="pcf", method="clik2",
covmodel=list(model="stable", alpha=0.5))
  if(require("RandomFields")) {
    husfit5 <- kppm(HussalgAalborg~I(x^3)+I(y^3)+I(x^2)+I(y^2),
    "LGCP", statistic="pcf", model="gauss", method="clik2",
    covmodel=list(model="stable", alpha=0.5)
                        control=list(maxit=100), algorithm="BFGS")}
```

Dette giver en log Gaussisk Cox proces med følgende parametre.

```
$trend
  (Intercept)        I(x^3)        I(y^3)        I(x^2)        I(y^2)
-2.229980e+05 -7.494374e-15 -1.759690e-15  6.268562e-09  1.668769e-08

$var
[1] 5.560511

$scale
[1] 15547.44
```

Figur 6.31: *Parametre for den fittede inhomogene log Gaussiske Cox proces med stabil korrelation.*

Parametrene på Figur 6.31 specificerer en log Gaussisk Cox proces med kovariansfunktion

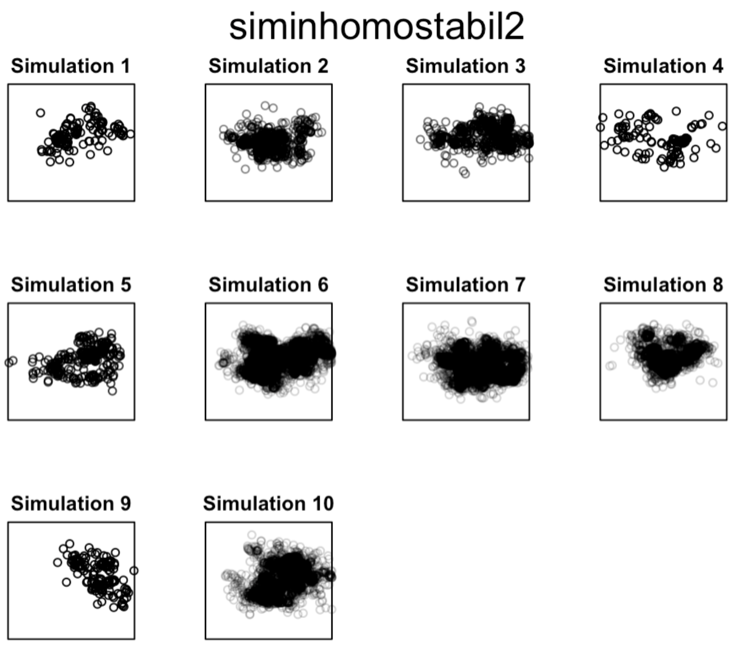
$$c(\xi) = 5.560511 \exp(-\|\xi/15547.44\|^{1/2})$$

og intensitet

$$\begin{aligned}\rho(\xi) = {} & -7.494374 \cdot 10^{-15} x^3 - 1.759690 \cdot 10^{-15} y^3 + 6.268562 \cdot 10^{-9} x^2 \\ & + 1.668769 \cdot 10^{-8} y^2 - 2.229980 \cdot 10^5.\end{aligned}$$

Bemærk at også for modellerne med 3. grads intensitetsfunktion er intensitetsfunktionen den samme i det eksponentielle, Gaussiske og stabile tilfælde.

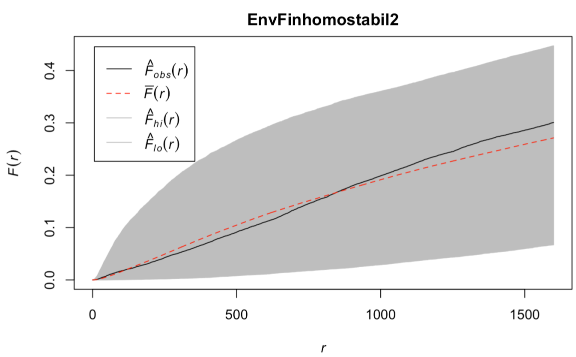
Den inhomogene log Gaussiske Cox proces specificeret ved parametrene på Figur 6.31 simuleres 10 gange, hvilket producerer 10 punktmønstre, som ses på Figur 6.32.



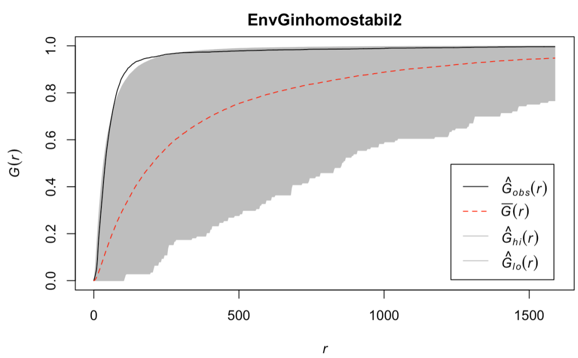
Figur 6.32: *10 simulationer af den inhomogene log Gaussiske Cox proces med stabil korrelation.*

Antallet af punkterne i simulationerne varierer fra 131 punkter for Simulation 1 til 17016 for Simulation 6. Disse punktmønstre udviser, ligesom for den eksponentielle og Gaussiske model, clustertendenser. Det ses, hvordan punkterne forekommer clusterede omkring centrum af observationsvinduet, hvilket minder om punktmønsteret for hussalg i Aalborg Kommune.

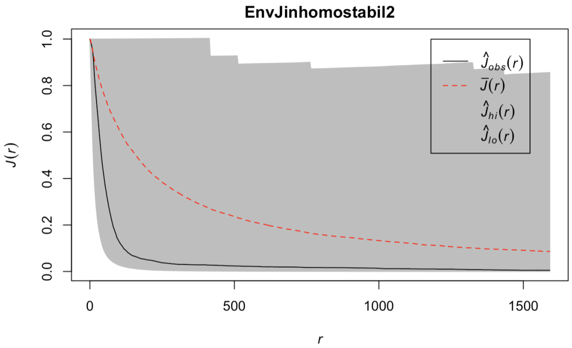
For at undersøge hvor godt modellen givet ved parametrene på Figur 6.31 passer på data, anvendes envelopes for F-, G- og J-funktionen. Envelopes for modellen givet ved parametrene på Figur 6.31 ses på Figur 6.33.



(a) *F-funktionen.*



(b) *G-funktionen.*



(c) *J-funktionen.*

Figur 6.33: *Plot af F-, G- og J-funktionen for den inhomogene log Gaussiske Cox proces med stabil korrelation med tilhørende envelopes.*

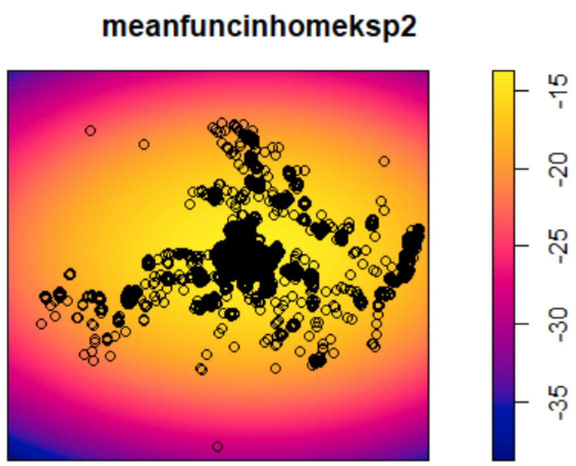
På Figur 6.33a og 6.33c ses at både F- og J-funktionen er indenfor de tilhørende envelopes. Dermed vil ingen værdi af r, for disse to funktioner, resultere i at nulhypotesen afvises, givet at r er valgt på forhånd.

På Figur 6.33b ses at G-funktionen er udenfor envelopes for få værdier af r. Dermed vil nulhypotesen kun blive afvist for få værdier af r, givet at r er valgt på forhånd.

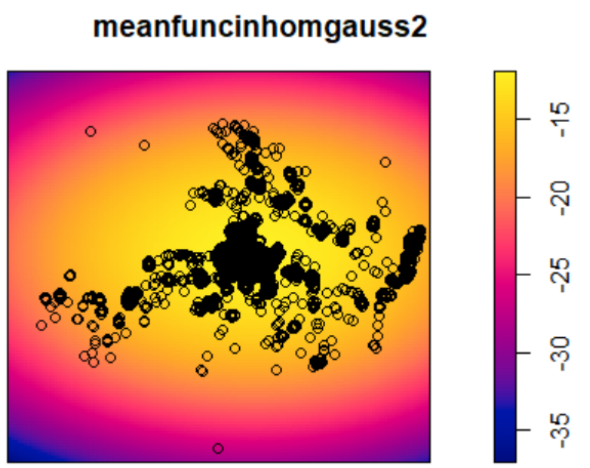
Dette viser, at den stabile model givet ud fra parametrene på Figur 6.31 passer godt på data for de fleste værdier af r.

6.3.3.4 Middelværdifunktioner for modellerne med tredje grads intensitet

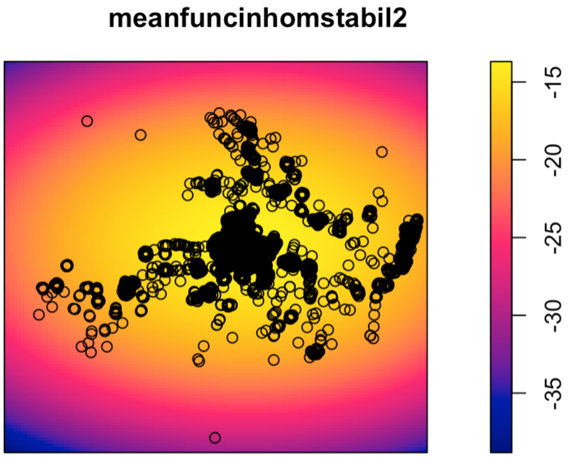
I dette underafsnit undersøges middelværdifunktionerne for de Gaussisk stokastiske felter tilhørende modellerne med 3. grads intensitet. Et plot af middelværdifunktionerne for den eksponentielle, Gaussiske og stabile model kan ses på Figur 6.34 sammen med punktmønsteret for hussalg i Aalborg Kommune.



(a) *Eksponentiel korrelation.*



(b) *Gaussisk korrelation.*



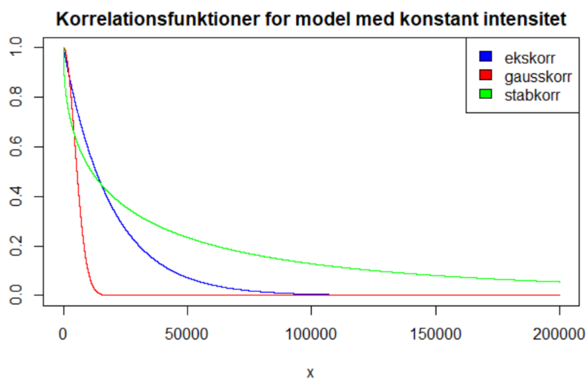
(c) *Stabil korrelation.*

Figur 6.34: *Plot af middelværdifunktionerne for den log Gaussiske Cox proces med tredje grads intensitet.*

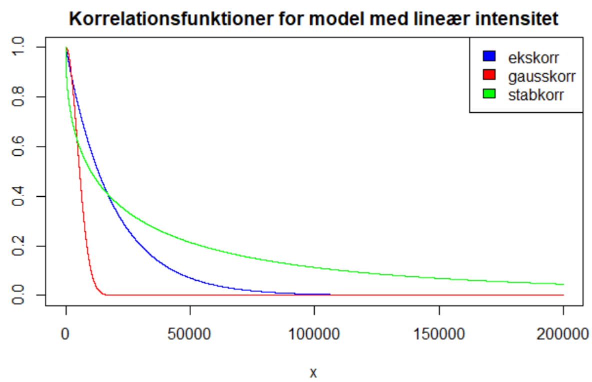
Det ses hvordan middelværdierne på Figur 6.34 stort set er ens. De tre middelværdifunktioner på Figur 6.34a-6.34c antager de højeste værdier omkring midten af observationsvinduet. Dette stemmer overens med, hvordan intensitetsfunktionen for punktmønsteret bestående af hussalg i Aalborg Kommune på Figur 6.3 opfører sig. Bemærk at værdierne for middelværdifunktionerne varierer fra -35 til -15, hvilket indikerer større grad af inhomogenitet end for middelværdifunktionerne for de lineære modeller på Figur 6.24.

6.3.4 Korrelationsfunktioner for modellerne

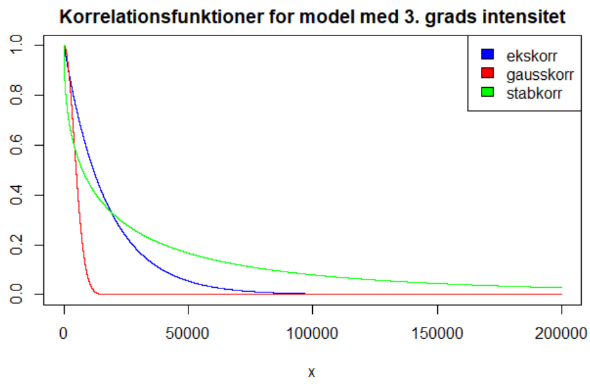
I dette afsnit samles alle korrelationsfunktionerne for de ni forskellige modeller til sammenligning. På Figur 6.35 ses tre plots, som repræsenterer korrelationsfunktionerne for hvert Gaussisk stokastisk felt tilhørende de ni forskellige log Gaussiske Cox processer modelleret i Afsnit 6.3.1-6.3.3.



(a) *Konstant intensitet.*



(b) *Lineær intensitet.*



(c) *Tredje grads intensitet.*

Figur 6.35: *Plot af korrelationsfunktionerne for de Gaussisk stokastiske felter tilhørende de ni forskellige modeller.*

Ved at sammenligne Figurerne 6.35a, 6.35b og 6.35c ses at graferne for henholdsvis de eksponentielle, Gaussiske og stabile korrelationsfunktioner stort set er ens for hver af de tre overordnede modeltyper givet ud fra intensitetsfunktionen. Dette indikerer, at ændringen i intensitetsfunktionen for den log Gaussiske Cox proces ikke påvirker korrelationsfunktionen for det underliggende Gaussisk stokastiske felt specielt meget. Omvendt er det set, hvordan ændringen af korrelationsfunktionerne for hver af de tre intensitetstyper ikke påvirker intensitetsfunktionen.

Ud fra envelopes lader det til at modellerne med eksponentiel korrelation opfører sig bedre end modellerne med stabil og Gaussisk korrelation. I den sammenhæng ses på Figur 6.35, at korrelationsfunktionen går mod nul omkring $r = 100000$ for alle tre modeller med eksponentiel korrelation. Dette indikerer, at måden denne type korrelation aftager på, er den mest passende for punktmønsteret. Det tyder på, at den Gaussiske og stabile korrelation modellerer henholdsvis for meget og for lidt korrelation over længere afstande i forhold til punktmønsteret.

6.3.5 Opsummering af resultater

I Afsnit 6.3 er der fittet ni forskellige log Gaussiske Cox processer til punktmønsteret for hussalg i Aalborg Kommune set på Figur 6.2. Der er fittet modeller med tre forskellige klasser af intensitetsfunktioner hver med tre forskellige korrelationsfunktioner: Den eksponentielle, den Gaussiske og den stabile.

For de tre forskellige intensitetsfunktionsklasser påvirkede ændringen af korrelationsfunktionen ikke intensiteten. Det kan forklares ved måden intensitetsfunktionen estimeres på i henholdsvis det homogene og de inhomogene tilfælde. Desuden er det fælles for hver af de tre intensitetsfunktioner, at modellerne med den Gaussiske korrelationsfunktion har mindst korrelation over større afstande. Derudover har modellerne med den stabile korrelationsfunktion mest korrelation over større afstande for hver af de tre intensitetsfunktioner.

Det er set, hvordan ændringen af intensitetsfunktionen ikke virker til at påvirke korrelationen i det underliggende Gaussisk stokastiske felt specielt meget. Det ses dog, at ændringen i intensiteten påvirker middelværdifunktionen.

Ud af de tre typer af intensitetsfunktioner er det modellerne, hvor log intensiteten har formen $x^3 + y^3 + x^2 + y^2$, at middelværdistrukturen svarer mest til intensitetsplottet for punktmønsteret over hussalg i Aalborg Kommune på Figur 6.3. I den forbindelse ses også, at det er simulationerne for disse log Gaussiske Cox processer, set på Figur 6.26, 6.29 og 6.32, som 'minder mest om' punktmønsteret for hussalg i Aalborg Kommune.

For de log Gaussiske Cox processer med eksponentiel korrelation, ses det for både den homogene og de inhomogene modeller, at F-, G- og J-funktionen ligger indenfor deres tilhørende envelopes i alle tre tilfælde. Det betyder, at de log Gaussiske Cox processer med eksponentiel korrelation passer godt på punktmønsteret for hussalg i Aalborg Kommune, for alle tre modeller af intensitetsfunktionen.

For processerne med stabil korrelationsfunktion ses for den homogene og de inhomogene modeller, at F-funktionen er indenfor de tilhørende envelopes for alle tre modeller. I det homogene tilfælde og det inhomogene tilfælde med 3. grads intensitetsfunktion er G-funktionen udenfor envelopes for nogle værdier af r. Dog er J-funktionen indenfor for alle værdier af r for disse to modeller. For den inhomogene log Gaussiske Cox proces med lineær intensitet er J-funktionen udenfor de tilhørende envelopes for nogle værdier af r. Dette betyder, at de log Gaussiske Cox processer med stabil korrelation passer forholdvis godt på punktmønsteret for alle tre modeller af intensitetsfunktionen. Med dette menes, at de stabile modeller passer godt på datasættet for bestemte værdier af r. Derfor passer modellerne ikke lige så godt på data som modellerne med eksponentiel korrelation.

For de log Gaussiske Cox processer med Gaussisk korrelationsfunktion ses det i tilfældene med konstant og lineær intensitet, at F-, G- og J-funktionen ligger udenfor deres tilhørende envelopes for de fleste værdier af r. For modellen med tredjegrads intensitetsfunktion er F-funktionen udenfor de tilhørende envelopes for næsten alle værdier af r. G-funktionen er indenfor de tilhørende envelopes for næsten alle værdier af r, og J-funktionen er indenfor envelopes for alle værdier af r. Dette betyder, at modellerne med Gaussisk korrelation ikke passer særlig godt på data i det homogene og inhomogene tilfælde med lineær intensitet. Det virker dog til, at modellen med Gaussisk korrelation og tredje grads intensitet passer bedre på data end de Gaussiske modeller med konstant og lineær intensitetsfunktion.

Overordnet set virker de log Gaussiske Cox processer med eksponentiel korrelation til, at være de modeller der passer bedst på punktmønsteret for hussalg i Aalborg Kommune. Dette kan skyldes, at modellerne med Gaussisk og stabil korrelation modellerer henholdsvis for lidt og for meget korrelation over længere afstande i forhold til punktmønsteret.

Konklusion 7

Dette speciale har beskæftiget sig med rumlige punktprocesser, med fokus på log Gaussiske Cox processer. Med henblik på at besvare problemformuleringen, er der introduceret teori omkring Poisson processer og rumlige punktprocesser for at etablere den nødvendige viden krævet for at definere Cox processer og log Gaussiske Cox processer. I den forbindelse er eksistens og entydigheden af Poisson processen blevet bevist.

Det er desuden bevist, at intensitetsfunktionen og parkorrelationsfunktionen for log Gaussiske Cox processer kan udtrykkes udelukkende gennem middelværdien og kovariansen for det underliggende Gaussisk stokastiske felt. I forlængelse heraf er det bevist, at den log Gaussiske Cox proces er homogen, hvis og kun hvis at det underliggende Gaussisk stokastiske felt er stationært.

For at kunne tilpasse de log Gaussiske Cox processer til data, er composite likelihood metoden blevet introduceret. Denne gør brug af andenordens produkttætheden til at konstruere en composite likelihood, som maksimeres ved at finde nulpunktet for den tilhørende scorefunktion. Det er blevet vist, at composite likelihood estimatoren er konsistent for homogene log Gaussiske Cox processer med power eksponentiel korrelation.

Ved anvendelse af composite likelihood metoden er der tilpasset ni forskellige log Gaussiske Cox processer til punktmønsteret for hussalg i Aalborg Kommune. De ni modeller består af tre modeller med forskellige intensitetsfunktioner, hver med tre forskellige korrelationsfunktioner. Der er anvendt punktvise envelopes, for at vurdere hvor godt disse modeller er tilpasset punktmønsteret. Envelopes af denne type fungerer som en punktvis hypotesetest, der kræver at hvert punkt fastsættes, før teststørrelsen udregnes. Denne teststørrelse er givet som en af de summary statistics defineret i Kapitel 3.

Det viste sig, at modellerne med eksponentiel korrelation passer godt på datasættet for alle tre intensitetsfunktioner, og at de tre modeller med stabil korrelation også passer forholdvis godt på data. Modellerne med stabil korrelation tilpasser dog ikke punktmønsteret lige så godt som modellerne med eksponentiel korrelation. De log Gaussiske Cox processer med Gaussisk korrelation passede ikke specielt godt på data. Dog viste modellen med tredje grads intensitetsfunktion sig at være en bedre model for punktmønsteret end de andre modeller i det Gaussiske tilfælde. Altså er det set ud fra punktvise envelopes, at størstedelen af de log Gaussiske Cox processer, undersøgt i specialet, passer godt på punktmønsteret for hussalg i Aalborg Kommune.

Litteratur

[1] Petter Abrahamsen. *A Review of Gaussian Random Fields and Correlation Functions*. Norsk Regnesentral/Norwegian Computing Center, second edition, 1997. ISBN 8253904355, 9788253904351.

[2] Peter Olofsson & Mikael Anderson. *Probability, Statistics and Stochastic Processes*. John Wiley and Sons, Inc., second edition, 2012. ISBN 0470889748, 9780470889749.

[3] Tilahun A Muche & Agegnehu A Atena. Investigating triangular numbers with greatest integer function, sequences and double factorial. *Asia Pacific Journal of Multidisciplinary Research*, 4(4), November 2016.

[4] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton, Fla. : Chapman and Hall/CRC, first edition, 2016. ISBN 9781482210200.

[5] Artur Gramacki. *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Springer International Publishing, first edition, 2018. ISBN 978-3-319-71687-9.

[6] Yongtao Guan. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476), December 2006.

[7] William Kleiber. High resolution simulation of nonstationary gaussian random fields. *Computational statistics and data analysis*, 101, September 2016.

[8] Jesper Møller, Rasmus Waagepetersen, and Anne Randi Syversveen. Log gaussian cox processes. *Scandinavian Journal of Statistics*, 25(3), September 1998.

[9] Zbyněk Pawlas. *Spatial Modelling and spatial statistics*, 2008. Tilgængelig på http://web.math.ku.dk/~mogens/rumlig.pdf. Hentet d. 27/11-2019.

[10] Ebbe Thue Poulsen. *Funktioner af en og flere variable*. Institut for Matematik, Aarhus Universitet, first edition, 2015. ISBN 8712038261.

[11] Morten Grud Rasmussen. *Punktmængdetopologi, metriske rum, fuldstændighed*, 2015. Tilgængelig på http://people.math.aau.dk/~morteng/Homepage_files/E15A1/a1fl1512.pdf Hentet d. 2/12-2019.

[12] Steen Thorbjørnsen. *Grundlæggende mål-og integralteori*. Aarhus Universitetsforlag, første edition, 2014. ISBN 978-87-7124-508-0.

[13] D.J. Daley & D. Vere-Jones. *An Introduction to the Theory of Point Processes; Volume 2: General Theory and Structure*. Springer, New York, second edition, 2008. ISBN 978-0-387-21337-8, 978-0-387-49835-5.

[14] Jesper Møller & Rasmus Plenge Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, first edition, 2003. ISBN 1-58488-265-4.

[15] Wikipedia. *Particular values of the gamma function*, 2019. Tilgængelig på https://en.wikipedia.org/wiki/Particular_values_of_the_gamma_function. Hentet d. 27/9-2019.

Appendiks

Målteori A

A.1 σ-algebra og mål

I dette appendiks er der tilføjet visse definitioner og lemmaer, der ligger udenfor specialets teoretiske ramme, men som stadig anvendes i forbindelse med visse resultater.

Først introduceres målteoretiske begreber, for at give mening til definitionen af punktprocesser. Først defineres en σ-algebra. Hertil benyttes en ikke-tom mængde X.

Definition A.1.1 (σ-algebra)
Et system $\mathcal{E}$ af delmængder af X kaldes en σ-algebra i X, hvis det opfylder følgende betingelser.

1. $X \in \mathcal{E}$.
2. For alle mængder $A \in \mathcal{E}$ gælder $A^c \in \mathcal{E}$.
3. Hvis $(A_n)_{n\in\mathbb{N}}$ er en følge af mængder i $\mathcal{E}$, da gælder at $\cup_{n\in\mathbb{N}}(A_n) \in \mathcal{E}$.

[12][s. 1]

Mængderne i $\mathcal{E}$ kaldes $\mathcal{E}$-målelige mængder. Parret $(X, \mathcal{E})$ kaldes da et måleligt rum. Ud fra dette kan begrebet *mål* nu defineres.

Definition A.1.2 (Mål)
Lad $(X, \mathcal{E})$ være et måleligt rum. Et mål på $(X, \mathcal{E})$ er da en afbildning $\mu : (X, \mathcal{E}) \to [0, \infty]$, som opfylder to betingelser.

1. $\mu(\emptyset) = 0$.
2. For enhver følge af disjunkte mængder $(A_n)_{n\in\mathbb{N}}$ i ξ gælder at

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n).$$

Hvis μ er et mål på $(X, \mathcal{E})$, kaldes $(X, \mathcal{E}, \mu)$ et målrum [12][s. 13].

Hvis $(X, \mathcal{E}, \mu)$ er et målrum, og $p : X \to \{\text{sandt, falsk}\}$ er en egenskab blandt elementerne i X. Da siges p at være opfyldt *μ-næsten overalt* hvis

$$\mu\big(\{x \in X | p(x) = \text{falsk}\}\big) = 0.$$

Altså at p er sand undtagen i en μ-nulmængde. Hvis μ er et sandsynlighedsmål, siges p at være opfyldt *μ-næsten sikkert* eller *med sandsynlighed et*.

For at bevise Sætning 2.1.4 skal følgende lemmaer anvendes. Givet sandsynlighedsrummet $(\Omega, \mathcal{F}, \mathcal{P})$ defineres $\mathcal{A}$ som en mængde af delmængder $A \subseteq \Omega$ og $\sigma(\mathcal{A})$ skrives som σ-algebraen frembragt af $\mathcal{A}$.

Lemma A.1.3
Lad μ_1 og μ_2 være mål defineret på Ω udstyret med σ-algebraen $\mathcal{F} = \sigma(\mathcal{A})$ for en fællesmængdestabil mængde af delmængder $\mathcal{A}$. Hvis $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ på $\mathcal{A}$, så er $\mu_1 = \mu_2$ på $\mathcal{F}$ [14][s. 241].

Bevis
For bevis se [12][s. 29]. ■

For at indføre Lemma A.1.5 defineres først en *tæt* delmængde.

Definition A.1.4 (Tæt)
En delmængde A af rummet X siges at være tæt (i X), hvis enhver ikke-tom åben delmængde $U \subseteq X$ opfylder at $A \cap U \neq \emptyset$ [11] [s. 2].

En mængde A kaldes *separabel* hvis A indeholder en tællelig tæt mængde. Lad nu $\mathcal{N}_{le}$ være defineret som i Definition 2.1.2. Da gives

$$\mathcal{N}_{le}^0 = \{\{x \in N_{le} : n(x_B) = 0\} : B \in \mathcal{B}_0\} \tag{A.1}$$

som klassen af void hændelser.

Lemma A.1.5
Lad $S \subseteq \mathbb{R}^d$ være separabel. Da gælder at $\mathcal{N}_{le} = \sigma(\mathcal{N}_{le}^0)$ [14][s. 243].

Bevis
For bevis se [14][s. 243]. ■

Følgende lemma anvendes i forbindelse med beviset for Slivnyak-Meckes Sætning 2.2.16, hvor der desuden gøres brug af definitionen på mål.

Lemma A.1.6
Lad $S \subseteq \mathbb{R}^d$ og lad N_{le} være defineret som i Definition 2.1.2.

(i) Lad μ og ν være mål på $S \times N_{le}$. Da er $\mu = \nu$, givet at $\mu(B \times N_{le}) = \nu(B \times N_{le}) < \infty$ og $\mu(B \times F) = \nu(B \times F)$, for alle begrænsede $B \subseteq S$ og alle $F \in \mathcal{N}^0_{le}$.

(ii) Lad μ og ν være mål på $N_{le} \times N_{le}$. Da er $\mu = \nu$, givet at $\mu(N_{le} \times N_{le}) = \nu(N_{le} \times N_{le}) < \infty$ og $\mu(F_1 \times F_2) = \nu(F_1 \times F_2)$, for alle $F_1, F_2 \in \mathcal{N}^0_{le}$.

[14][s. 245]

Bevis
For bevis se [14][s. 245]. ■

A.2 Ergodicitet

Dette afsnit er baseret på [13].
Følgende definition viser sig nødvendig for at opskrive et resultat for asymptotiske egenskaber for composite likelihood estimatoren. Lad først $M_S^\#$ være rummet bestående af alle begrænsede endelige mål på $\mathcal{B}(S)$, hvor $\mathcal{B}$ noterer Borel-algebraen og $S = \mathbb{R}^d$. Lad da

$$N_S^\# = \{\mu \in M_S^\# \mid \mu \text{ er tællemål } \} \subseteq M_S^\#.$$

Definer da

$$T_u x = x + u \quad \text{og} \quad T_u A = A + u = \{x + u \mid x \in A\}$$

for $x, u \in S$ og $A \in \mathcal{B}(S)$. Operatoren T_u inducerer en transformation τ_u af $M_S^\#$ og $N_S^\#$ gennem

$$(\tau_u \mu)(A) = \mu(T_u A)$$

for et mål μ i $M_S^\#$ eller $N_S^\#$ og $A \in \mathcal{B}(S)$. Herefter lades $\mathbb{U}^d_{2a}$ være den d-dimensionelle hyperkube med sidelængde $2a$ og hjørnepunkter $(\pm a, \dots, \pm a)$ og $\mathcal{P}$ være et sandsynlighedsmål for en punktproces på S.

Definition A.2.1 (Ergodisk)
En punktproces X på S siges at være ergodisk, hvis der for alle $V, W \in \mathcal{B}(N_S^\#)$ gælder

$$\frac{1}{l(\mathbb{U}^d_a)} \int_{\mathbb{U}^d_a} \Big(\mathcal{P}\big((\tau_x)(V \cap W)\big) - \mathcal{P}(V)\mathcal{P}(W)\Big) dx \to 0 \quad \text{for} \quad a \to \infty,$$

hvor $l()$ noterer Lebesgue målet på $\mathbb{R}^d$.

Stokastiske felter B

B.1 Stokastiske felter

Dette er baseret på [1][s. 4].
Dette afsnit omhandler stokastiske felter introduceret i Definition 4.1.1. Disse noteres for nemhedens skyld ved $Z = Z_{\mathbf{t}} = Z(\mathbf{t}, \omega)$ for $\mathbf{t} \in \mathbb{R}^n$. Generelt kan et stokastisk felt beskrives ud fra dets *endelige dimensionelle fordelinger*, som er givet

$$F_{\mathbf{t}_1,\dots,\mathbf{t}_k}(z_1, \dots, z_k) = P(Z_{\mathbf{t}_1} \leq z_1, \dots, Z_{\mathbf{t}_k} \leq z_k). \tag{B.1}$$

Der er dog to betingelser til disse endelig dimensionelle fordelinger. Ved en permutation π på indeksmængden $\{1, \dots, k\}$ skal *symmetribetingelsen*

$$F_{\mathbf{t}_1,\dots,\mathbf{t}_k}(z_1, \dots, z_k) = F_{\mathbf{t}_{\pi 1},\dots,\mathbf{t}_{\pi k}}(z_{\pi 1}, \dots, z_{\pi k}) \tag{B.2}$$

og *kompatibilitetsbetingelsen*

$$F_{\mathbf{t}_1,\dots,\mathbf{t}_{k-1}}(z_1, \dots, z_{k-1}) = F_{\mathbf{t}_1,\dots,\mathbf{t}_{k-1},\mathbf{t}_k}(z_1, \dots, z_{k-1}, \infty) \tag{B.3}$$

være opfyldt for de endelig dimensionelle fordelinger. Jævnfør [1][s. 4], er betingelserne (B.2) og (B.3) opfyldt for et stokastisk felt givet som i Definition 4.1.1. Det omvendte viser sig også at være sandt. Det vil sige, hvis der findes fordelingsfunktioner som opfylder (B.2) og (B.3), er disse endelig dimensionelle fordelinger for et stokastisk felt jævnfør [1][s. 4]. Dette resultat er hvad der kendes som Kolmogorovs Eksistens Sætning.

B.1.1 Gaussisk stokastiske felter

Dette underafsnit er baseret på [1][s. 7-8].
Lad et Gaussisk stokastisk felt Y være defineret som i Definition 4.2.1. Per definition er de endelig dimensionelle fordelinger for Y multivariate normalfordelinger. Disse kan bestemmes ud fra multivariate normalfordelings tætheder

$$p_{\mathbf{t}_1,\dots,\mathbf{t}_k}(x_1, \dots, x_k) = |2\pi\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m})\Big)$$

hvor $\mathbf{m}^T = \big[m(\mathbf{t}_1), \dots, m(\mathbf{t}_k)\big]$ er middelværdivektoren og $\Sigma_{ij} = c(\mathbf{t}_i, \mathbf{t}_j)$ er indgange i kovariansmatricen. Dermed kan de endelig dimensionelle fordelinger for Gaussisk stokastiske felter specificeres ud fra middelværdien m og kovariansfunktionen c jævnfør [1][s. 7-8].

R-kode C

C.1 R-kode til dataklargøring

Nedenstående R-script omdanner længde-og breddegradskoordinater til UTM koordinater, og lagrer til sidst delmængder af boligssalg fra Aalborg Kommune i en ny CSV-fil.

Listing C.1: Omregning til UTM koordinater.

```
library(dawa)
HOME <- read.csv("Redigeret-Dataudtraek␣2004-2016.csv", sep = ";",
stringsAsFactors=FALSE,
                  encoding = "latin1")

koordinater <- read.csv("koordinater.csv")

nyHOME=HOME
nyHOME$id=0
nyHOME$xkoor=0
nyHOME$ykoor=0

for (i in 159230:length(koordinater$id)) {
  ad <- HOME$FuldAdresse[i]
  post <- HOME$Postnr[i]
  by <- HOME$Bynavn[i]
  tempadr <- cbind(ad,post,by)
  adr <- toString(tempadr)
  v <- datavask(adr,"adresser")
  adrid <- v$resultater[[1]]$adresse$id
  h <- adresser(id=adrid, struktur="mini",srid=25832)
    if(is.null(h[1][[1]])==TRUE){next}
  koordinater$id[i] <- adrid
  koordinater$xkoor[i] <- h[[1]]$x
  koordinater$ykoor[i] <- h[[1]]$y
}

koordinater <- subset(koordinater, select = -c(X.2,X.1,X))

write.csv(koordinater, file = "koordinater.csv", row.names = FALSE)
```

```
HOME$id=koordinater$id
HOME$xkoor=koordinater$xkoor
HOME$ykoor=koordinater$ykoor

Aalborg <- subset(HOME, Kommune=="Aalborg")

Aalborg <- subset(Aalborg, xkoor!=0)

write.csv(Aalborg, file = "Aalborg.csv", row.names = FALSE)
```

Herefter indlæses CSV filen for datasættet i et nyt R-script.

Listing C.2: Indlæsning af Aalborg datasættet.

```
Aalborg <- read.csv("Aalborg.csv")
```

Før at `spatstat`-pakken kan bearbejde datasættet, skal det omdannes til `ppp`-format, hvilket står for *planar point pattern*. For at danne dette punktmønster i $\mathbb{R}^2$ udtrækkes x og y-koordinaterne af datasættet.

Listing C.3: Udtræk af x- og y-koordinater.

```
u <- as.vector(Aalborg["xkoor"])
v <- as.vector(Aalborg["ykoor"])
c <- as.numeric(unlist(u))
d <- as.numeric(unlist(v))
```

Herefter dannes et observationsvindue W, som skal indeholde alle punkter fra datasættet. Grænserne er fundet ud fra maksimums- og minimumsværdier i x- og y-koordinaterne.

Listing C.4: Observationsvinduet.

```
ut <- as.vector(c(520000,585000))
vt <- as.vector(c(6220000,6346000))
W <- owin(ut,vt)
```

Kommandoen `ppp` laver ud fra dataen og observationsviduet et punktmønster i $\mathbb{R}^2$, og `unique` fjerner dobbeltpunkter i datasættet. Punktmønsteret plottes gennem kommandoen `plot`.

Listing C.5: Klargøring til plot af punktmønster.

```
HuspriserAalborg <-  ppp(c,d, W)
HuspriserAalborgun <- unique(HuspriserAalborg)
plot(HuspriserAalborgun)
```

Plottet produceret af koden i Listing C.5 kan ses på Figur 6.1. Det bemærkes, at der forekommer to outliers i punktmønsteret, og disse fjernes.

Listing C.6: Outliers fjernes.

```
 Aalborgny <- subset(Aalborg,ykoor!=min(ykoor))
Aalborgny1 <- subset(Aalborgny,ykoor!=min(ykoor))
```

Herefter dannes det nye punktmønster ud fra datasættet uden outliers.

Listing C.7: Punktmønsteret uden outliers.

```
a <- as.numeric(unlist(x))
b <- as.numeric(unlist(y))

xt <- as.vector(c(520000,585000))
yt <- as.vector(c(6290000,6350000))
O <- owin(xt,yt)

HuspriserAA <- ppp(a,b, O)
HussalgAalborg <- unique.ppp(HuspriserAA)

plot(HussalgAalborg)
```

Plottet produceret af Listing C.7 kan ses på Figur 6.2.